



**Universidad  
Norbert Wiener**

**FACULTAD DE INGENIERÍA Y NEGOCIOS  
ESCUELA ACADÉMICO PROFESIONAL DE INGENIERÍAS**

**Tesis**

**Perfilamiento y limpieza para optimizar la gestión de datos la  
plataforma web de una empresa de ventas, Lima 2020  
Para optar el título profesional de Ingeniero de Sistemas e  
Informática**

**AUTORA**

Br. Gonzales León, Andrea Carolina

ORCID: 0000-0002-7691-5819

**LÍNEA DE INVESTIGACIÓN GENERAL DE LA UNIVERSIDAD**

Ingenierías de Sistemas e Informática, Industrial y Gestión Empresarial y  
Ambiental

**LÍNEA DE INVESTIGACIÓN ESPECÍFICA DE LA UNIVERSIDAD**

Seguridad en Base de Datos

**LIMA - PERÚ**

**2020**

## **Miembros del Jurado**

Presidente del Jurado

Dr. David Flores Zafra

ORCID: 0000-0001-5846-325X

Secretario

Mg. Luis Enrique Ramirez Pacheco

ORCID: 0000-0001-7134-0457

Vocal

Mtro. Cesar Antonio Porras Ramirez

ORCID: 0000-0002-6882-5194

Asesor metodólogo

Dr. Fernando Alexis Nolzco Labajos

ORCID: 0000-0001-8910-222X

Asesor temático

Mg. Walter Amador Chávez Alvarado

ORCID: 0000-0001-8614-482X

### **Dedicatoria**

A mis padres y hermana por su apoyo contante en mi vida, y por creer en mi capacidad para cumplir mis metas.

A mi familia y amigos que me brindaron sus consejos para poder alcanzar y cumplir mis objetivos.

### **Agradecimiento**

Agradezco a Dios por guiarme en este camino, para poder lograr mis proyectos. A mi casa de estudios la Universidad Norbert Wiener por haber aportado en mi aprendizaje, y a la facultad de ingeniería por guiar esta investigación.

## Declaración de autoría

 Universidad Norbert Wiener	<b>DECLARACIÓN DE AUTORIA</b>	
	CÓDIGO: UPNW-EES-FOR-017	VERSIÓN: 01 REVISIÓN: 01

Yo, Gonzales León Andrea Carolina estudiante de la escuela académica profesional de Ingenierías de la Universidad Privada Norbert Wiener, declaro que el trabajo académico titulado: "Perfilamiento y limpieza para optimizar la gestión de datos la plataforma web de una empresa de ventas, Lima 2020" para la obtención del Título Profesional de Ingeniería de Sistemas e Informática es de mi autoría y declaro lo siguiente:

1. He mencionado todas las fuentes utilizadas, identificando correctamente las citas textuales o paráfrasis provenientes de otras fuentes.
2. No he utilizado ninguna otra fuente distinta de aquella señalada en el trabajo.
3. Autorizo a que mi trabajo puede ser revisado en búsqueda de plagios.
4. De encontrarse uso de material intelectual ajeno sin el debido reconocimiento de su fuente y/o autor, me someto a las sanciones que determina los procedimientos establecidos por la UPNW.



.....  
Firma  
Gonzales León, Andrea Carolina  
DNI: 70546668



Huella

Lima, 30 de Junio de 2020.

## Índice

	<b>Pág.</b>
Miembros del Jurado	ii
Dedicatoria	iii
Agradecimiento	iv
Declaración de autoría	v
Índice	vi
Índice de tablas	viii
Índice de figuras	ix
Índice de cuadros	xi
Resumen	xii
Abstract	xiii
I. INTRODUCCIÓN	14
II. MÉTODO	21
2.1. Enfoque y tipo	21
2.2. Población, muestra y unidades informantes	22
2.3. Categorías, sub categorías apriorísticas y emergentes	23
2.4. Técnicas e instrumentos de recolección de datos	23
2.5. Proceso de recolección de datos	24
2.6. Método de análisis de datos	24
III. RESULTADOS	25
3.1. Descripción de resultados cuantitativos	25
3.2. Descripción de resultados cualitativos	35
3.3 Diagnóstico	39
3.4. Propuesta	42
IV. DISCUSIÓN, CONCLUSIONES Y RECOMENDACIONES	59
4.1 Discusión	59
4.2 Conclusiones	62
4.3 Recomendaciones	64
REFERENCIAS	66
ANEXOS	71
Anexo 1: Matriz de la investigación	72
Anexo 2: Evidencias de la propuesta	73

Anexo 3: Instrumento cuantitativo	75
Anexo 4: Instrumento cualitativo	76
Anexo 5: Base de datos (instrumento cuantitativo)	77
Anexo 6: Transcripción de las entrevistas o informe del análisis documental	78
Anexo 7: Pantallazos del Atlas. Ti	84
Anexo 8: Fichas de validación de la propuesta	85
Anexo 9: Matrices de trabajo	87

## Índice de tablas

	<b>Pág.</b>
Tabla 1 Matriz de categorización de gestión de datos	23
Tabla 2 Matriz de Sub Categorías emergentes	23
Tabla 3 Frecuencia (f) y porcentaje (%) de los patrones del campo número de consulta de enero, febrero y marzo del año 2019	25
Tabla 4 Frecuencia (f) y porcentaje (%) de los patrones en el campo número de orden enero, febrero y marzo del año 2019	27
Tabla 5 Frecuencia (f) y porcentaje (%) de los patrones de DNI de usuario pdv	28
Tabla 6 Frecuencia (f) y porcentaje (%) de los registros de atención ingresados de enero a junio del año 2019	30
Tabla 7 Frecuencia (f) y porcentaje (%) del acceso a la plataforma web de enero a junio del año 2019	32
Tabla 8 Frecuencia (f) y porcentaje (%) de consultas de usuarios pdv de enero a junio del año 2019.	33

## Índice de figuras

	<b>Pág.</b>
Figura 1. Porcentaje (%) de los patrones del campo número de consulta de enero, febrero y marzo del año 2019	26
Figura 2. Porcentaje (%) de los patrones en el campo número de orden enero, febrero y marzo del año 2019	28
Figura 3. Porcentaje (%) de los patrones de DNI de usuario pdv	29
Figura 4. Porcentaje (%) de los registros de atención ingresados de enero a junio del año 2019	31
Figura 5. Porcentaje (%) del acceso a la plataforma web de enero a junio del año 2019	33
Figura 6. Porcentaje (%) de consultas de usuarios pdv de enero a junio del 2019	34
Figura 7. Red de sub categoría calidad de datos	35
Figura 8. Red de sub categoría almacenamiento de datos	37
Figura 9. Red de sub categoría recuperación de información	38
Figura 10. Análisis mixto de la sub categoría calidad de datos	40
Figura 11. Análisis mixto de la sub categoría almacenamiento de datos	41
Figura 12. Análisis mixto de la sub categoría recuperación de información	42
Figura 13. Documento de perfilamiento de datos - descripción y dimensiones	49
Figura 14. Documento de perfilamiento de datos - reglas de perfilamiento y calidad	49
Figura 15. Perfilamiento de los datos de una tabla	49
Figura 16. Vista general del perfilamiento de los datos de un campo	50
Figura 17. Patrones	50
Figura 18. Reglas en el perfilamiento de datos	51
Figura 19. Crear regla en el perfilamiento de datos	51
Figura 20. Aplicar regla en el perfilamiento de datos	52
Figura 21. Inicio de regla específica	52
Figura 22. Plataforma de la regla específica	53
Figura 23. Creación de regla específica	53
Figura 24. Editar perfilamiento de la tabla	54
Figura 25. Aplicación de la regla específica	54
Figura 26. Comparación de perfilamiento después de la limpieza	54
Figura 27. Creación del scorecarding	55

Figura 28. Renombrar campos para el scorecarding	55
Figura 29. Seleccionar campos para el scorecarding	56
Figura 30. Validar campos para el scorecarding	56
Figura 31. Documento de perfilamiento de datos - objetivos	57
Figura 32. Ingreso de métricas	57
Figura 33. Vista de scorecarding inicial	58
Figura 34. Vista de scorecarding final	58

## Índice de cuadros

	<b>Pág.</b>
Cuadro 1. Matriz de diagnóstico y priorización de los problemas.	43
Cuadro 2. Matriz de direccionalidad de la propuesta	45
Cuadro 3. Matriz de tácticas, actividades y cronograma	47

## Resumen

El presente estudio, titulado Perfilamiento y limpieza para optimizar la gestión de datos la plataforma web de una empresa de ventas, Lima 2020; tiene como objetivo general proponer el perfilamiento y limpieza para optimizar la gestión de datos; y los objetivos específicos son analizar la calidad de los datos, diagnosticar la efectividad del almacenamiento y recuperación de datos e identificar si la calidad de los datos, el almacenamiento o la recuperación de los datos perjudica la eficiencia en la plataforma web de una empresa de ventas.

Se empleó el sintagma holístico con un enfoque mixto de tipo proyectiva, con nivel comprensivo y adaptando el método deductivo e inductivo. Se estableció como muestra a tres colaboradores de la empresa, y la información registrada. Se dispuso una guía como instrumento para elaborar la entrevista. Se aplicó Excel y ATLAS.ti 8 para procesar la información.

Elaborada la triangulación de los resultados, se diagnosticó la presencia de anomalías en los datos, que afectan la calidad de los mismos. Se brindó una propuesta que mejora y optimiza la gestión de datos, aplicando perfilamiento y limpieza de datos. Lo que permitirá, que la empresa emplee su información con confianza, mejorando procesos y brindando un mejor servicio, para lograr la fidelidad por parte de sus clientes.

**Palabras clave:** Gestión de datos, Calidad de Datos, Almacenamiento de Datos, Recuperación de Datos, Perfilamiento y limpieza de Datos.

## Abstract

The present study, with the title Profiling and cleaning to optimize the data management of the web platform of a sales company, Lima 2020; has the objective of proposing the profiling and cleaning to optimize the data management; and it has as specific objectives that are to analyze data quality, diagnose the effectiveness of data storage and data retrieve and identify whether data quality, data storage or data retrieval impairs efficiency on a sales company's web platform.

It was used the holistic syntagma whit a mixed projective approach, with a comprehensive level and adapting the deductive and inductive method. It has the area of systems as a population, and as a sample, three collaborators of the company, and the registered information. An interview guide was used as an instrument to prepare the interview. It was used the Excel and ATLAS.ti 8 to process the information.

With the triangulation of the results, the presence of anomalies in the data was diagnosed, which affects the quality of the data. A proposal was provided that improves and optimizes data management, applying data profiling and cleaning. Allowing the company to use its own information with confidence, improving processes and providing a better service, to achieve loyalty from its customers.

**Key words:** Data management, Data Quality, Data Storage, Data Retrieve, Data profiling and data cleaning.

# I. INTRODUCCIÓN

En nuestro país, la gestión de datos está presente en diferentes empresas públicas y privadas, porque reconocen la importancia de los datos; por lo cual, son responsables de asignarles su valor. Lamentablemente, muchas empresas son indiferentes y no le dan el valor que corresponde. No obstante, con el cambio en la relación de las empresas y su público a causa de las nuevas tecnologías, la ausencia de buenas prácticas para la gestión de datos puede perjudicar la comunicación entre ambos. De la misma forma, la calidad y el contenido en los sistemas de información no cumplen el desempeño esperado; a causa de la falta de interés por fortalecer la información en los repositorios; en consecuencia, la posibilidad de explotar la enorme cantidad de datos para generar información, se reduce. Razón por la cual, se considera necesario validar la seguridad y confiabilidad de un sistema informático, por esto se debe verificar la información obtenida, para poder determinar la calidad de los datos (Curioso, Henríquez & Espinoza, 2018; Vargas y otros, 2018; Pereira, Portilla & Rodríguez, 2019).

Por otra parte, la gestión de datos incorrectos también es un problema, porque es un riesgo para cualquier empresa, puesto que, la presencia de datos o información incorrecta puede afectar su reputación, dañando la relación con los clientes y proveedores, y el sustento económico de la misma. Simultáneamente, la carencia en mejoras de gestión de datos, dificulta la obtención de información confiable para la toma de decisiones y mejora en las operaciones. Sobre todo, en las Instituciones de Educación Superior, que a pesar de ser organizaciones que genera conocimiento, no manejan una adecuada gestión y control de calidad de datos, dejando a los mismos propensos a encontrarse dispersos, duplicados, incompletos o en diferentes formatos. Por otra parte, la ausencia de medidas y controles adecuados para manejar la gran cantidad de información disponible, puede abrumar al usuario que la maneja, perdiendo la oportunidad de obtener el conocimiento que los datos y la información pueden brindar (Álvarez, Romero, Cadenas, Coronado & Rodríguez, 2016; Osorio, Guerrero & González, 2017; Deloitte Consulting Group S.C., 2018).

Para poder obtener antecedentes de la investigación, se revisaron artículos y tesis de origen nacional e internacional. Para empezar, el primer estudio a mencionar, hecho en

Cuba, nos explican que el sistema estadístico informatizado, empleado por el departamento de cuadros de la Escuela Nacional de Salud Pública (ENSAP) para recoger datos de docentes, no cumple los requerimientos solicitados por los jefes del departamento de docentes; debido a que no pueden disponer de un flujo de información completo de cada profesor; en consecuencia, impide el desarrollo de estrategias de crecimiento individual. Además, no poseen una base de datos estructurada, en la cual puedan registrar los datos del profesor, tanto personales como relacionados con su labor (Santana, Muñoz, O'farrill, Martínez, & Martínez, 2017). Así mismo, en la tesis elaborada por Cabrera (2019) detectó las deficiencias en los procesos de gestión de datos a causa de la dificultad de organizar y controlar el gran volumen de datos de biopsias oncológicas que se generan cada año en el Centro Oncológico de Holguín, con información almacenada de forma física y riesgo de pérdida. Por consiguiente, los datos registrados y con posibles errores, dificulta el trabajo de los especialistas del centro y afecta el bienestar de los pacientes. De igual manera, en la tesis realizada por González & Hernández (2017) se utilizó como muestra, a los usuarios de las áreas del negocio que solicitan reportes de información a *Business Intelligence* (BI) para determinar que la empresa presenta problemas en sus grandes volúmenes de información y la forma de acceder a los mismos, causando pérdidas económicas. Además, el gran número de inconsistencias que presentan sus datos, dañan la confiabilidad en los reportes de información; por consiguiente, realizan gastos para el reproceso de los mismos en el área de BI. Por otra parte, en la elaboración de la tesis, Jiménez (2017) utilizó como muestra la tabla de clientes de la base de datos de una empresa del sector financiero y determinó que la problemática en la gestión de datos corresponde a inconsistencia en los datos de clientes, puesto que la misma presenta ausencia de limpieza e integridad de datos, impidiendo el análisis para la búsqueda de soluciones y tomando decisiones que puede perjudicar y poner en riesgo a la empresa. Así mismo, el manejo de un elevado volumen de datos, ausente de calidad de datos puede perjudicar a la empresa con pérdidas económicas y el cierre de la misma.

Se debe agregar que, en Perú, Balaguer & Arce (2019) en su tesis realizada en el Banco BBVA Continental, determinó que diferentes entidades presentan desconfianza en la información que manejan, a causa de la baja seguridad y confiabilidad en sus datos, puesto que los responsables de garantizar la calidad de información y estándares de confiabilidad no cumplen su rol. De igual modo, la ausencia de conocimiento en estándares

y principios de calidad de datos, normas, sistemas de medición, etc. es otro factor que garantiza que la información que utilizan no es segura, confiable y auténtica, y puede ser perjudicial en caso se utilice para la toma de decisiones. Así mismo, Sandoval (2018) en la tesis elaborada en una Institución Educativa Limeña, logró identificar inconsistencia en los datos de alumnos, puesto que presentan un bajo nivel de calidad de datos e información. Por esto, es prioridad el mantener una buena calidad, puesto que son recursos valiosos; sin embargo, algunas empresas ignoran el impacto y las consecuencias de una baja calidad de datos, que puede afectar el desempeño, la relación con sus clientes, la toma de decisiones y los planes estratégicos en una empresa. Por otro lado, en la tesis elaborada por Chilón (2018) para la empresa Corporación el Dorado, dedicada a la venta de materiales de construcción y vidriería, detectó el manejo deficiente de la información, razón por la cual presenta constantes quejas de sus clientes y sus trabajadores, causados por la insatisfacción al cliente y dando como resultado la devolución de equipos; y la dificultad para cumplir sus funciones, en las áreas de operaciones y producción, a causa de la deficiente calidad y manejo de la información. Así mismo, presentan obstáculos en el flujo de la información dentro de la empresa, y problemas en la gestión documentaria relacionados a duplicidad de datos externos e internos. Así, por ejemplo, la duplicidad en atención al cliente. De igual modo, Dueñas (2018) utilizó como muestra la Base de Datos del Instituto de Informática de la Universidad Nacional San Antonio de Abad del Cusco (UNSAAC), observó la presencia de inconsistencias a causa de campos en blanco, datos incorrectos, duplicados, datos con formatos diferentes, etc; que dañan la unificación y migración de los datos de un repositorio a otro, por la ausencia de una adecuada limpieza de datos.

Para sustentar la investigación, se utiliza la teoría de la información, en donde nos orienta a identificar la información como un elemento necesario, para elaborar y estructurar propuestas de soluciones y mejoras (Campbell citado en Aladro, 2011). Asimismo, se respalda la importancia de la información, orientando a incursionar en la investigación de la misma, para que el conocimiento obtenido pueda ser encaminado a un entorno en donde sea compartido (Valbuena citado en Monsalve, 2003). De modo que, esta teoría le permitirá a la investigación, lograr que la empresa entienda la importancia de la información e investigación de la misma, puesto que proporciona conocimiento variado, para aportar propuestas de solución y/o optimización en beneficio de la empresa. Pero para

ello, la información que manejan requiere seguridad y confiabilidad, y eso se obtiene manejando datos de calidad.

Por otra parte, la teoría de la decisión, respaldada por Aguiar (2004) apoya el estudio de acciones y nuevas perspectivas de decisión, para seleccionar entre un conjunto de opciones, aquella que proporcione el mejor resultado según sus preferencias, y aplicarlas en diferentes entornos, que pueden ser: afrontando problemas frecuentes que impactan la vida diaria y entornos informativos en el cual un individuo se pueda desplegar. Por lo cual, la teoría aportará en la investigación, a la empresa y sus organizadores, la toma de decisiones con base a los conocimientos que poseen, logrando analizar sus problemas y reconocer el estado en el que se encuentran, para identificar soluciones y nuevas propuestas que se puedan implementar, seleccionando entre todas las opciones posibles, aquella que beneficie y cumpla los requerimientos de la empresa.

La presente investigación tiene como categoría principal la gestión de datos, que se entiende como el encargado de administrar y aprovechar los datos empleados en procedimientos, procesos y proyectos, relacionados directa o indirectamente con los datos y buscando la mejora, la entrega, el control y la protección de los mismos; debido que junto a la información, son activos de valor para el negocio. Así mismo, la gestión de datos se emplea como una herramienta, que beneficia a la empresa en la gestión de sus distintas áreas, generando valor y beneficio (DAMA International, 2009; Osorio, Guerrero & González, 2017; Portilla, 2017). Por otra parte, también se encarga de la construcción de información, empleando datos medibles para obtener información valiosa que pueda ser aprovechada en beneficio de la empresa. Asimismo, se considera un factor clave, debido al efecto que puede tener en la creación o gestión del conocimiento (Escolar, 2002; Pereira, Portilla & Rodríguez, 2019).

También dispone de sub categorías apriorísticas, siendo una de ella la calidad de datos, considerada como una característica que no solo cumple normas y requerimientos sino también, las expectativas del cliente; por lo cual, es de gran valor garantizar la semejanza de los datos con la realidad, para acrecentar la confiabilidad en los mismos al momento de brindar conocimiento y en la toma de decisiones. Es por esto que, es necesario emplear herramientas para identificar irregularidades que afectan el contenido de los datos,

luego tratarlos y mejorar la calidad de los mismos; para así evitar que la efectividad y eficiencia de la organización se vean afectadas (Fernández , 2008; Valverde, Marotta & Vallespir, 2012; Caro, Fuentes & Soto, 2013; Osorio, Guerrero & González, 2017). Por otra parte, está el almacenamiento de datos que se puede visualizar desde dos enfoques: la estructura de los sistemas de almacenamiento y la administración de datos; además, con los avances tecnológicos y el aumento de información, los sistemas de almacenamiento deben poder guardar y administrar grandes volúmenes de datos, sin presentar inconvenientes, para que los datos puedan estar disponibles posteriormente, sin presentar alguna anomalía, y ser compartido entre diferentes usuarios en la empresa. Razón por la cual, la empresa y sus responsables deben validar la confiabilidad en los datos, y asegurar la eficiencia en el almacenamiento de datos, para que las tareas de búsqueda y adquisición de datos, para la solución de problemas, sea efectuada rápidamente (Vázquez, 2015; Instituto Nacional de Ciberseguridad, 2016). En último lugar, se tiene como sub categoría la recuperación de la información, con un rol de gran valor, puesto que ofrece a las empresas facilidades en obtener información, para ejecutar tareas y determinar la eficiencia o el fracaso de las operaciones. Por lo cual, se debe organizar la información almacenada, facilitando la localización de datos relevantes con características específicas, en un almacén de información o base de datos, capaces de brindar solución a un problema (Martínez, 2004; Salvador & Arquero, 2006; Bordignon & Tolosa, 2007; Vázquez, De León, Velasco, Villa & Briceño, 2017).

Por otro lado, también dispone de sub categorías emergentes, siendo una de ellas el perfilamiento de datos, considerado como el primer paso para realizar la mejora en los datos. Se encarga de cumplir dos funciones primordiales; en primer lugar, obtener resultados estadísticos de cada columna, por ejemplo la cantidad de valores nulos, los valores distintos dentro de la columna, tipo de dato o patrones más frecuentes de los datos en la columna, para luego analizar y evaluar la calidad de los datos. En segundo lugar, se encarga de examinar los valores de datos que pueden presentar dificultades para ser calculados, porque involucran múltiples columnas, presentan inclusión, dependencias funcionales y de orden. Ambas funciones, contribuyen para validar la presencia de errores y anomalías en los datos. Además, el perfilamiento de datos es considerada como una actividad importante para el profesional de IT (DAMA International, 2009; Naumann, 2014; Abedjan, Golab, & Naumann, 2017). Por otra parte, está la limpieza de datos (Data

cleaning o scrubbing), y se puede considerar que su objetivo es mejorar la calidad de los datos. Razón por la cual, se busca identificar errores o inconsistencias, por ejemplo campos vacíos, datos inválidos y/o datos duplicados que puedan estar presentes en archivos o base de datos, para minimizar el impacto negativo que pueda afectar a la información o para luego ser corregidos y brindar datos precisos y consistentes. Además, para ser aplicada con éxito, se debe cumplir su proceso, que incluye 3 etapas, identificar el error, realizar un diagnóstico y corregir anomalías. Por otro lado, la ausencia de una correcta limpieza de datos, puede afectar a otros datos o los resultados de un análisis que emplearon datos que contienen errores o anomalías (Rahm & Hai Do, 2000; Van den Broeck, Argeseanu, Eeckels, & Herbst, 2005; Osborne, 2013).

La investigación presenta la justificación basada en teorías, dado que aportan sustento; empezando por la teoría de la información, que respalda la importancia de la información para la organización, y el valor que le corresponde dentro de la misma; puesto que, emplearla de forma correcta, puede proveer de valioso conocimiento del rendimiento de la organización. Así mismo, la teoría de la decisión expone que, en base a la información obtenida y su correcta aplicación, se pueden visualizar opciones no perceptibles hasta el momento y tomar decisiones que pueden favorecer a la organización y a sus colaboradores.

Por otra parte, la justificación práctica en la investigación, se basa en la inspección de la data del servicio que ofrece la empresa, para poder identificar anomalías que puede presentar. También, permitirá la localización del origen de los problemas de calidad en entrada de datos o interfaces incorrectas. Se debe agregar que, con el aporte de la investigación, se espera que las anomalías identificadas en los datos sean corregidas, y cuando sean procesados, proporcionen información segura para un óptimo análisis del rendimiento de la organización y del servicio que ofrecen, o para la implementación de mejoras.

Finalmente, para la justificación metodológica se decidió utilizar la investigación holística porque el análisis de los datos de las tablas principales, obtenidas del modelamiento de datos del servicio de atención al cliente que ofrece la empresa, y las anomalías que presentan; corresponde a un enfoque mixto. Además, las soluciones a

proponer se basarán en las reglas del negocio y en propuestas brindadas por el encargado de la revisión de los datos. Llegados a este punto, se busca solucionar el problema de confiabilidad de los datos en su servicio de atención al punto de venta, dando a conocer las irregularidades en los datos para que sean corregidos y luego utilizados con mayor confianza por la empresa y sus colaboradores.

El problema general formulado en la investigación es ¿Cómo mejorar la gestión de datos de la plataforma web de una empresa de ventas?; así mismo, se determinó que los problemas específicos son: A) ¿Cómo validar la presencia de anomalías en los datos en la plataforma web de una empresa de ventas? B) ¿Cuál es la efectividad del almacenamiento de datos en la plataforma web de una empresa de ventas? C) ¿Cuál es la efectividad de la recuperación de datos en la plataforma web de una empresa de ventas? D) ¿Qué factores causan las incidencias que perjudican la eficiencia en la plataforma web de una empresa de ventas?

Dicho lo anterior, se determinó como objetivo general proponer la optimización de la gestión de datos de la plataforma web de una empresa de ventas, Lima 2020; así mismo, los objetivos específicos son: A) Analizar la calidad de los datos en la plataforma web de una empresa de ventas. B) Diagnosticar la efectividad del almacenamiento de datos en la plataforma web de una empresa de ventas. C) Diagnosticar la efectividad de la recuperación de datos en la plataforma web de una empresa de ventas. D) Identificar si la calidad de los datos, el almacenamiento o la recuperación de los datos perjudica la eficiencia en la plataforma web de una empresa de ventas.

## **II. MÉTODO**

### **2.1. Enfoque y tipo**

#### **2.1.1. Enfoque Mixto**

Proporciona una visualización más amplia de la investigación, logrando la unificación e implementación de los puntos positivos de los enfoques cualitativos y cuantitativos, capaces de proporcionar información más completa a la investigación. De manera semejante, la aplicación de una entrevista a los colaboradores y el análisis a los datos obtenidos de la empresa, puede otorgar mayor información, capaz de brindar una comprensión del problema de estudio que viene afectando a la empresa y a sus colaboradores (Hernández, Fernández, & Baptista, 2014).

#### **2.1.2. Sintagma Holístico**

Consiste en orientar al investigador en la aplicación de metodologías más completas y efectivas, ampliando su percepción de las situaciones y manejando distintos procesos orientados al crecimiento de la empresa (Hurtado, 2000). De manera que, puede proporcionar un enfoque global del problema que presenta, la gestión de datos de la plataforma web de una empresa de ventas, determinando los procesos o personas que se encuentran afectados, para luego proponer soluciones capaces de optimizar los datos, la información que proporcionan, los procesos en los que son empleados y la confianza por parte del personal asignado a su aplicación.

#### **2.1.3. Tipo proyectiva**

Planea escenarios futuros, para identificar oportunidades y definir acciones; además, es fundamental realizar la indagación previa (Hurtado, 2000). De manera que, con la aplicación de la investigación proyectiva, se pueden otorgar propuestas de solución, para optimizar la confiabilidad en la gestión de datos en la empresa, con previa búsqueda de fundamentos para la propuesta.

#### **2.1.4. Método Deductivo e Inductivo**

Aplicados para obtener mayor conocimiento, por medio de diferentes procedimientos, durante el desarrollo de la investigación. La aplicación de la metodología inductiva y

deductiva; es empleada al proporcionar nueva información que será obtenida del análisis de los datos; y la búsqueda de situaciones o hechos reales similares al problema por el que pasa la empresa, para aplicar sus teorías, principio, leyes, etc, para ser adaptados en soluciones o situaciones particulares (Bernal, 2010).

## **2.2. Población, muestra y unidades informantes**

### **2.2.1. Población**

La población es el conjunto o agrupación de elementos o individuos que presentan características semejantes a las cuales está enfocada la investigación (Bernal, 2010). Dicho lo anterior, se determina, que el conjunto o agrupación asignado en el trabajo, es la fuente requerida que posee las características necesarias para la investigación y para proporcionar una respuesta a la problemática principal en la investigación; por lo cual, se asignó como la población, a los colaboradores que pertenecen al área de Tecnología y Transformación Digital, dentro del área de sistemas, de la empresa de ventas; además; ellos serán sujetos para la entrevista, y aportarán los datos para obtener información válida de una fuente confiable.

### **2.2.2. Muestra**

La muestra es la agrupación de personas u objetos que forman parte de la población, siendo una cantidad determinada, que brinda información a la investigación; efectuando la medición y observación en base a los objetivos planteados (Bernal, 2010). Por consiguiente, en el presente trabajo de investigación, se seleccionó como muestra a tres colaboradores confirmados de la empresa, encargados de la base de datos, el diseño de la plataforma web y cualquier cambio necesario a la misma. De igual modo, se ha tomado como muestra la información almacenada, ingresada por la plataforma web y manejada por el área de sistemas.

### **2.2.3. Unidades informantes**

Es la defunción práctica y sencilla de las características que tiene en común la muestra de estudio, para poder brindar una respuesta más concreta y minuciosa al problema planteado en la investigación (Hurtado, 2000). Por lo cual, en el presente trabajo de investigación, se emplea a 3 colaboradores que son el Analista Programador, el Analista de Operaciones y un Programador, como unidades informantes y obtener conocimiento relacionado a la

empresa, el proyecto en el que se encuentran y los datos que gestionan. Por otro lado, también contaremos con dos objetos como una de las unidades informantes, que son los datos ingresados durante la atención a los puntos de venta, y un documento en Excel, con los registros de atención a los puntos de venta, obtenido de la plataforma web.

### 2.3. Categorías, sub categorías apriorísticas y emergentes

Tabla 1

*Matriz de categorización de gestión de datos*

Categorización	Sub Categoría Apriorística	Indicadores
Gestión de Datos	Calidad de Datos	Consistencia de datos
		Integridad de datos
		Conformidad con datos
	Almacenamiento de datos	Capacidad de Almacenamiento
	Recuperación de Información	Accesibilidad
		Disponibilidad

Tabla 2

*Matriz de Sub Categorías emergentes*

Categorización	Sub Categoría Emergentes
Gestión de Datos	Perfilamiento
	Limpieza

### 2.4. Técnicas e instrumentos de recolección de datos

#### 2.4.1. Técnicas

La entrevista es una técnica empleada para recolectar información relevante, por medio de la conversación, a base de preguntas y respuestas (Niño, 2011). Además, se puede obtener información adicional por medio de puntos de vista, opiniones o ideas de los participantes respecto a los problemas que presentan, para exponer una conclusión más completa para la investigación. Razón por la cual, se aplicará la entrevista a los trabajadores responsables de la gestión de datos y el uso de los mismos, para cumplir con los objetivos planteados en la investigación.

#### **2.4.2. Instrumentos**

La guía de entrevista es un documento que contiene temas y preguntas a analizar, relacionadas al objetivo planteado en la investigación; siendo una guía para no salir del contexto, y direccionar la conversación, evitando que el sujeto a encuestar, se aburra o incomode, y pueda brindar mayor información (Acevedo & López, 2004). Además, también puede contener notas con puntos importantes a tratar. De esta forma, la entrevista se desarrollará de forma fluida y armoniosa; con la libertad de poder variar o aumentar preguntas, para lograr recolectar mayor cantidad de información para la investigación.

#### **2.5. Proceso de recolección de datos**

Se detalla los procedimientos efectuados para la investigación:

Solicitar autorización para emplear datos de la empresa.

Elaboración del registro documental, guía de entrevista y ficha de entrevista.

Revisión y aprobación de la ficha de entrevista.

Aplicación de las entrevistas virtuales a los colaboradores.

Análisis de los datos ingresados durante la atención a los puntos de venta, para obtener los resultados cuantitativos.

Análisis de la información obtenida por la entrevista, para obtener los resultados cualitativos.

Triangulación de los resultados cuantitativos y cualitativos para el análisis mixto, y obtener un diagnóstico.

Elaboración de la propuesta.

#### **2.6. Método de análisis de datos**

El análisis de datos, se realizó empleando la información obtenida de las entrevistas y el registro documental. Se utilizó la herramienta Excel para analizar la información que brindó la empresa de ventas, logrando obtener tablas y gráficos que permiten visualizar los resultados cuantitativos, correspondiente a cada sub categoría; también se empleó la herramienta ATLAS.ti 8 para analizar las respuestas obtenidas de las entrevistas a los colaboradores de la empresa de ventas, logrando obtener resultados cualitativos, con una interpretación eficaz de la información mediante el uso de indicadores y redes. La misma herramienta fue utilizada en la triangulación de los resultados cualitativos y cuantitativos, para obtener un análisis más certero y una apreciación más eficaz de la información, mediante el uso de indicadores y redes.

### III. RESULTADOS

#### 3.1. Descripción de resultados cuantitativos

A continuación, presentaremos resultados cuantitativos del estudio. Lo primero que se analizará será el campo número de consulta, del registro que se obtiene luego de brindar la atención al punto de venta (pdv) por medio de su plataforma web. Presentamos la tabla 3, donde se observa que el campo número de consulta presenta patrones de 8, 9 y 10 dígitos numéricos en sus datos para los meses de enero, febrero y marzo del año 2019. Se contabiliza 30 000 registros de atención al pdv en cada mes, visualizando la frecuencia (f) y el porcentaje (%) de los patrones de los datos obtenidos del campo número de consulta. Siendo la frecuencia, el total de registros con datos que tienen los patrones indicados. Este campo permite revisar las incidencias registradas que presentan los puntos de venta durante la atención al cliente.

Tabla 3

*Frecuencia (f) y porcentaje (%) de los patrones del campo número de consulta de enero, febrero y marzo del año 2019*

Patrones	Enero		Febrero		Marzo	
	f	%	f	%	f	%
8 dígitos numéricos	2879	9	3178	11	3403	11
9 dígitos numéricos	26314	88	26822	89	26597	89
10 dígitos numéricos	807	3	0	0	0	0
Total	30000	100	30000	100	30000	100

*Fuente:* Elaboración propia (2020)

En la figura 1 se visualiza los porcentajes (%) de los patrones de datos ingresados en el campo número de consulta correspondiente a la tabla 3. Este campo se debe completar ingresando 9 dígitos numéricos; sin embargo, inicialmente no se dio a los usuarios las indicaciones para completar el campo, razón por la cual, se observa datos con diferentes patrones en el primer trimestre del 2019. Se evidencia que, en enero del 2019, de 30 000 registros, el 88% de los datos fueron ingresados correctamente con 9 dígitos numéricos, y el 12% de los datos fueron ingresados de forma incorrecta con 8 y 10 dígitos

numéricos. En Febrero y Marzo del 2019, de 30 000 registros en cada mes, el 89% de los datos fueron ingresados correctamente con 9 dígitos numéricos, y el 11% de los datos fueron ingresados de forma incorrecta con 8 dígitos numéricos.

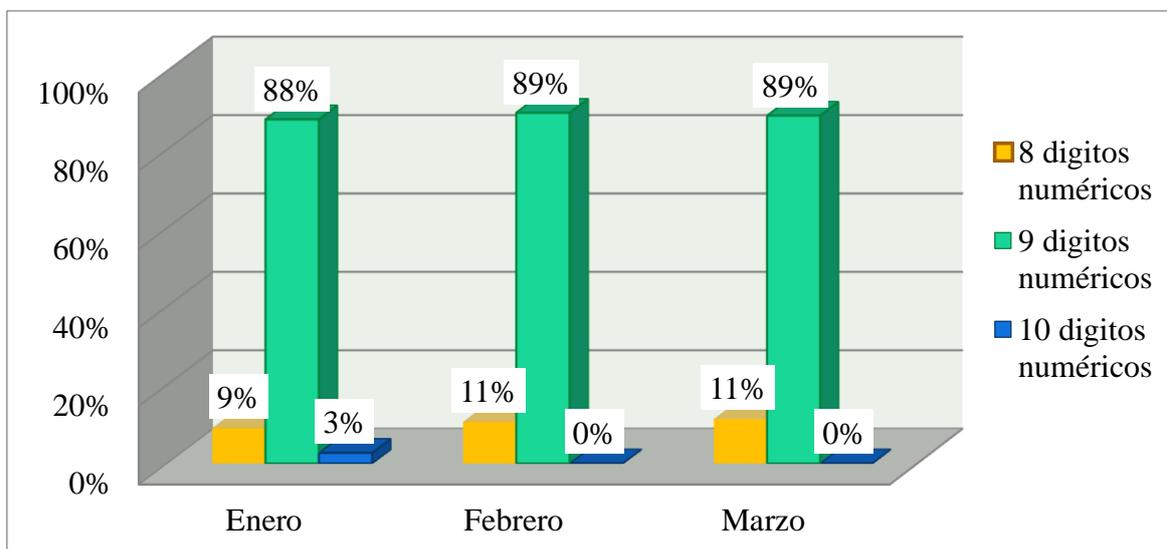


Figura 1. Porcentaje (%) de los patrones del campo número de consulta de enero, febrero y marzo del año 2019

Fuente: Elaboración propia (2020)

Luego se analizará el campo número de orden, el cual no es obligatorio que el usuario ingrese datos en registro al brindar la atención al pdv por medio de su plataforma web; sin embargo, una vez registro se considera un dato importante. Presentamos la tabla 4, donde se observa que el campo número de orden presenta diferentes patrones en sus datos en los meses de enero, febrero y marzo del año 2019. En primer lugar, como no es un campo obligatorio puede contener campos vacíos que va a estar representado por la palabra Vacío; en segundo lugar, presenta un patrón de 10 dígitos numéricos representado con 9(10); en tercer lugar, presenta un patrón de 10 dígitos alfanuméricos con letra mayúscula representado con 9(9)X y por último presenta un patrón de 10 dígitos alfanuméricos con letra minúscula representado con 9(9)x. Se contabiliza 30 000 registros de atención al pdv en cada mes, visualizando la frecuencia (f) y el porcentaje (%) de los patrones de los datos obtenidos.

En la figura 2 se visualizan los porcentajes (%) de los patrones de los datos ingresados en el campo número de orden, correspondiente a la tabla 4. Este campo es

completado de dos formas, como no es obligatorio, puede estar vacío, pero en caso se ingresa datos en este campo, debe contener 10 dígitos alfanuméricos con letra mayúscula; sin embargo, inicialmente no se dio a los usuarios las indicaciones para completar el campo; razón por la cual, se observa datos con diferentes patrones en el primer trimestre del 2019. Se evidencia que en Enero del 2019, de 30 000 registros, el 68% de los registros tienen este campo vacío, siendo aceptable porque no es un campo obligatorio; el 24% de los datos fueron ingresados de forma correcta con 10 dígitos alfanuméricos que emplea letra mayúscula; el 8% de los datos fueron ingresados de forma incorrecta con 10 dígitos numéricos y 10 dígitos alfanuméricos que emplea letra minúscula. En Febrero del 2019, de 30 000 registros, el 64% de los registros tienen este campo vacío, siendo aceptable porque no es un campo obligatorio; el 30% de los datos fueron ingresados de forma correcta con 10 dígitos alfanuméricos que emplea letra mayúscula; el 6% de los datos fueron ingresados de forma incorrecta con 10 dígitos numéricos y 10 dígitos alfanuméricos que emplea letra minúscula. En Marzo del 2019, de 30 000 registros, el 62% de los registros tienen este campo vacío, siendo aceptable porque no es un campo obligatorio; el 33% de los datos fueron ingresados de forma correcta con 10 dígitos alfanuméricos que emplea letra mayúscula; el 5% de los datos fueron ingresados de forma incorrecta con 10 dígitos numéricos y 10 dígitos alfanuméricos que emplea letra minúscula.

Tabla 4

*Frecuencia (f) y porcentaje (%) de los patrones en el campo número de orden enero, febrero y marzo del año 2019*

Patrones	Enero		Febrero		Marzo	
	f	%	f	%	f	%
Vacío	20520	68	19156	64	18712	62
9(10)	490	2	446	1	360	1
9(9)X	7182	24	8971	30	9855	33
9(9)x	1808	6	1427	5	1073	4
Total	30000	100	30000	100	30000	100

*Fuente:* Elaboración propia (2020)

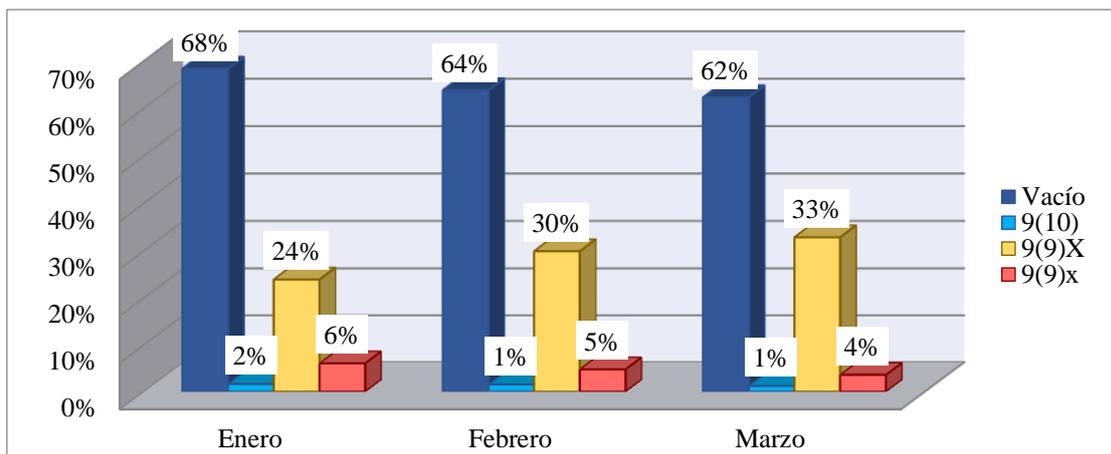


Figura 2. Porcentaje (%) de los patrones en el campo número de orden enero, febrero y marzo del año 2019

Fuente: Elaboración propia (2020)

A continuación, se analizará el campo Documento Nacional de Identidad (DNI) de los usuarios pdv. Los usuarios pdv son vendedores, que emplean la plataforma web para solicitar ayuda a los usuarios agente o asesor en caso presenten alguna irregularidad durante la venta, y sus datos son registrados en dos bases de datos diferentes; que son la base de datos principal de la empresa, representada como Base de Datos A, y en la base de datos que pertenece a la plataforma web, representada como Base de Datos B. En la Base de Datos A, algunos datos son almacenados con 7 dígitos numéricos, a pesar que este campo debe contener datos con 8 dígitos numéricos. Presentamos la tabla 5, donde se puede observar la Base de Datos A y la Base de Datos B y su frecuencia (f) y porcentaje (%) de los patrones de 7 y 8 dígitos numéricos de los datos del campo DNI de los usuarios pdv. Siendo la frecuencia (f), el total de datos que tienen los patrones indicados.

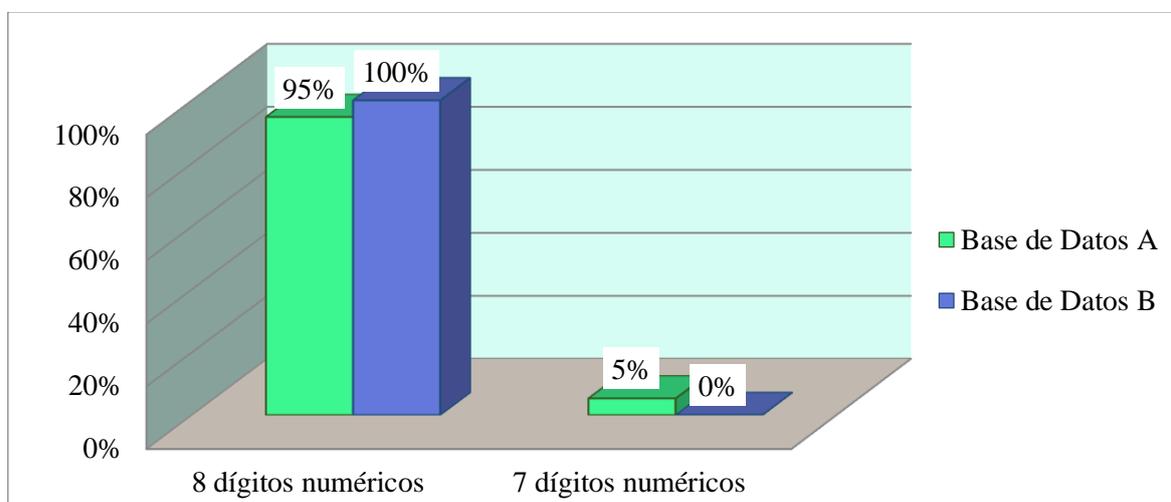
Tabla 5

Frecuencia (f) y porcentaje (%) de los patrones de DNI de usuario pdv

Patrones	Base de Datos A		Base de Datos B	
	f	%	f	%
7 dígitos numéricos	252	5	0	0
8 dígitos numéricos	4608	95	4860	100
Total	4860	100	4860	100

Fuente: Elaboración propia (2020)

En la figura 3 se visualiza los porcentajes (%) de los patrones del campo DNI de los usuarios pdv, correspondiente a la tabla 5. En este campo se registra el número de identidad de cada usuario y está asignado como llave primaria (clave principal) en la tabla donde se almacena los datos del usuario. Por esta razón, se deben ingresar 8 dígitos numéricos correctamente digitados y completos. Sin embargo, en la base de datos de la plataforma web, representada como Base de Datos A, este campo se encuentra habilitado para almacenar números enteros, dificultando el almacenamiento de los datos que tiene un 0 (cero) como primer dígito, porque no los considera y registra solo 7 dígitos numéricos. Razón por la cual, se observa datos con patrones de 7 y 8 dígitos numéricos en usuarios que se encontraban laborando durante el primer trimestre del 2019. Dicho lo anterior, se evidencia que en la Base de Datos A, de 4860 usuarios pdv, el 95% de los DNI's fueron registrados de forma correcta con 8 dígitos numéricos; sin embargo, el 5% de los DNI's fueron registrados de forma incorrecta con 7 dígitos numéricos. Por otro lado, en la Base de Datos B, de 4869 usuarios pdv, el 100% de los DNI's fueron registrados de forma correcta con 8 dígitos numéricos. Lo que evidencia, una inconsistencia en dos bases de datos que se relacionan y comparten información.



*Figura 3.* Porcentaje (%) de los patrones de DNI de usuario pdv

*Fuente:* Elaboración propia (2020)

A continuación, se analizará el total de registros de atención a los pdv, que ingresaron durante los primeros 6 meses del año 2019. Presentamos la Tabla 6, donde se puede observar, durante los meses de Enero a Junio, los Registros de Atención realizados mensualmente y su frecuencia (f) y porcentaje (%), siendo la frecuencia (f) el total de

Registros de Atención contabilizados en el mes; el Espacio de Almacenamiento que ocupa en *gigabyte* (GB); y la Media del total de registros durante los primeros 6 meses del año 2019 con su porcentaje (%).

Tabla 6

*Frecuencia (f) y porcentaje (%) de los registros de atención ingresados de enero a junio del año 2019*

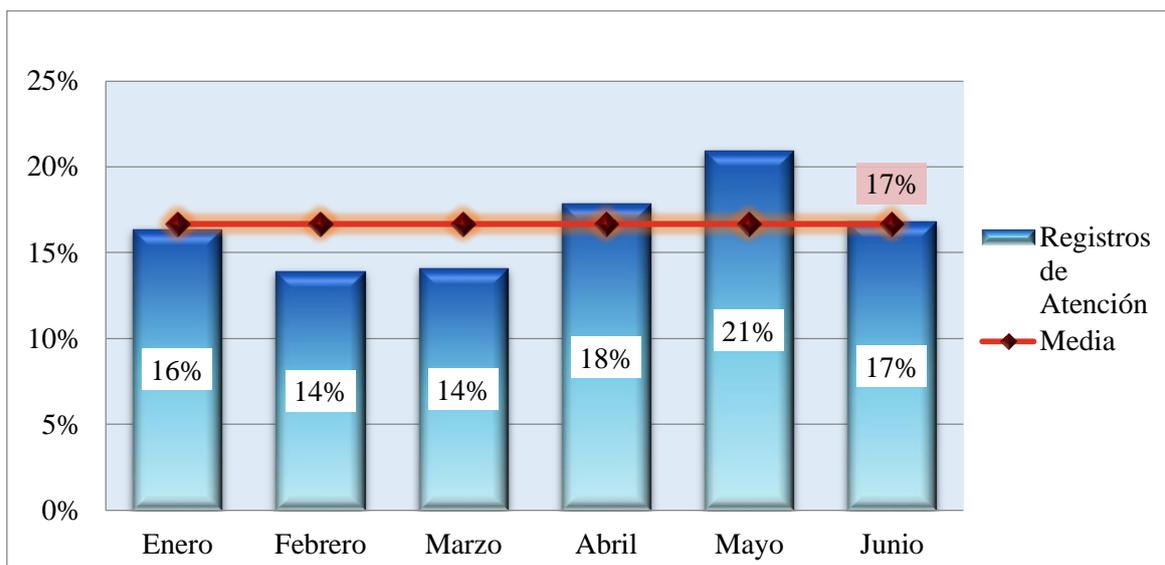
Mes	Registros de Atención		Espacio de Almacenamiento	Media	
	f	%	GB	Promedio	%
Enero	131338	16	3	133608	17
Febrero	111546	14	3	133608	17
Marzo	112998	14	3	133608	17
Abril	143093	18	4	133608	17
Mayo	167971	21	4	133608	17
Junio	134703	17	3	133608	17
Total	801649	100	20		

*Fuente:* Elaboración propia (2020)

En la tabla 6 nos muestra que del total Registros de Atención almacenados durante los 6 primeros meses del año 2019, que equivalen a 20 GB de Espacio de Almacenamiento, el 16% de los Registros de Atención almacenados durante el mes de Enero, equivalen a 3GB de memoria en Espacio de Almacenamiento; el 14% de los registros almacenado durante el mes de Febrero y también durante el mes de Marzo, equivalen a 3GB de memoria en Espacio de Almacenamiento en cada mes; el 18% de los registros almacenados durante el mes de Abril, equivalen a 3GB de memoria en Espacio de Almacenamiento; el 21% de los Registros de Atención almacenados durante el mes de Mayo, equivalen a 4GB de memoria en Espacio de Almacenamiento; y el 17% de los registros almacenados durante el mes de Junio, equivalen a 3GB de memoria en Espacio de Almacenamiento. Tomando las medidas necesarias para salvaguardar la información.

A continuación, en la figura 4 se muestra los porcentajes (%) del total de Registros de Atención contabilizados, correspondiente a la tabla 6, que nos indica el total de usuarios pdv atendidos durante el mes y a los que se les brindó la ayuda solicitada. Se puede visualizar que se tiene una cantidad de registros aceptable en cada mes, teniendo una media del 17%, que corresponde a los primeros 6 meses del año 2019. Además, se puede observar

que el total de registros de cada mes oscila la media en  $\pm 3\%$ , con excepción en el mes de mayo.



*Figura 4.* Porcentaje (%) de los registros de atención ingresados de enero a junio del año 2019

*Fuente:* Elaboración propia (2020)

A continuación, se analizará el acceso a la plataforma web, y a la información que se almacena; centrándonos en los movimientos para ingresar de forma exitosa, y a los errores que comete el usuario pdv. Presentamos la tabla 7 donde se puede observar, durante los meses de Enero a Junio, los movimientos para acceder a la plataforma web. Se tiene 5 movimientos; el primero es el ingreso exitoso por parte del usuario pdv, que es representado con la palabra INGRESAR (*LOGIN*); por otro lado, los siguientes 4 movimientos que son los errores que comete el usuario al momento de ingresar a la Plataforma Web, son representados con FALLA POR CATCHA, FALLA CONTRASEÑA, FALLA POR BLOQUEO, USUARIO INCORRECTO. Además, la suma de todos los movimientos, que son los ingresos exitosos y con error por parte del usuario pdv, dan el total de ingresos registrados realizados por los usuarios pdv, representado con Total, porque a pesar de que ellos cometan algún error al momento de acceder a la plataforma web, al final, logran acceder a la misma para poder realizar sus consultas. También, se visualiza la frecuencia (*f*) y el porcentaje (%) de los movimientos indicados para acceder a la plataforma web y el total.

Tabla 7

*Frecuencia (f) y porcentaje (%) del acceso a la plataforma web de enero a junio del año 2019*

MES	INGRESAR (LOGIN)		FALLA POR CATCHA		FALLA CONTRASEÑA		FALLA POR BLOQUEO		USUARIO INCORRECTO		TOTAL	
	f	%	f	%	f	%	f	%	f	%	f	%
Enero	56673	53	25061	24	19487	19	2723	3	534	1	104478	100
Febrero	50749	53	23370	24	19165	20	2815	3	412	0	96511	100
Marzo	52736	54	23341	24	18574	19	2643	3	290	0	97584	100
Abril	56709	52	25848	24	21739	20	3091	3	1459	1	108846	100
Mayo	61753	52	27276	23	22968	19	2982	2	4503	4	119482	100
Junio	56595	54	22194	21	20246	19	2492	2	3653	3	105180	100

*Fuente:* Elaboración propia (2020)

En la figura 5 se puede visualizar los porcentajes (%), correspondientes a la tabla 7, de los movimientos realizados por los usuarios pdv al ingresar su usuario, contraseña y el captcha, como medida de seguridad. Se evidencia que en Enero del 2019, de 104478 registros de usuarios pdv que ingresaron a la plataforma web, el 53% ingresó de forma exitosa; y el 47% de los registros indican que el usuario pdv cometió algún error al intentar ingresar a la plataforma web; en Febrero del 2019, de 96511 registros de usuarios pdv que ingresaron a la plataforma web, el 53% fue exitoso; y el 47% de los registros indican que el usuario pdv cometió algún error; en Marzo del 2019, de 97584 registros de ingreso a la plataforma web, el 54% fue exitoso; y el 46% de los registros indican que el usuario pdv cometió algún error al intentar ingresar a la plataforma web; en Abril del 2019, de 108846 registros de ingreso a la plataforma web, el 52% fue exitoso; y el 48% de los registros indican que el usuario pdv cometió algún error al intentar ingresar a la plataforma web; en Mayo del 2019, de 119482 registros de ingreso a la plataforma web, el 52% fue exitoso; y el 48% de los registros indican que el usuario pdv cometió algún error; en Junio del 2019, de 105180 registros de ingreso a la plataforma web, el 55% fue exitoso; y el 45% de los registros indican que el usuario pdv cometió algún error al intentar ingresar a la plataforma web. Se debe considerar como error del usuario el ingreso de contraseña incorrecta, captcha incorrecto, usuario incorrecto y bloqueo de usuario por ingresar mal su usuario o contraseña incorrectos.

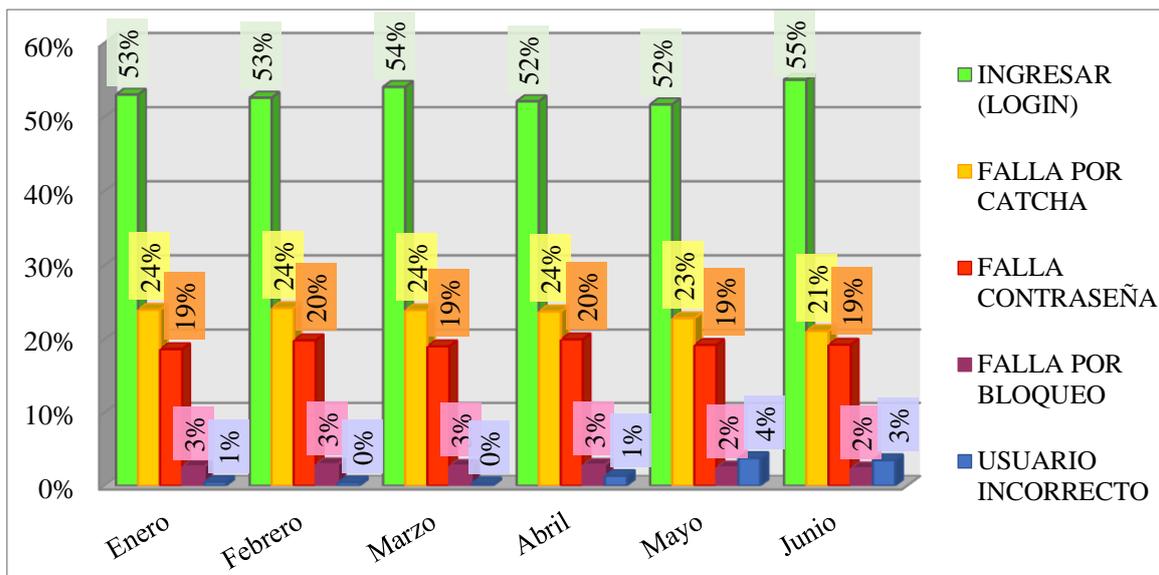


Figura 5. Porcentaje (%) del acceso a la plataforma web de enero a junio del año 2019

Fuente: Elaboración propia (2020)

A continuación se analizará el total de Consultas realizadas a los datos ingresados a la plataforma web, por parte de los usuarios pdv. Presentamos la Tabla 8, donde se puede observar, la frecuencia (f) y porcentaje (%) de las Consultas realizadas por los usuarios pdv, siendo la frecuencia (f) el total de Consultas realizadas en el mes, durante los meses de Enero a Junio. También se observa la Media, que se obtiene por el promedio de las Consultas realizadas durante los primeros 6 meses del año 2019 y su porcentaje (%) correspondiente.

Tabla 8

Frecuencia (f) y porcentaje (%) de consultas de usuarios pdv de enero a junio del año 2019.

Mes	Consultas		Media	
	f	%	Promedio	%
Enero	4912	4	19920	17
Febrero	14802	12	19920	17
Marzo	16527	14	19920	17
Abril	24873	21	19920	17
Mayo	31596	26	19920	17
Junio	26809	22	19920	17
Total	119519			

Fuente: Elaboración propia (2020)

En la figura 6 se muestra los porcentajes (%) de los registros de las consultas realizadas por los usuarios pdv a la plataforma web, correspondiente a la Tabla 8, para asegurar la disponibilidad de la información para los usuarios pdv; logrando evidenciar que el total de consultas ejecutadas durante los primeros seis meses del 2019 tiene una media del 17%, obtenida del promedio de consultas realizadas durante los primeros 6 meses del año 2019; además, también se observa que del total de consultas ejecutadas, un 4% de las consultas fue realizado durante el mes de enero y se encuentra lejos de la media, siendo el primer mes con menor porcentaje de consultas; por otro lado se observa un mejora del 12% en las consultas durante el mes de febrero, de igual forma se puede observar un mejora del 14% en las consultas realizadas durante el mes de marzo, observando que en los meses de febrero y marzo los porcentaje se acercan a la media; en el mes de abril también se visualiza una mejoría del 21% de consultas realizadas, en mayo se aprecia la mejoría en un 26% de las consultas realizadas, observando que en los meses de abril y mayo los porcentajes sobrepasaron la media; sin embargo durante el mes de junio se observa un 22% de consultas realizas, y a pesar de que su porcentaje sobrepasa la media, es menor al mes anterior. Lo que evidencia que hay una gran oscilación alrededor de la media, en aproximadamente un  $\pm 9\%$ . También se determinó, que a partir del mes de Febrero, los usuarios pdv empezaron a realizar cónsulas de forma más constantes porque la información que buscaban se encontraba disponible y era más detallada.

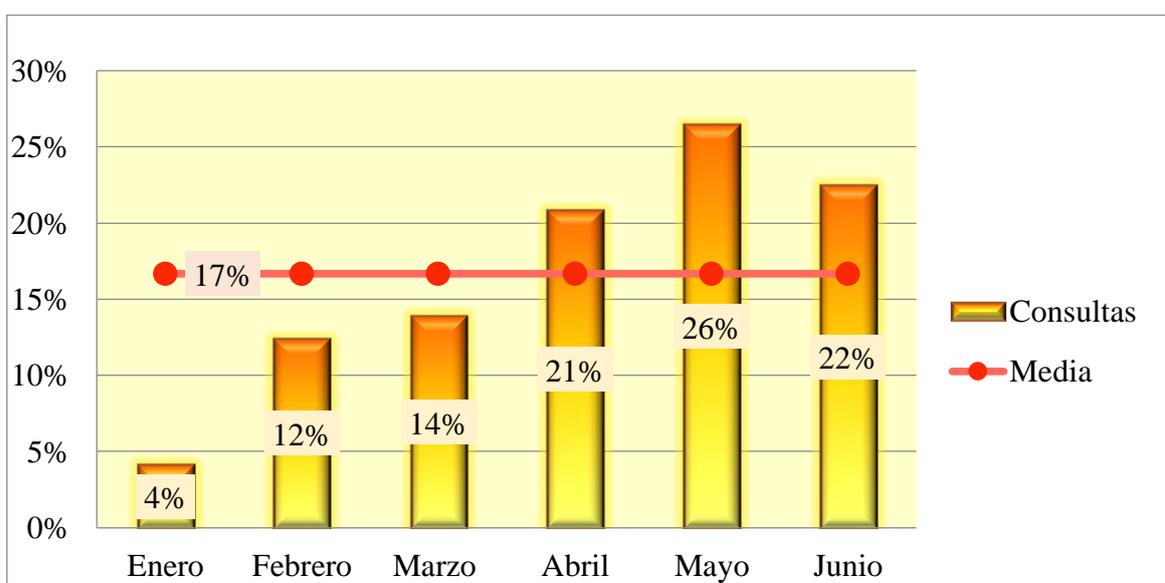


Figura 6. Porcentaje (%) de consultas de usuarios pdv de enero a junio del 2019

Fuente: Elaboración propia (2020)

### 3.2. Descripción de resultados cualitativos

A continuación, presentaremos los resultados cualitativos del estudio. Lo primero a analizar será la red correspondiente a la primera sub categoría Calidad de Datos, obtenida después de relacionar las opiniones de cada uno de los entrevistados con los indicadores propios de la primera sub categoría.

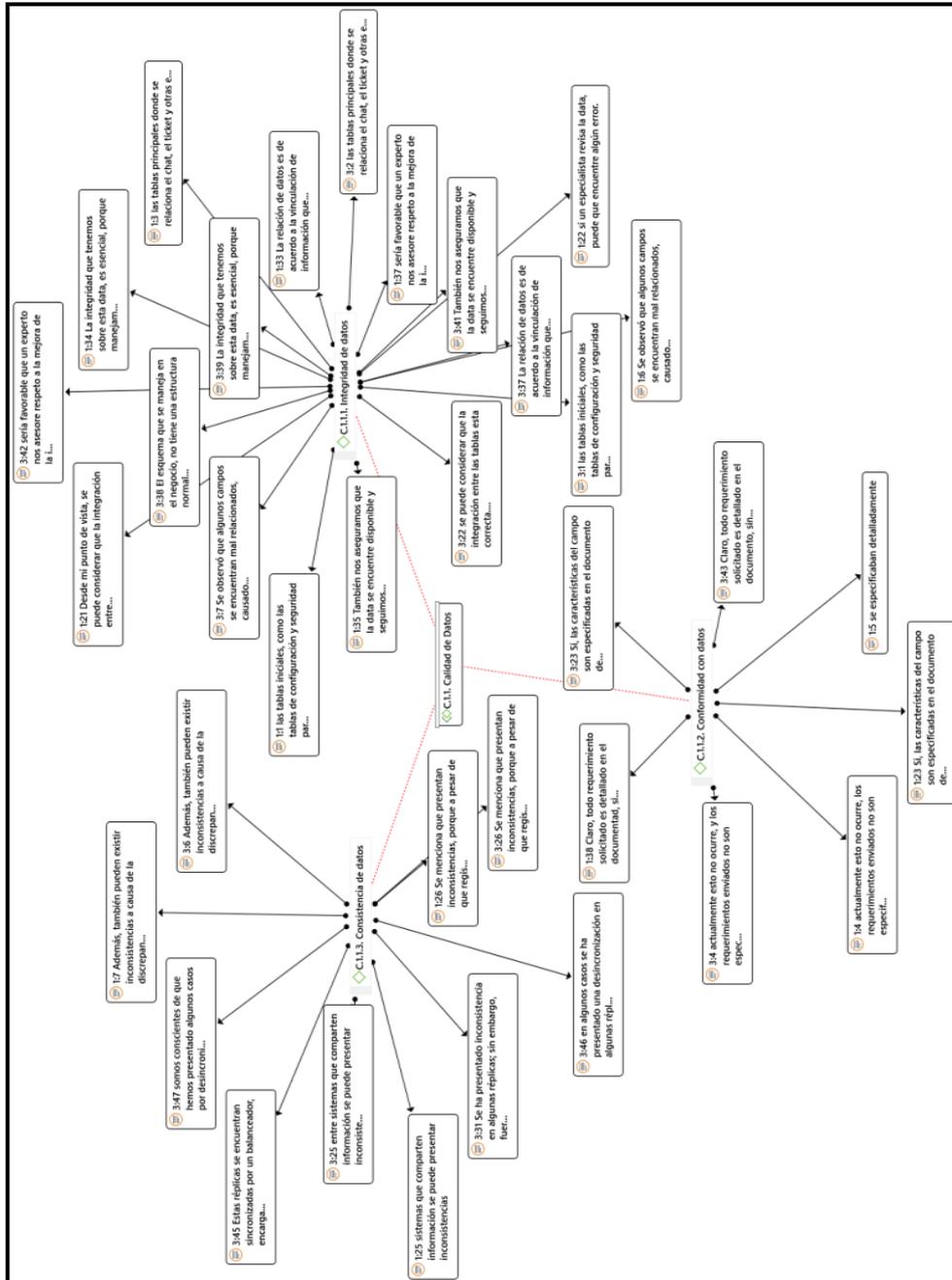


Figura 7. Red de sub categoría calidad de datos

Fuente: Elaboración propia (2020)

En la figura 7 se visualiza la Red de sub categoría Calidad de Datos obtenida por medio del análisis cualitativo, logrando identificar las deficiencias que presenta la base de datos de la plataforma web, que se emplea para brindar servicio de ayuda a los pdv. Se logró determinar que la base de datos presenta tablas, tales como usuario, usuario adicional y vendedor, en donde su llave primaria (clave principal), que es el campo que permite la relación entre tablas, de la propia base de datos u otras, almacena el dato incompleto y no permite una correcta integración con otras bases de datos empleadas en la plataforma web, provoca inconsistencias como por ejemplo dificulta para la búsqueda de información y/o ingreso a la plataforma web. Por otro lado, la descoordinación al momento de aplicar requerimientos, como la creación de nuevos campos, afecta el uso de los datos, porque las características del campo y su contenido pueden variar, y los patrones de los datos pueden diferir. Se determinó también, que tuvieron casos de desincronización entre sus réplicas, causando que los datos no se almacenen de la misma manera en todas las réplicas de la base de datos, sin embargo; por el nivel de importancia que tienen, los problemas fueron atendidos de forma inmediata. Logrando determinar, que su calidad de datos es buena, sin embargo, por parte de los mismos entrevistados, es recomendable la asesoría por parte de especialistas, para identificar otras inconsistencias y poder corregirlas.

A continuación, se analizará la red correspondiente a la segunda sub categoría Almacenamiento de Datos, obtenida después de relacionar las opiniones de cada uno de los tres entrevistados con los indicadores propios de la segunda sub categoría.

En la figura 8 se visualiza la Red de sub categoría Almacenamiento de Datos obtenida por medio del análisis cualitativo, logrando identificar que la empresa maneja servidores con el suficiente espacio para soportar el constante ingreso de datos, a la base de datos original y sus 8 réplicas, distribuidos en servidores físicos y virtuales. Sin embargo, con el uso contante de la plataforma web, es inevitable el incremento en el volumen de datos; razón por la cual, es necesario aumentar el espacio para el almacenamiento. Además, el manejo de estándares para salvaguardar la información les permite conservar la data histórica, en caso sea requerida en algún momento. Por esta razón, se está proyectando posibles soluciones para aumentar el espacio de almacenamiento. Por ejemplo, la compra de otro servidor o migrar su data a la nube, siendo la opción más factible para mitigar

costos. Por otro lado, cualquier problema que se presente en las bases de datos o los servidores, por ejemplo que el sistema se sature por consultas o desactualización en las réplicas, sino se atiende de forma inmediata, se verán reflejadas en la plataforma web, afectando a los usuarios y dificultando sus actividades.

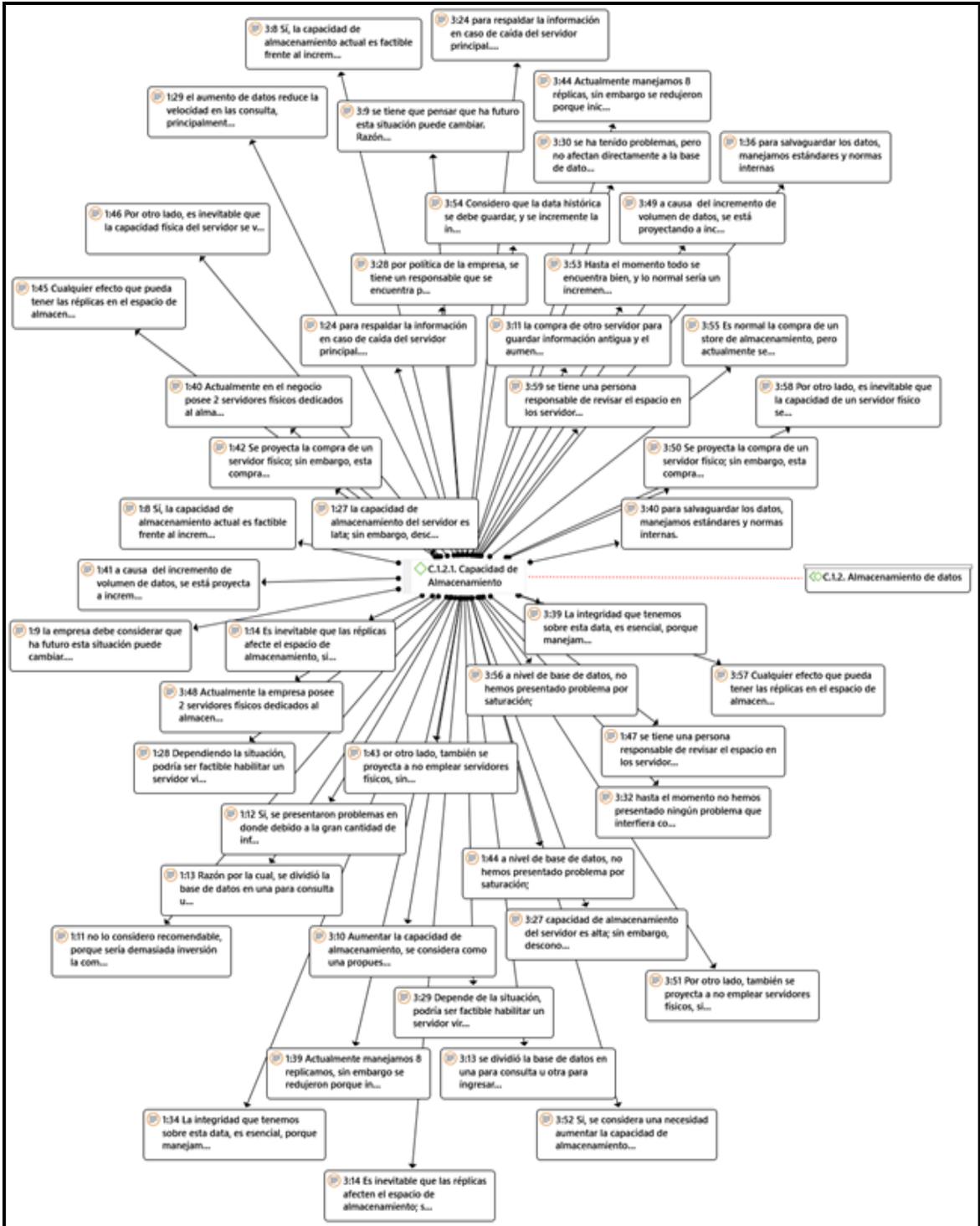


Figura 8. Red de sub categoría almacenamiento de datos

Fuente: Elaboración propia (2020)

Por último, se analizará la red correspondiente a la tercera sub categoría Recuperación de Información, obtenida después de relacionar las opiniones de cada uno de los tres entrevistados con los indicadores propios de la tercera sub categoría.

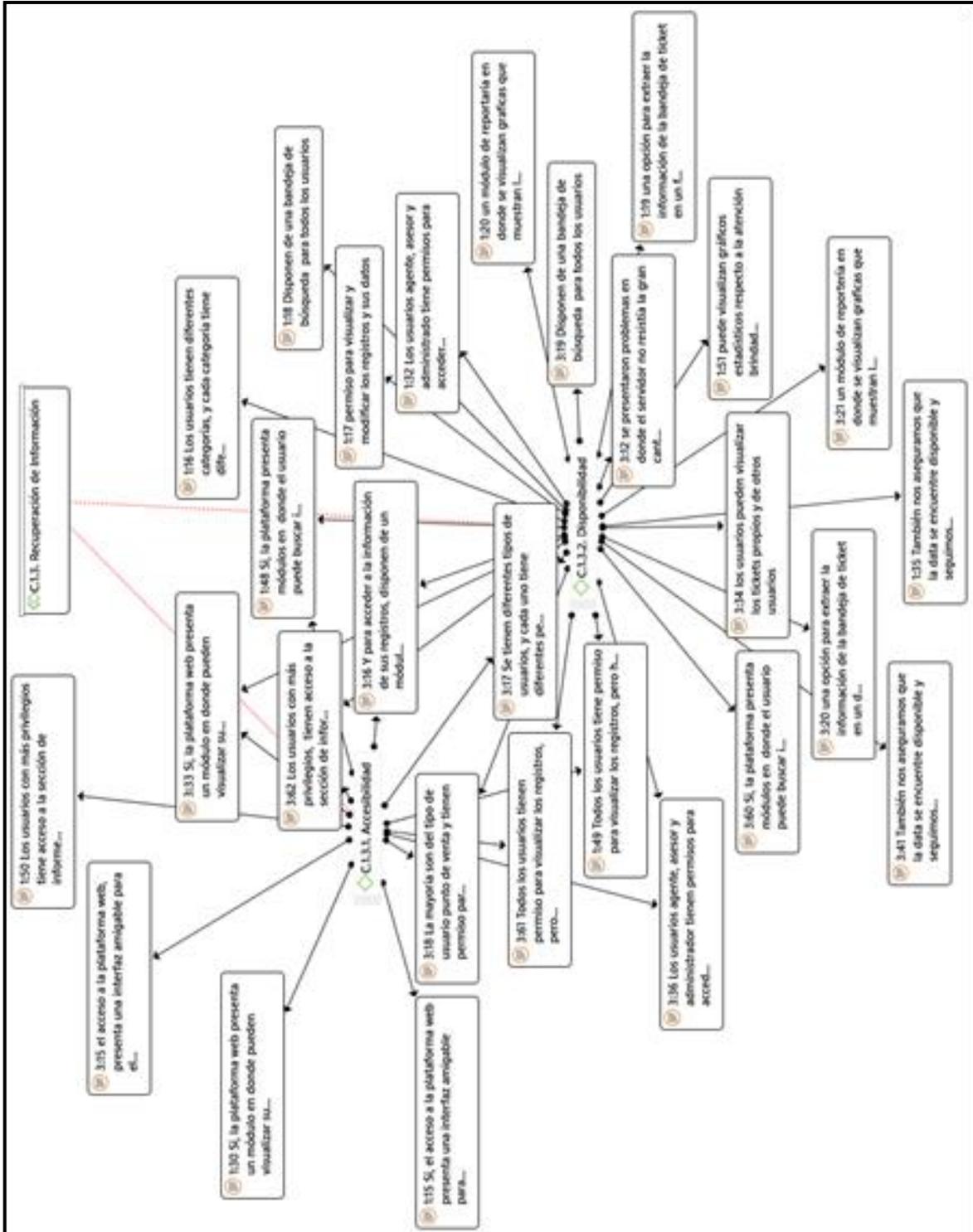


Figura 9. Red de sub categoría recuperación de información

Fuente: Elaboración propia (2020)

En la figura 9 se visualiza la Red de sub categoría Recuperación de Información, obtenida por medio del análisis cualitativo, logrando determinar que el acceso a la plataforma web, y a la información registrada, es amigable para el usuario; y también mantiene su seguridad para no poner en riesgo la información que contiene. Por otro lado, todos los usuarios disponen de una bandeja de búsqueda, para realizar consultas requeridas y poder visualizarla, pero no todos los usuarios tienen permiso para modificar los datos disponibles en la plataforma web. Los permisos son asignados según el tipo de usuario; en total se tiene 4 tipos de usuarios, los cuales son usuario punto de venta, usuario agente, usuario asesor y usuario administrador; asignados de acuerdo a las actividades de cada colaborador. Además, los usuarios con mayores permisos, como usuario asesor o administrador, disponen de un módulo para poder visualizar gráficos que determinan el nivel de atención por medio de indicadores; y tiene habilitada la opción para poder exportar, de la plataforma web, un documento Excel con información de los registros realizados, y analizar los datos para obtener conocimiento o brindar propuestas para la mejora del servicio.

### **3.3 Diagnóstico**

En cuanto a la triangulación del análisis cuantitativo y cualitativo de la sub categoría Calidad de Datos, que se observa en la figura 10, se interpreta que la gestión de datos de la plataforma web presenta problemas de calidad en los datos, tales como la ausencia de conformidad, causada por la descoordinación al momento de aplicar requerimientos, principalmente en la creación de nuevos campos, porque los datos no presentan un formato fijo al momento de ser ingresados y pueden presentar diferentes patrones. Además, la integración de datos también se ve afectada, porque los campos con diferentes patrones en sus datos pueden ser datos incompletos o incorrectos, y afectan la relación entre tablas o bases de datos. A pesar de ello, se observa que el porcentaje de datos con patrones incorrectos es menor al porcentaje de datos con patrones correctos; sin embargo, esto afecta al momento de emplear el campo como un indicador de medición, puesto que puede perjudicar el análisis. Por otro lado, también se encuentra afectada la consistencia e integración, en campos que pertenecen a bases de datos diferentes empleadas en la plataforma web, que se relacionan y comparten datos.

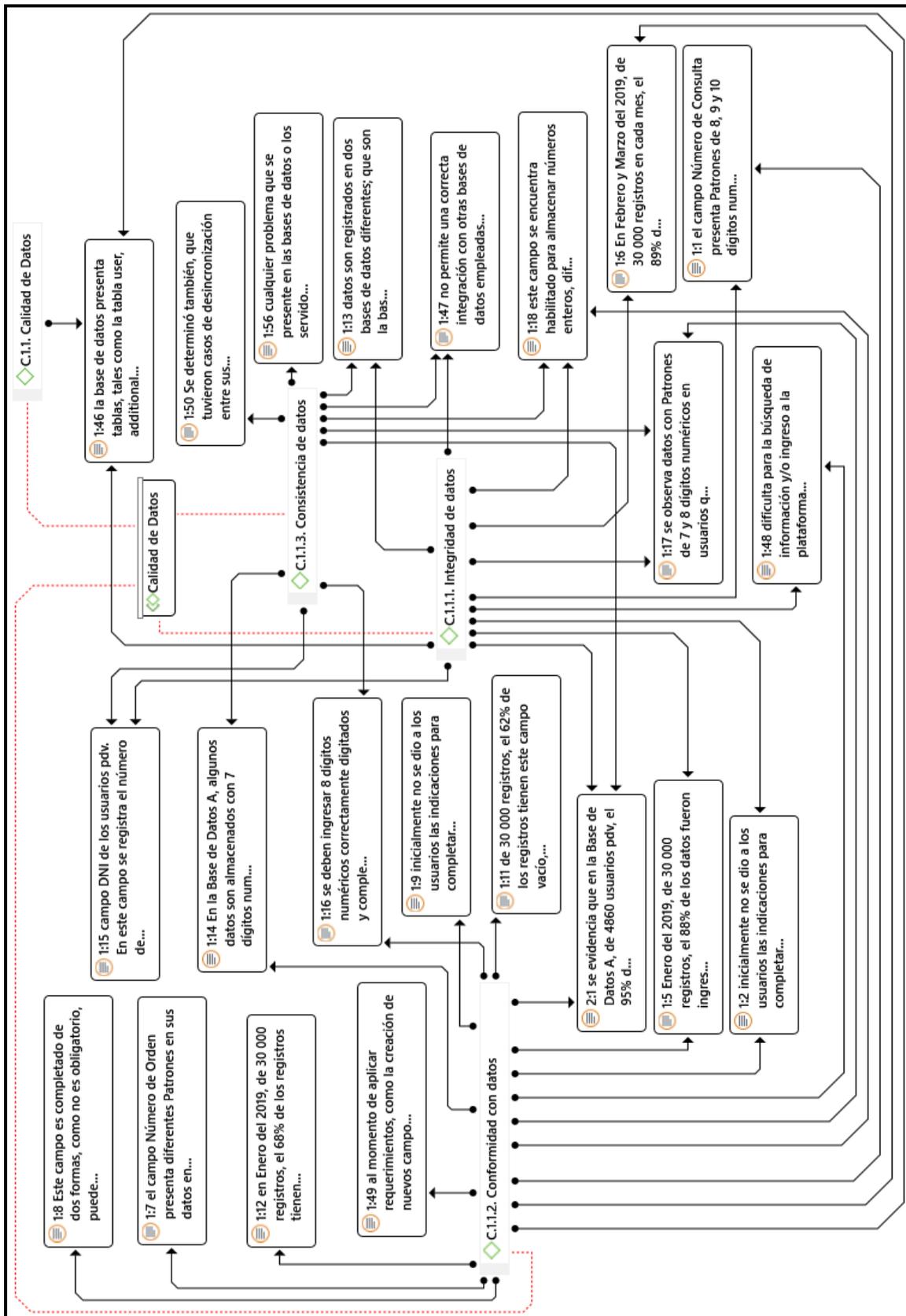


Figura 10. Análisis mixto de la sub categoría calidad de datos

Fuente: Elaboración propia (2020)

En cuanto a la triangulación del análisis cuantitativo y cualitativo de la sub categoría Almacenamiento de Datos, que se observa en la figura 11, se interpreta que la gestión de datos de la plataforma web, actualmente no presenta problemas en el almacenamiento de datos. Según el análisis, el espacio asignado para almacenar la cantidad de registros que ingresan mensualmente, al emplear la plataforma web, brinda el abasto necesario para los datos. Por otro lado, se determinó que es inevitable el incremento en el volumen de datos; y se proyecta ampliar el espacio de almacenamiento; mientras tanto manejan estándares para salvaguardar la información, y conservar la data histórica, en caso sea requerida en algún momento.

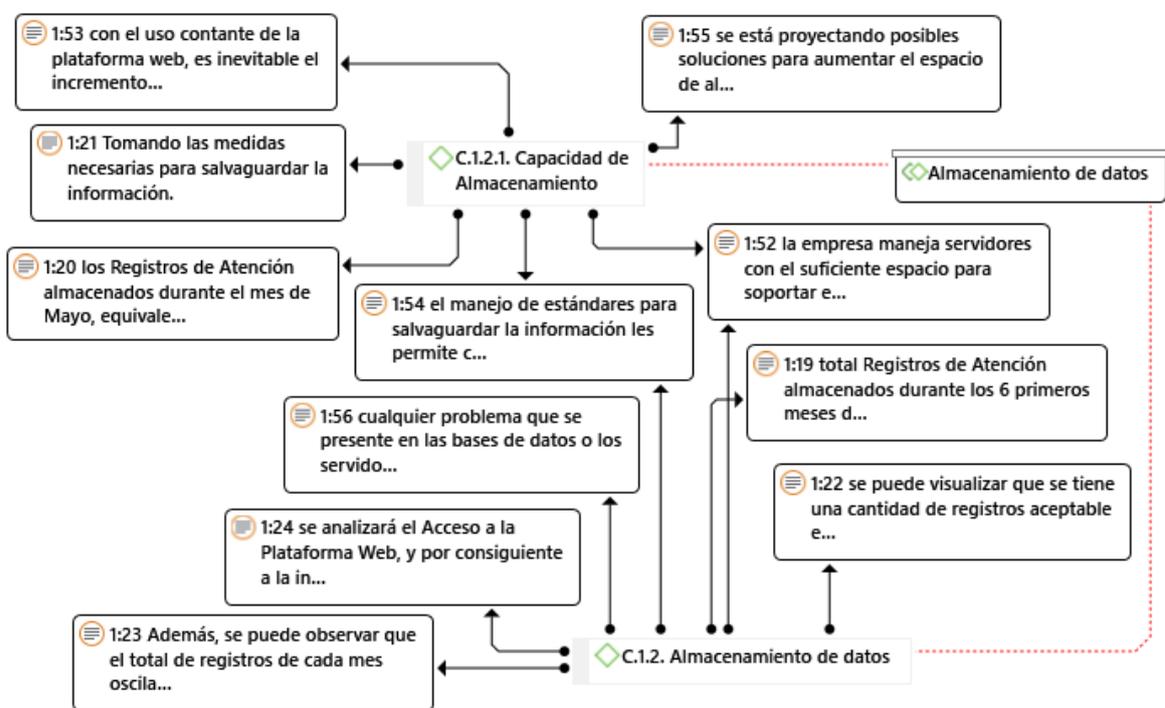


Figura 11. Análisis mixto de la sub categoría almacenamiento de datos

Fuente: Elaboración propia (2020)

En cuanto a la triangulación del análisis cuantitativo y cualitativo de la sub categoría Recuperación de Información, que se observa en la figura 12, se interpreta que la plataforma web actualmente no presenta problemas, al recuperar la información de la base de datos. Según el análisis, la plataforma web brinda un acceso amigable para que el usuario pueda acceder a la información registrada, y mantiene un nivel de seguridad para no poner en riesgo la información. Además, se puede afirmar que durante los primero 6 meses del año 2019, todos los usuarios lograron ingresar a la plataforma web, registrando

ingresos exitosos y con errores, por ejemplo el ingreso de contraseña incorrecta, *captcha* incorrecto, usuario incorrecto; logrando validar que no se puede ingresar fácilmente a la plataforma web sin los accesos correctos. Por otro lado, la plataforma web dispone de una bandeja de búsqueda, accesible para todos los usuarios, que permite realizar consultas y poder visualizar información. También se adicionaron mejoras para mayor facilidad al usuario y poder obtener mayor información; razón por la cual, se observa que durante los primeros seis meses del 2019 el porcentaje de consultas fue mejorando en cada mes.

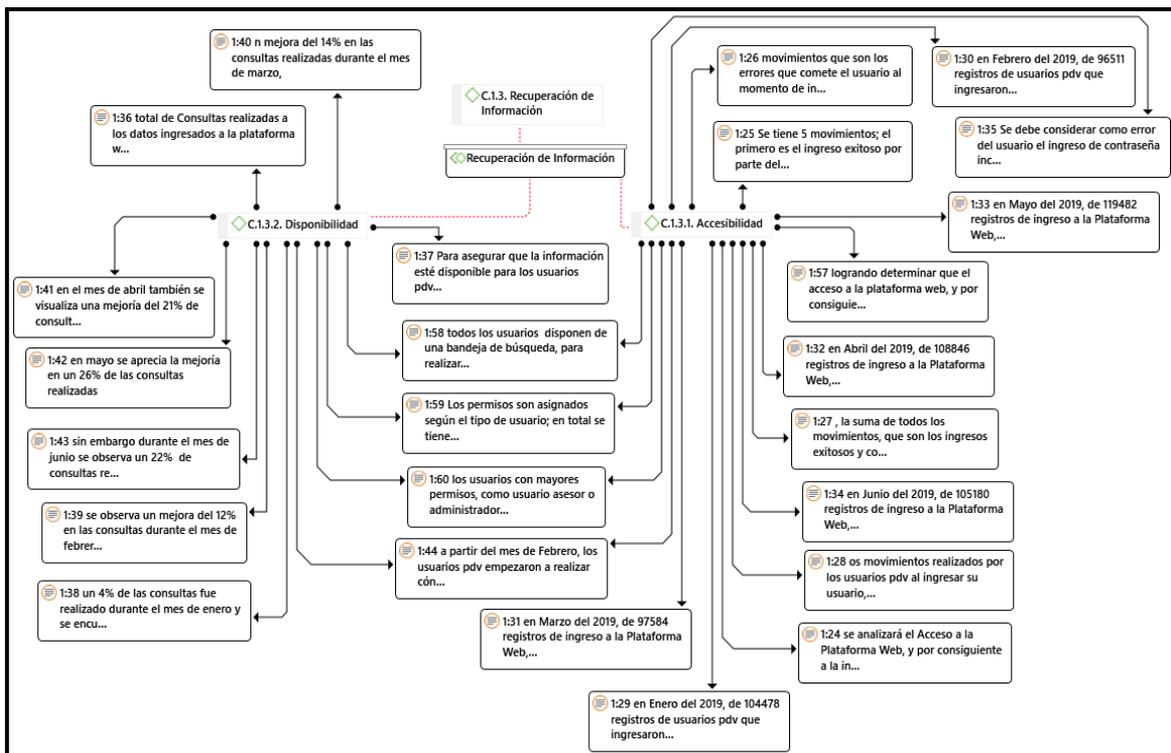


Figura 12. Análisis mixto de la sub categoría recuperación de información

Fuente: Elaboración propia (2020)

### 3.4. Propuesta

#### 3.4.1. Priorización de los problemas

En la siguiente tabla, se puede visualizar el diagnóstico para identificar los problemas principales.

Diagnóstico	Priorización de los problemas
1. Fallas al momento de diseñar la base de datos	Problema 1. Integridad de Datos
2. Fallas al diseñar la relación entre las tablas.	
3. Descoordinación cuando se crean campos	
4. Errores en el registro de datos	
5. Campos necesarios para relacionar tablas con datos diferentes patrones o incompletos	
6. Problemas al relacionar datos	
7. Campos con información compartida, contienen datos diferentes	Problema 2. Consistencia de datos
8. Datos con diferentes patrones.	
9. Datos no estandarizados	
10. Formato de datos no están establecidos.	
11. Campos de tablas sin un formato establecidos en sus datos.	Problema 3. Conformidad con datos
12. Campos con datos almacenado que presentan formatos incorrectos.	
13. Campos con datos que presentan patrones diferentes	
14. Campos con datos que presentan patrones incompletos	
15. Datos con características no definidas.	

*Cuadro 1.* Matriz de diagnóstico y priorización de los problemas.

*Fuente:* Elaboración propia (2020)

### 3.4.2. Consolidación del problema

En el presente trabajo, se diagnosticó que la empresa de ventas, presenta problemas en la gestión de datos correspondiente a la base de datos de la plataforma web. Se logró diagnosticar, por medio de un análisis cuantitativo y cualitativo, que los problemas más resaltantes son causados por la integración de los datos, la consistencia de los datos y la conformidad de los datos. Por lo cual, se determinó que el principal problema que presenta la gestión de datos, es por la baja calidad de sus datos.

### 3.4.3. Fundamentos de la propuesta

La propuesta tiene como base la teoría de la información, considerando a la información como un factor necesario, para elaborar y estructurar propuestas de mejoras y soluciones (Aladro, 2011; Monsalve, 2003). Además, es de gran importancia, porque brinda conocimiento, empleando datos correctamente registrados y almacenados, que puedan

brindan una mejor propuesta de solución o/y optimización en beneficio de la empresa. Es por esto que, se propone una mejora en los datos y la información, que emplea la empresa para sus respectivos reportes y aplicación de mejoras. Así mismo, se aplica la teoría de calidad total, que recomienda proponer un cambio enfocado en mejorar la organización y corregir los errores, causados por la carencia de conocimiento y/o atención, y lograr un estándar de cero defectos (Aguiar, 2004). Es por esto que, se propone un análisis de calidad en los datos y lograr determinar su estado actual, para luego realizar la optimización correspondiente a los datos.

#### **3.4.4. Categoría solución**

La propuesta, tiene como categoría solución, el perfilamiento y la limpieza de datos. Se puede considerar al perfilamiento de datos, como el primer paso para realizar la mejora en los datos. Se encarga de cumplir dos funciones primordiales; en primer lugar, obtener resultados estadísticos de cada columna, por ejemplo, la cantidad de valores nulos, los valores distintos dentro de la columna, tipo de dato o patrones más frecuentes de los datos en la columna, para luego analizar y evaluar la calidad de los datos. En segundo lugar, se encarga de examinar los valores de datos que pueden presentar dificultades para ser calculados, porque involucran múltiples columnas, presentan inclusión, dependencias funcionales y de orden. Ambas funciones, contribuyen para validar la presencia de errores y anomalías en los datos. Además, el perfilamiento de datos es considerada como una actividad importante para el profesional de IT (DAMA International, 2009; Naumann, 2014; Abedjan, Golab, & Naumann, 2017). Por otra parte, se entiende que el objetivo de la limpieza de datos (*Data cleaning* o *scrubbing*), es mejorar la calidad de los datos. Razón por la cual se busca identificar errores o inconsistencias; por ejemplo, campos vacíos, datos inválidos y/o datos duplicados que puedan estar presentes en archivos o base de datos, para minimizar el impacto negativo que pueda afectar a la información o para luego ser corregidos y brindar datos precisos y consistentes. Además, para ser aplicada con éxito, se debe cumplir su proceso, que incluye 3 etapas, identificar el error, realizar un diagnóstico y corregir anomalías. Por otro lado, la ausencia de una correcta limpieza de datos, puede afectar a otros datos o los resultados de un análisis que emplearon datos que contienen errores o anomalías (Rahm & Hai Do, 2000; Van den Broeck, Argeseanu, Eeckels, & Herbst, 2005; Osborne, 2013).

### 3.4.5. Direccionalidad de la propuesta

En la siguiente tabla, se puede visualizar la construcción de la propuesta, estableciendo los objetivos, las estrategias, las tácticas y las actividades.

Objetivo	Estrategia	Táctica	KPI
Objetivo 1. Definir herramienta y diagnosticar los datos.	Estrategia 1. Obtener acceso a los datos, y luego analizar el contenido y la estructura de los campos para identificar los errores e inconsistencias que contaminan la base de datos y luego identificar la fuente de origen de las mismas. También, se debe incorporar una herramienta con capaz de detectar los problemas de calidad de datos.	Táctica 1. Seleccionar la herramienta para realizar el perfilamiento y limpieza de los datos, y brinda los accesos y una guía para los datos.	KPI 1. Seleccionar herramienta.
		Táctica 2. Determinar las reglas de perfilamiento y calidad. Luego realizar el diagnóstico de los datos.	KPI 2. Establecer reglas de perfilamiento y calidad.
Objetivo 2. Realizar perfilamiento y diseñar medidas para optimizar los datos.	Estrategia 2. Realizar un análisis de forma exhaustiva a las columnas, y determinar las reglas de perfilamiento.	Táctica 3. Medir la calidad de los datos, empleando el perfilamiento a la tabla y sus campos.	KPI 3. Perfilamiento en los campos
		Táctica 4. Emplear las reglas y aplicarlas como consultas.	KPI 4. Adaptación de reglas.
Objetivo 3. Realizar la limpieza de los datos y análisis de calidad.	Estrategia 3. Emplear la herramienta seleccionada, para ejecutar la limpieza de los datos y validar el cumplimiento de las 6 dimensiones de la calidad (Precisión, Completitud, Consistencia, Conformidad, Duplicación, Integridad) determinados en cada campo. Luego, determinar los objetivos de la calidad de datos y asegurar su cumplimiento.	Táctica 5. Realizar la limpieza a los campos analizados.	KPI 5. Limpieza de datos.
		Táctica 6. Validar la mejora con base a las dimensiones de calidad.	KPI 6. Verificación

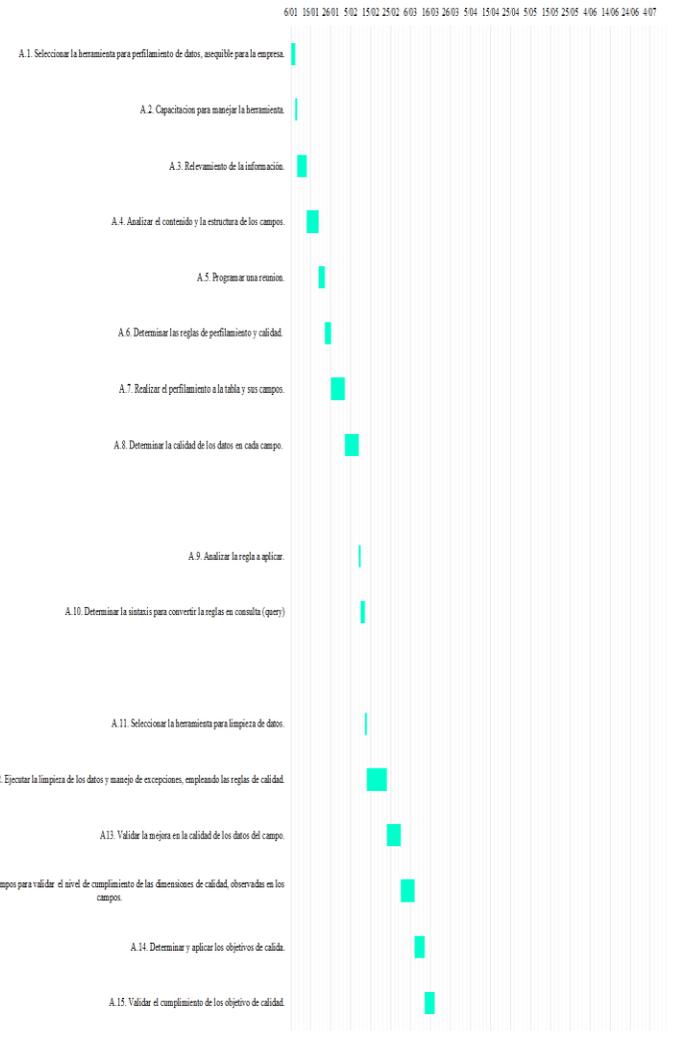
Cuadro 2. Matriz de direccionalidad de la propuesta

Fuente: Elaboración propia (2020)

### 3.4.6. Actividades y cronograma

En la siguiente tabla, se puede visualizar el diagrama de actividades con un periodo de tiempo acordado para aplicar la propuesta.

	Diagnostico	Priorización de los problemas	Consolidación del problema	Fundamento de la propuesta	Numero de la categoría solución	Objetivo	Estrategia	Táctica	Actividades	Inicio	Días	Fin	Responsable/s	Presupuesto de la implementación	Evidencia	
Cuantitativo	1.- Falta al momento de diseñar la base de datos	Problema 1. Integridad de Datos	En el presente trabajo, se diagnosticó que la empresa de ventas, presenta problemas en la gestión de datos de la plataforma web.	T e o r i a l a d e m e n t e C o n s i s t e n c i a d e d a t o	P e r f i l t r a r l o s d a t o s	Objetivo 1. DEFINIR HERRAMIENTA Y DIAGNOSTICAR LOS DATOS	Estrategia 1. Obtener acceso a los datos, y luego analizar el contenido y la estructura de los campos para identificar los errores e inconsistencias que contaminan la base de datos y luego identificar la fuente de orígenes de las mismas. También, se debe incorporar una herramienta con capacidad de detectar los problemas de calidad de datos.	Táctica 1. Seleccionar la herramienta para perfilamiento de datos, asegurable para la empresa. Táctica 2. Determinar las reglas de perfilamiento y calidad. Luego realizar el diagnóstico de los datos.	A.1. Seleccionar la herramienta para perfilamiento de datos, asegurable para la empresa. A.2. Capacitación para manejar la herramienta. A.3. Relevamiento de la información. A.4. Analizar el contenido y la estructura de los campos. A.5. Programar una reunión. A.6. Determinar las reglas de perfilamiento y calidad.	6/01/2020 8/01/2020 9/01/2020 14/01/2020 20/01/2020 23/01/2020	2 1 5 6 3 3	8/01/2020 9/01/2020 14/01/2020 20/01/2020 23/01/2020 26/01/2020	a. Profesional de Administración de Datos b. Profesional de Gestión de Datos a. Profesional de Gestión de Datos a. Profesional de Administración de Datos b. Profesional de Gestión de Datos a. Profesional de Administración de Datos b. Profesional de Gestión de Datos c. Usuarios	15000 9000 100 5 0 2	Evidencia 1. Seleccionar herramienta. Establecer reglas de perfilamiento y calidad.	
	2.- Falta al diseñar la relación entre las tablas	Problema 2. Consistencia de datos diferentes. Problema 3. Conformidad con datos	Se logró diagnosticar, por medio de un análisis cuantitativo y cualitativo, que los problemas más resaltantes son causados por la integración de los datos, la consistencia de los datos y la conformidad de los datos. Por lo cual, se determinó que el principal problema que presenta la gestión de datos, es por la baja calidad de sus datos.	d e f i n i t i v o s	d e t e r m i n a r l o s d a t o s	Objetivo 2. REALIZAR PERFILAMIENTO Y DISEÑAR MEDIDAS PARA OPTIMIZAR LOS DATOS	Estrategia 2. Realizar un análisis de forma exhaustiva a las columnas, y determinar las reglas de perfilamiento.	Táctica 3. Medir la calidad de los datos, empleando el perfilamiento a la tabla y sus campos. Táctica 4. Emplear las reglas y aplicarlas como consultas.	A.7. Realizar el perfilamiento a la tabla y sus campos. A.8. Determinar la calidad de los datos en cada campo. A.9. Analizar la regla a aplicar. A.10. Determinar la sintaxis para convertir la regla en consulta (query)	26/01/2020 2/02/2020 9/02/2020 9/02/2020 10/02/2020	7 7 0 0 2	2/02/2020 9/02/2020 10/02/2020 12/02/2020	a. Profesional de Administración de Datos b. Profesional de Gestión de Datos a. Profesional de Administración de Datos b. Profesional de Gestión de Datos	0 0 0 0 0	Evidencia 2. Perfilamiento en los campos. Adaptación de reglas. A.9. Analizar la regla a aplicar.	
	3.- Descoordinación cuando se crean campos		Evidencia 2. Perfilamiento en los campos. Adaptación de reglas.													
	4.- Errores en el registro de datos		A.10. Determinar la sintaxis para convertir la regla en consulta (query)													
	5.- Campos necesarios para relacionar tablas con datos diferentes patrones o incompletos.		0													
	6.- Problemas al relacionar datos.		2													
	7.- Campos con información compartida, contienen datos diferentes.		0													
	8.- Datos con diferentes patrones.		0													
	9.- Datos no estandarizados.		0													
10.- Formato de datos no están establecidos.	0															
Cualitativo	11.- Campos de tablas sin un formato establecidos en sus datos.	Problema 3. Conformidad con datos				Objetivo 3. REALIZAR LA LIMPIEZA DE LOS DATOS Y ANALISIS DE CALIDAD	Estrategia 3. Por medio de la herramienta seleccionada, se ejecuta la limpieza de los datos y se debe evaluar la mejora de métricas por las reglas de calidad. Para finalizar se valida el cumplimiento de las 6 dimensiones de la calidad (Precisión, Complejidad, Consistencia, Conformidad, Duplicación, Integridad) determinados en cada campo. Luego poder definir los objetivos de la calidad de datos y determinar su cumplimiento.	Táctica 5. Realizar la limpieza a los campos analizados. Táctica 6. Validar la mejora con base a las dimensiones de calidad.	A.11. Seleccionar la herramienta para limpieza de datos. A.12. Ejecutar la limpieza de los datos y manejo de excepciones, empleando las reglas de calidad. A.13. Validar la mejora en la calidad de los datos del campo. A.13. Analizar los campos para validar el nivel de cumplimiento de las dimensiones de calidad, observadas en los campos. A.14. Determinar y aplicar los objetivos de calidad. A.15. Validar el cumplimiento de los objetivos de calidad.	12/02/2020 13/02/2020 23/02/2020 1/03/2020 8/03/2020 13/03/2020	1 10 7 7 5 5	13/02/2020 23/02/2020 1/03/2020 8/03/2020 13/03/2020 18/03/2020	a. Profesional de Administración de Datos b. Profesional de Gestión de Datos a. Profesional de Administración de Datos b. Profesional de Gestión de Datos	0 0 0 0 0 0	Evidencia 3. Limpieza de datos. Verificación. A.12. Ejecutar la limpieza de los datos y manejo de excepciones, empleando las reglas de calidad. A.13. Validar la mejora en la calidad de los datos del campo. A.14. Determinar y aplicar los objetivos de calidad. A.15. Validar el cumplimiento de los objetivos de calidad.	
	12.- Campos con datos almacenados que presentan formatos incorrectos.															
	13.- Campos con datos que presentan patrones diferentes.															
	14.- Campos con datos que presentan patrones incompletos.															
	15.- Datos con características no definidas.															



Cuadro 3. Matriz de tácticas, actividades y cronograma

Fuente: Elaboración propia (2020)

## **Evidencia 1. Seleccionar herramienta y establecer reglas de perfilamiento y calidad.**

1.1. Se selecciona la herramienta más asequible para la empresa, de una lista de opciones, y su guía para que los colaboradores puedan aprender las funciones de la herramienta. Las opciones son:

- ✓ Data Quality Informatica
- ✓ Data Quality Microsoft
- ✓ Data Quality BDM
- ✓ Data Quality Talend
- ✓ Data Quality SIGMA

1.2. Luego, se efectúa el relevamiento de información, introduciendo al Profesional de Gestión de datos, en la estructura de la base de datos y su contenido, para luego ser analizados, y por medio de reuniones con el Profesional de Administración de Datos y usuarios, relacionar los problemas existentes con base a las 6 dimensiones de calidad que son:

Compleitud (*Completeness*): Son los datos nulos, en banco o con información desfavorable.

Conformidad (*Conformity*): Adaptar un dato a un formato ya establecido. Por ejemplo: los numero de celular, tienen un formato de 9 dígitos, sin guiones ni espacios

Consistencia (*Consistency*): Es la relación entre dos o más conjuntos de datos al momento de cruzarse y en el contenido.

Precisión/ Exactitud (*Accuracy*): La existencia de datos inválidos, con errores y obsoletos. Por ejemplo, fecha del último registro ingresado "1/3/1990".

Duplicación (*Duplicates*): Datos repetidos, que afectan la búsqueda y no aportan información.

Integridad (*Integrity*): Problemas en la integración de campos, que poseen clave primaria y foránea, y también en relación de datos que no tienen campos en común.

1.3. Las 6 dimensiones de calidad son empleadas en el documento de perfilamiento, en la parte de DIMENSIONES, para representar la anomalía que en el campo.

DOCUMENTO DISEÑO TECNICO - DATA PROFILING											
DESCRIPCIÓN						DIMENSION					
TABLA	CAMPO	TIPO DE DATO	PRECISION	NULL/NOT NULL	DETALLE	COMPLETITUD	CONFORMIDAD	CONSISTENCIA	EXACTITUD	DUPLICACIÓN	INTEGRIDAD
CUSTOMER_SHIPPING	COUNTRY	string	100	NULL	País				X		
CUSTOMER_SHIPPING	PHONE	string	100	NULL	Número de telefono	X	X				
CUSTOMER_SHIPPING	ADDR4	string	100	NULL	Contiene un código zip	X					

Figura 13. Documento de perfilamiento de datos - descripción y dimensiones

Fuente: Elaboración propia (2020)

1.4. También se registran las reglas de perfilamiento y calidad, necesarias para identificar el estado ideal del campo y aplicar las mejoras a las mejoras correspondientes.

REGLAS DE CALIDAD		
REGLAS DE PERFILAMIENTO	Primer Análisis	Segundo Análisis
Debe indicar el nombre del país con las tres primeras letras del país.	Los nombres de los países se deben representar con 3 dígitos que son las primeras letras del mismo.	
Debe contener datos numéricos.	Se debe separar los datos con valores no numéricos en otro campo	
Puede contener datos con 5 dígitos numéricos	Los datos con 4 dígitos numéricos, se debe agregar un "0" como primer dígito	
Puede contener datos con dígitos numéricos y alfabéticos en mayúscula.	-	

Figura 14. Documento de perfilamiento de datos - reglas de perfilamiento y calidad

Fuente: Elaboración propia (2020)

## Evidencia 2. Perfilamiento en los campos y adaptación de reglas.

2.1. Se emplea la herramienta Informatica Data Quality Analyst, para poder realizar el perfilamiento a los diferentes campos de la tabla seleccionada.

Filter By:	Columns and rules
Columns and rules: 15	Profile Run 3 of 3   15 of 15 columns   0 of 0 rules   1122 rows (All Rows)   05/30/2016 06:30:02 AM
Columns: 15	Sort By: Default   Ascending
Rules: 0	Columns and rules table with columns: Name, Null, Unique, Non-Unique, % Value, Pattern, Length, Data Type, Data Domain, Business Te
100% null: 0	PHONE: 1.33, 73.08, 25.58, 7812, 6, 26, 1.89%
100% unique: 1	
100% constant: 1	
Conflicting data type: 2	
Interrod data domains: 0	
Pattern outlier: 7	
Value Frequency outlier: 6	

Figura 15. Perfilamiento de los datos de una tabla

Fuente: Informatica Analyst 10 Lab Guide

2.2. En la siguiente actividad, para visualizar el perfilamiento individual de cada campo, se puede seleccionar un campo para visualizar su contenido y determinar su nivel de calidad por medio de la vista general (General), vista de valores (*Values*), tipo de dato (*Datatype*), vista de patrones (*Patterns*), vista previa de los datos (*Data Preview*). Por ejemplo, campos COUNTRY.

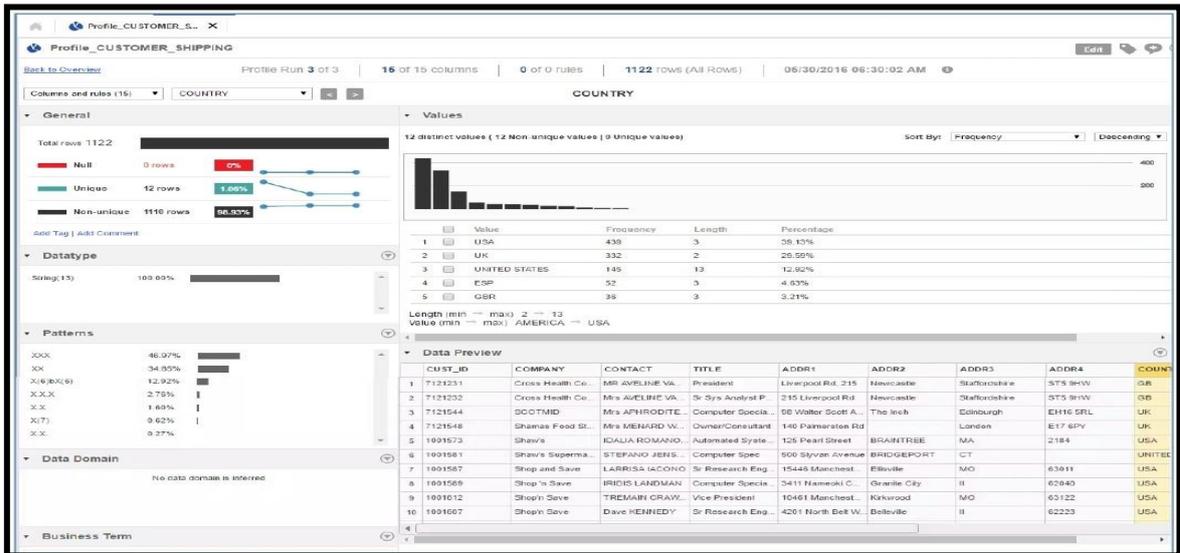


Figura 16. Vista general del perfilamiento de los datos de un campo

Fuente: Informatica Analyst 10 Lab Guide

2.3. También, visualizaremos el perfilamiento del campo ADDR4, el cual su regla de perfilamiento nos indica, que puede contener datos con 5 dígitos numéricos; razón por la cual, la regla de calidad nos indica que, a los datos con 4 dígitos numéricos se debe agregar un 0 como primer dígito.

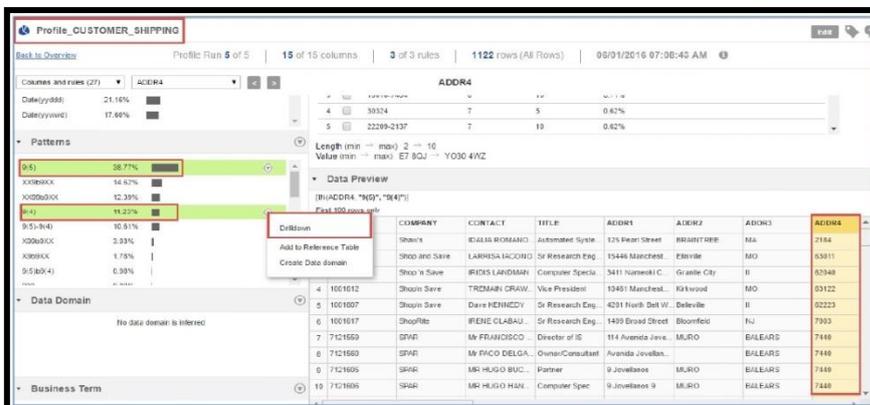


Figura 17. Patrones

Fuente: Informatica Analyst 10 Lab Guide

2.4. Para la siguiente actividad, se analiza la regla requerida. La herramienta nos brinda una opción que permite ingresar las reglas, creando una nueva regla.



Figura 18. Reglas en el perfilamiento de datos

Fuente: Informatica Analyst 10 Lab Guide

2.5. En la siguiente actividad, se analiza la regla, la cual nos indica que, a los datos con 4 dígitos numéricos se debe agregar un “0” como primer dígito. Luego, se transforma a una consulta (*query*), y se ingresa como una nueva regla.

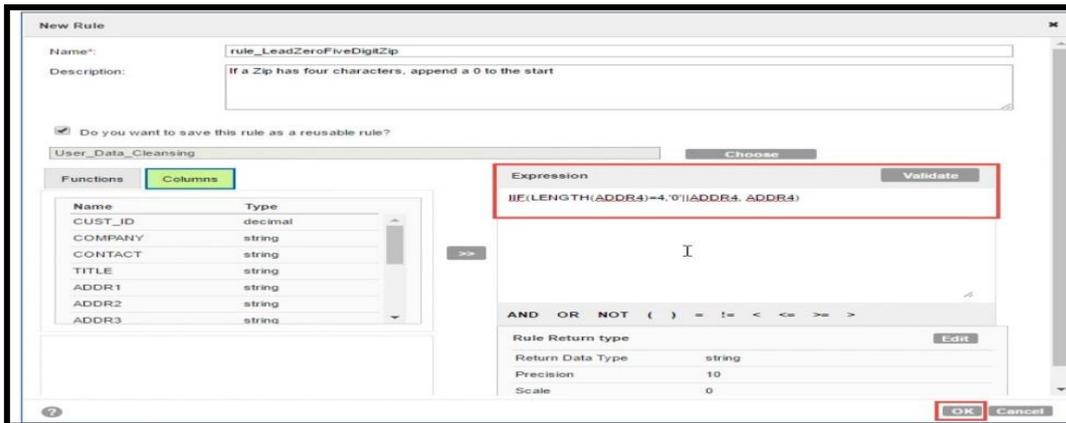


Figura 19. Crear regla en el perfilamiento de datos

Fuente: Informatica Analyst 10 Lab Guide

### Evidencia 3. Limpieza de datos y verificación

En la siguiente actividad, para el cumplimiento del objetivo, se continuará empleando la herramienta Informatica Data Quality Analyst; porque permite realizar la limpieza adecuada a los campos, aplicando las reglas calidad.

3.1. Ejecutada la regla para el campo ADDR4, explicada anteriormente, se puede observar como resultado final que los patrones con 4 dígitos numéricos desaparecieron, y los patrones con 5 dígitos numéricos aumentaron. Efectuando la limpieza de los datos de forma exitosa.

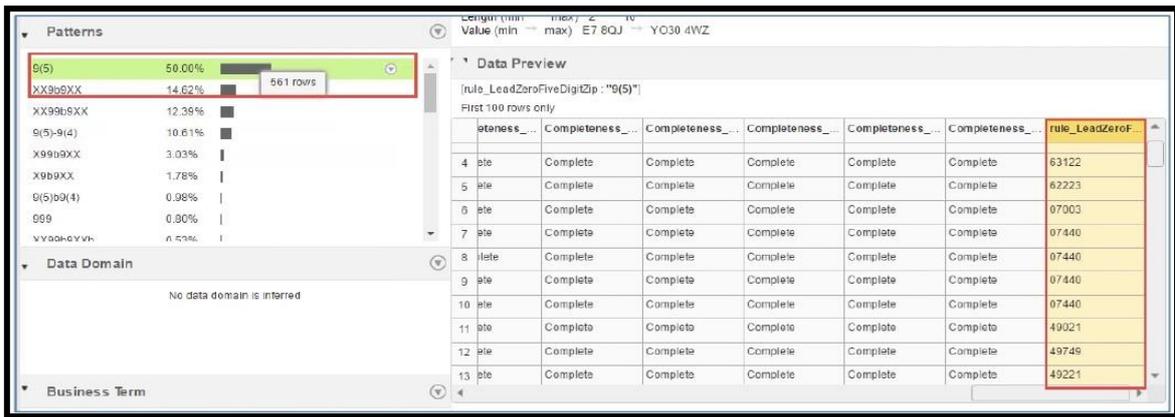


Figura 20. Aplicar regla en el perfilamiento de datos

Fuente: Informatica Analyst 10 Lab Guide

3.2. La herramienta también brinda opciones, para optimizar la limpieza de los datos. Por ejemplo, para limpiar el campo COUNTRY, se emplearán en primer lugar, la opción de tabla referencial (*Reference Table*); y luego la opción de regla específica (*Rule Specification*).

3.3. La tabla referencial (*Reference Table*), es empleada para estandarizar el valor de los datos que se encuentra en diferentes formatos dentro del campo. Por ejemplo, en campo COUNTRY, tiene datos como “USA” y “UNITED STATES” que equivalen a un valor. Y con la regla, se van a estandarizar con el valor “USA”. Sin embargo, la tabla referencial (*Reference Table*), no puede ser ejecutada, se guarda en una carpeta; y es empleada en una regla específica (*Rule Specification*) para ser ejecutada.

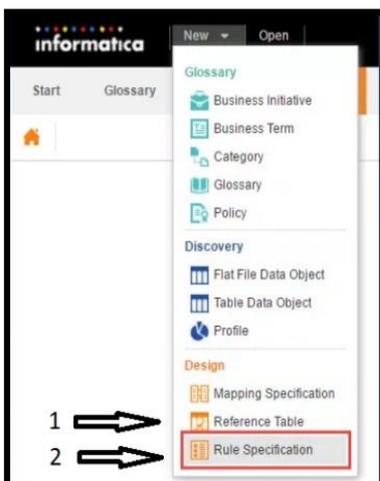


Figura 21. Inicio de regla específica

Fuente: Informatica Analyst 10 Lab Guide

3.4. Cuando se ingresa a la opción de crear una regla específica (*Rule Specification*), se debe agregar el nombre a la regla. Por ejemplo, rule\_StandardizationCountry.



Figura 22. Plataforma de la regla específica

Fuente: Informatica Analyst 10 Lab Guide

3.5. En esta actividad, es necesario completar los campos para que la regla se pueda ejecutar de forma exitosa. Razón por la cual, se debe indicar la tabla donde es necesario aplicar la regla (*Input*), la tabla referencial (*Reference Table*) con los valores a reemplazar (*Condition*) y la acción de reemplazar los valores (*Action*). Luego, se salva la regla.

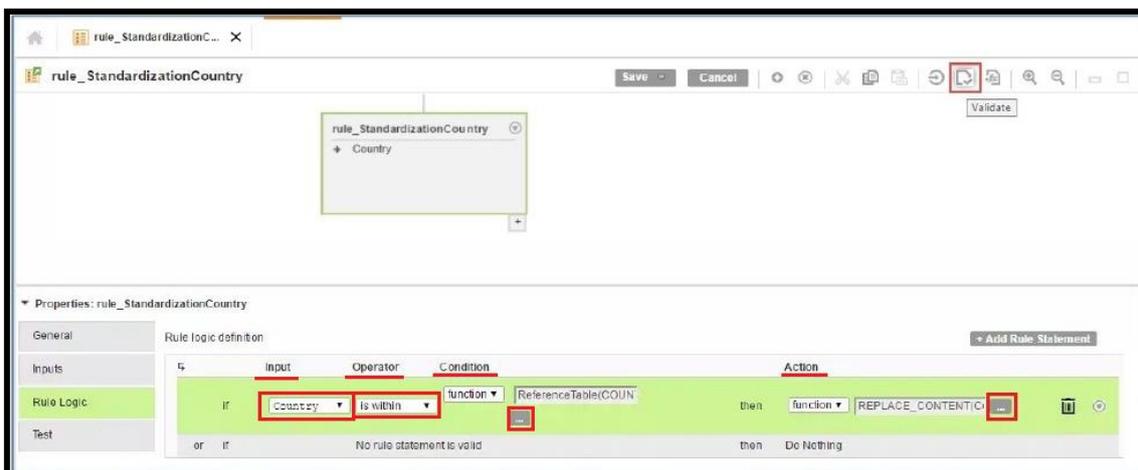


Figura 23. Creación de regla específica

Fuente: Informatica Analyst 10 Lab Guide

3.6. Se regresa al perfilamiento de la tabla, y se edita.

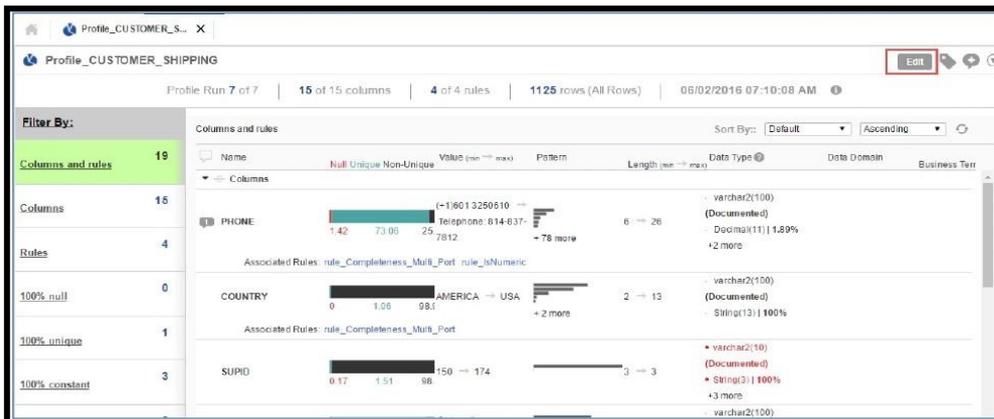


Figura 24. Editar perfilamiento de la tabla

Fuente: Informatica Analyst 10 Lab Guide

3.7. Luego, se aplica una regla existente, que será la regla específica (*Rule Specification*) creada anteriormente.

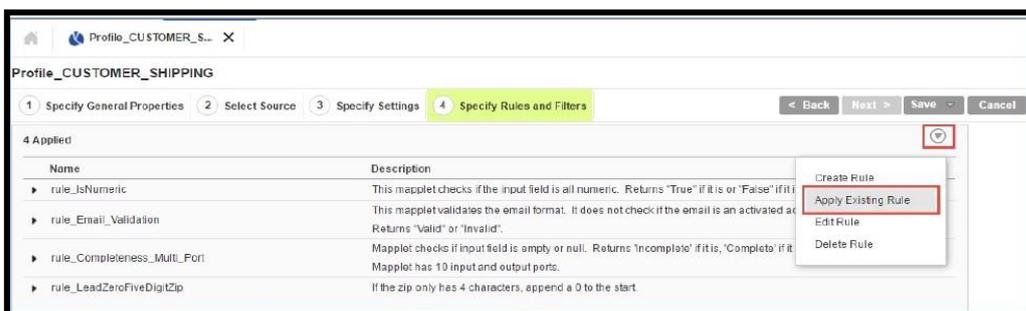


Figura 25. Aplicación de la regla específica

Fuente: Informatica Analyst 10 Lab Guide

3.8. A continuación, se puede validar la mejora en los datos, comparando la vista inicial del campo, y luego con la limpieza aplicada por medio de la regla.

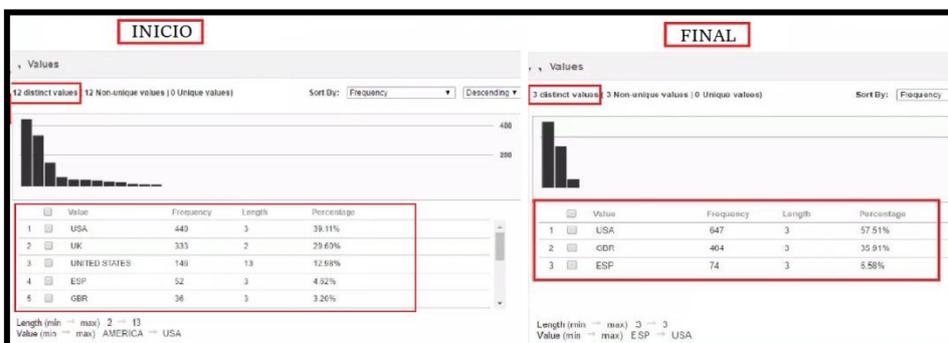


Figura 26. Comparación de perfilamiento después de la limpieza

Fuente: Informatica Analyst 10 Lab Guide

3.9. En la siguiente actividad, se valida el cumplimiento de las dimensiones de calidad, observadas en cada campo, y que fueron registrada en el documento de perfilamiento. Razón por la cual, se va a ejecutar la herramienta, seleccionando la opción agregar un *scorecarding*.

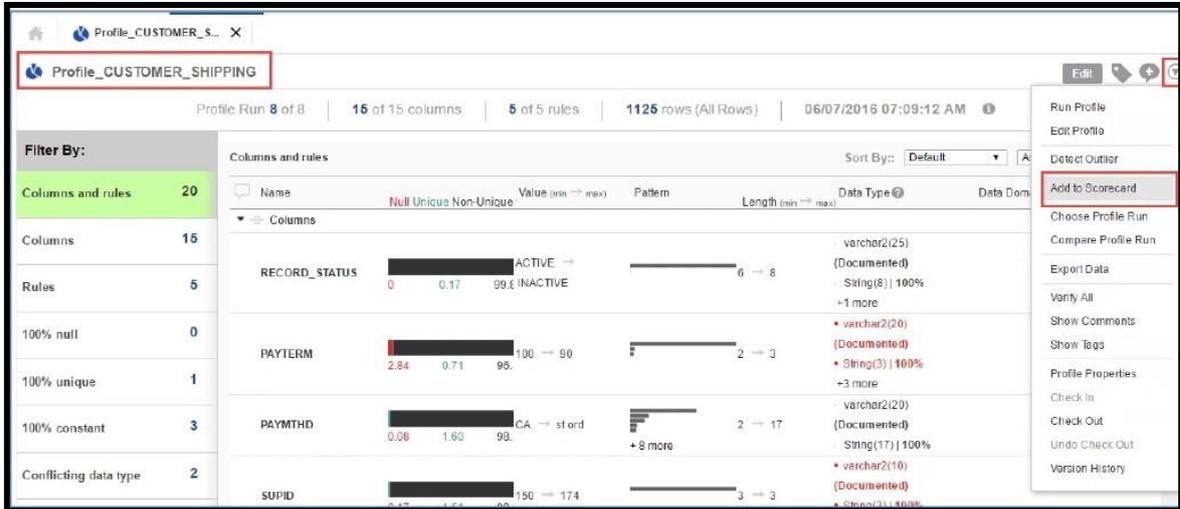


Figura 27. Creación del scorecarding

Fuente: Informatica Analyst 10 Lab Guide

3.10. En la primera vista para crear *scorecarding*, es asignar un nombre relacionado con la dimensión de calidad a evaluar en el campo. Por ejemplo, en el campo COUNTRY se va a evaluar su exactitud (*Accuracy*), por lo cual se le asigna el nombre de *Accuracy\_Country*

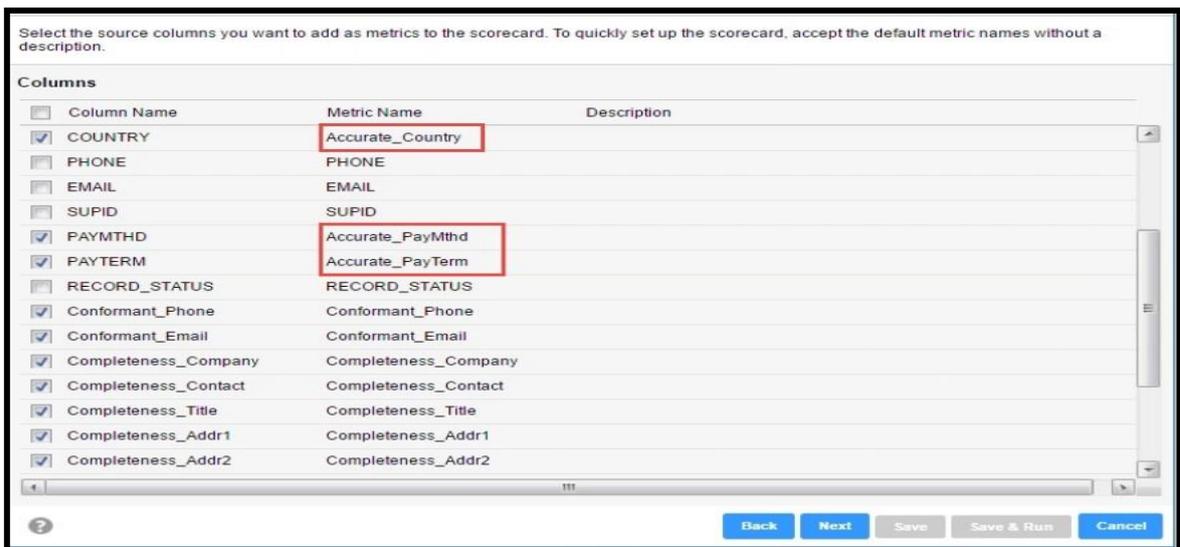


Figura 28. Renombrar campos para el scorecarding

Fuente: Informatica Analyst 10 Lab Guide

3.11. Luego, en la vista de valores (*Values*), se observa los datos únicos que posee el campo, para seleccionar aquellos considerados como válidos, y pasarlos a la vista de valores válidos (*Valid Values*).

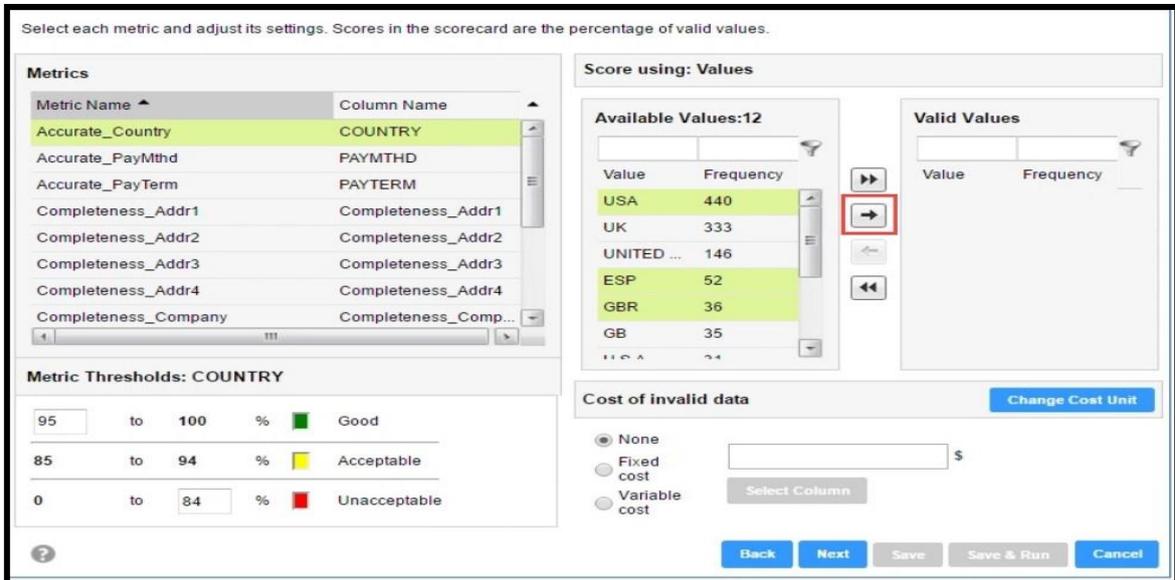


Figura 29. Seleccionar campos para el scorecarding

Fuente: Informatica Analyst 10 Lab Guide

3.12. A continuación se puede observar el resultado de la actividad anterior, y el cálculo del total de valores validos en el campo.

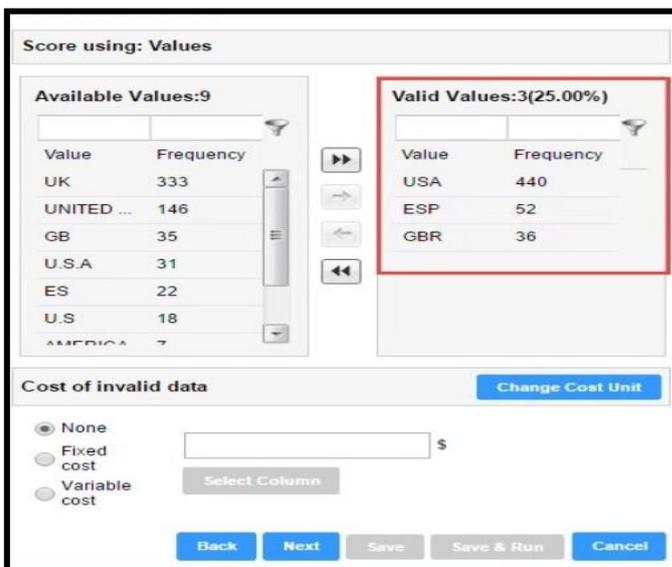


Figura 30. Validar campos para el scorecarding

Fuente: Informatica Analyst 10 Lab Guide

3.13. En la siguiente actividad, se emplea el documento de perfilamiento, que registra los objetivos de los campos en la dimensión asignada.

DOCUMENTO DISEÑO TECNICO - OBJETIVOS						
TABLA	CAMPO	DIMENSION	OBJETIVOS			
			Inaceptable (%)	Acceptable (%)	Bueno (%)	
CUSTOMER_SHIPPING	COUNTRY	EXACTITUD	0 a 84	85 a 94	95 a 100	
CUSTOMER_SHIPPING	PHONE	COMPLETITUD	0 a 84	85 a 94	95 a 100	
		CONFORMIDAD	0 a 84	85 a 94	95 a 100	
CUSTOMER_SHIPPING	ADDR4	COMPLETITUD	0 a 84	85 a 95	95 a 100	

Figura 31. Documento de perfilamiento de datos - objetivos

Fuente: Elaboración propia (2020)

3.14. Los objetivos de la empresa se ingresan en la vista métrica inicial (*Metric Thresholds*), y emplea la semaforización, para señalar si el nivel es bueno (*Good*), aceptable (*Acceptable*) o inaceptable (*Unacceptable*) en la dimensión de calidad asignada.

Select each metric and adjust its settings. Scores in the scorecard are the percentage of valid values.

Metric Name	Column Name
Accurate_Country	COUNTRY
Accurate_PayMthd	PAYMTHD
Accurate_PayTerm	PAYTERM
Completeness_Addr1	Completeness_Addr1
Completeness_Addr2	Completeness_Addr2
Completeness_Addr3	Completeness_Addr3
Completeness_Addr4	Completeness_Addr4
Completeness_Company	Completeness_Comp...

Score using: Values

Available Values:9		Valid Values:3(25.00%)	
Value	Frequency	Value	Frequency
UK	333	USA	440
UNITED ...	146	ESP	52
GB	35	GBR	36
U.S.A	31		
ES	22		
U.S	18		
AMERICA	7		

Cost of invalid data [Change Cost Unit](#)

None  
 Fixed cost  
 Variable cost

\$

**Metric Thresholds: COUNTRY**

95	to	100	%	<span style="color: green;">■</span>	Good
85	to	94	%	<span style="color: yellow;">■</span>	Acceptable
0	to	84	%	<span style="color: red;">■</span>	Unacceptable

Figura 32. Ingreso de métricas

Fuente: Informatica Analyst 10 Lab Guide

3.15. Para finalizar, la herramienta nos permite visualizar, la puntuación y el nivel de cumplimiento de la dimension de calidad asignado al campo, con base a los objetivos ingresados, para ser validadas. Por ejemplo, como se puede observar, *Accuracy\_Country* tiene un nivel de calidad inaceptable porque posee un 25% de valores válidos.



Figura 33. Vista de scorecarding inicial

Fuente: Informatica Analyst 10 Lab Guide

3.16. Para mejorar la puntuación asignada y obtener mejoras en la calidad de los datos, se recomienda repetir el proceso desde el punto 3.2. Por ejemplo, se puede observar la mejora de *Accuracy\_Country*; el cual se encontraba en un nivel inaceptable, y ahora se observa su mejora a un nivel aceptable.



Figura 34. Vista de scorecarding final

Fuente: Informatica Analyst 10 Lab Guide

## **IV. DISCUSIÓN, CONCLUSIONES Y RECOMENDACIONES**

### **4.1 Discusión**

En el presente trabajo se determinó que la empresa de ventas presenta problemas al manipular su gran volumen de datos, afectando la gestión de datos de su plataforma web. Por ello, fue necesario analizar la calidad, almacenamiento y recuperación de los datos en la plataforma web, para optimizar los procesos afectados y brindar una solución a la información afectada.

De los resultados cuantitativos y cualitativos, obtenidos de la investigación, se puede indicar que la calidad de los datos se encuentra afectada por la presencia de anomalías en los datos, impidiendo que puedan ser utilizados en análisis o en la solución de problemas. Por ello, decrece la confiabilidad en los datos porque carece de semejanza con la realidad y deja de ser considerada como una característica enfocada en cumplir las expectativas del cliente (Fernández, 2008; Valverde, Marotta & Vallespir, 2012; Caro, Fuentes & Soto, 2013; Osorio, Guerrero & González, 2017).

Al comparar el análisis de los resultados con otros trabajos de investigación, coincide con el trabajo de Chilón (2018) porque presenta una deficiente calidad de datos, que afecta el manejo de la información dentro de la empresa, y perjudica la relación con clientes y colaboradores. También, coincide con el trabajo de Balaguer & Arce (2019) porque no maneja correctamente el área encargada de asegurar la calidad de los datos; además, el desconocimiento en estándares y principios de calidad de datos, afecta la confiabilidad en los mismos para ser utilizados en la toma de decisiones. Igualmente, coincide con los trabajos de Dueñas (2018) y Sandoval (2018) porque presentan un bajo nivel de calidad causada por las inconsistencias en los datos, como campos en blanco, datos incorrectos, datos duplicados o datos con formatos diferentes, que afectan la información que maneja la empresa y el desempeño de la misma.

Por otro lado, se puede afirmar que el almacenamiento de datos en la empresa de venta, es empleado eficazmente, porque posee la capacidad de guardar y administrar grandes volúmenes de datos, para que puedan estar disponibles posteriormente (Vázquez,

2015; Instituto Nacional de Ciberseguridad, 2016). Sin embargo, se indicó que la presencia de anomalías afecta la calidad de datos; de modo que, coincide con el trabajo de González & Hernández (2017) porque presenta problemas en su gran volumen de información, a causa de inconsistencias presentes en los datos, dañando la confiabilidad en la información que emplean y almacenan.

También se puede asegurar que la recuperación de la información de la plataforma web, proporciona facilidades para que el usuario pueda acceder y localizar información relevante que se encuentra en un almacén de información o base de datos, y que podrían ser empleados para presentar la solución a un problema (Martínez, 2004; Salvador & Arquero, 2006; Bordignon & Tolosa, 2007; Vázquez, De León, Velasco, Villa & Briceño, 2017). Sin embargo, la información obtenida de una base de datos con anomalías, puede causar problemas a la empresa.

Por lo cual, también coincide con el trabajo de Chilón (2018) porque detectaron anomalías en sus datos, que afectaron los procesos en la empresa porque la información documentada presentaba errores. De igual modo, coincide con Santana, Muñoz, O'farrill, Martínez & Martínez (2017) porque detectaron que la información de la que disponían se encontraba incompleta, impidiendo el desarrollo de estrategias de crecimiento, de igual forma que en la empresa de ventas; sin embargo, la ausencia de información era causada por la presencia de anomalías o errores, que impedían su uso y reducían la confianza en la información.

Por último, el presente trabajo de investigación, coincide con el trabajo de Jiménez (2017) porque también presentan inconsistencias en sus datos, que impiden un análisis para la búsqueda de soluciones, dando lugar a problemas que ponen en peligro a la entidad. Sin embargo, no coincide con Cabrera (2019) por que la deficiente gestión de datos se debe a la ausencia de una organización y control de su gran volumen de datos, a diferencia que, en el presente trabajo brindó como resultado que la empresa de venta administra correctamente su gran volumen de datos. Esta situación, dificulta la aplicación de gestión de datos como una herramienta que brinda beneficio y ser considerado, junto a la información, como un activo de valor (DAMA International, 2009; Osorio, Guerrero & González, 2017; Portilla, 2017).

Dicho lo anterior, los trabajos de investigación citados proponen la limpieza de los datos para optimizar la calidad de los mismos; sin embargo, un paso importante a mencionar que se indica en el presente trabajo de investigación, es el perfilamiento de los datos; el cual es necesario para identificar los datos que requieren la limpieza. Lo que permitirá que la empresa limpie sus datos de forma eficiente, y emplearlos para brindar soluciones o en la toma de decisiones, mejorando los procesos en la empresa, y optimizando sus servicios para obtener satisfacción y fidelidad por parte de sus clientes y colaboradores.

## 4.2 Conclusiones

- Primera:** La propuesta del perfilamiento y limpieza de datos en la plataforma web de una empresa de ventas, busca la mejora en la gestión de datos, al analizar e identificar anomalías o irregularidades que presenta los datos; proporcionar una estadística del estado de los mismos; luego, realizar la limpieza correspondiente para que la información dañada sea corregida, y validar la optimización efectuada en los datos, proporcionando información más precisa y consistente.
- Segunda:** El análisis realizado con la información que proporcionó la empresa de venta, identificó la inconformidad e inconsistencia en los datos, por la presencia de anomalías o irregularidades en la base de datos, impactando en la integración de los datos; determinando la deficiente calidad de datos que maneja la plataforma web.
- Tercera:** En el diagnóstico, utilizando la información que brindó la empresa, se estableció que la empresa de ventas maneja de forma eficiente el almacenamiento de los datos de la plataforma web, manteniendo una estructura y organización para guardar su gran volumen de datos, y disponer de ellos posteriormente para brindar soluciones a problemas que se puedan presentar en la empresa de ventas.
- Cuarta:** En base al estudio de los resultados, se logró diagnosticar que la empresa de ventas maneja eficientemente la recuperación de los datos en la plataforma web, ofreciendo a los usuarios un acceso rápido y seguro, ingresando la clave y contraseña correspondiente del usuario, para localizar y adquirir fácilmente la información necesaria en la ejecución de tareas.
- Quinta:** Se identificó que la calidad de datos en la plataforma web es deficiente, puesto que sus datos presentan anomalías o irregularidades; en consecuencia, el almacenamiento y la recuperación de datos generan

desconfianza por el manejo de información imprecisa e irregular, perjudicando la eficiencia de la plataforma web de una empresa de ventas.

### 4.3 Recomendaciones

- Primera:** Implementar la propuesta de perfilamiento y limpieza de datos en la plataforma web, permitirá a la empresa de ventas evaluar e identificar el estado de la calidad de los datos, en los campos donde se observaron anomalías que afectan la correcta gestión de datos de la plataforma web, para luego realizar la limpieza respectiva y evaluar la mejora en la calidad de los datos, así como validar la autenticidad de los mismos, para ser empleados por la empresa.
- Segunda:** Realizar una minuciosa revisión a la base de datos de la plataforma web, para identificar todos los campos que presentan anomalías y no cumplan con las 6 dimensiones de calidad, para ser incluidas en la aplicación de la propuesta, y manejar una base de datos libre de anomalías o irregularidades, y con seguridad y confianza en su contenido.
- Tercera:** Realizar una copia de seguridad de toda la base de datos antes de aplicar la propuesta, incluidos tablas y campos que presenten datos dañados, nulos o en blanco, con el fin de poder realizar las comparaciones a los campos o tablas donde se encontraron datos que presentaban anomalías o irregularidades, y que fueron limpiados.
- Cuarta:** Crear una tabla que registre las consultas que realizan los usuarios. El modelo básico de la tabla puede contener los campos: id, dni, fecha, hora, módulo y campos adicionales que proporciona la empresa. También se recomienda validar que la información registrada en la base de datos optimizada, aplicada la propuesta de limpieza, se encuentre accesible y visible para el usuario en la plataforma web.
- Quinta:** Definir estándares de calidad para la gestión de datos de la plataforma web, con base a la información obtenida una vez aplicada la propuesta de perfilamiento y limpieza de datos, e implementarlas para prevenir que las

anomalías e irregularidades, identificadas y corregidas no vuelvan a presentarse en los nuevos datos.

## REFERENCIAS

- Abedjan, Z., Golab, L., & Naumann, F. (2017). Data Profiling: A Tutorial. *Conference: the 2017 ACM International Conference* (págs. 1747-1751). Chicago Illinois USA: Association for Computing Machinery New York NY United States. doi:10.1145/3035918.3054772
- Acevedo, A., & López, A. (2004). *El proceso de la entrevista: conceptos y modelos* (Cuarta ed.). México: Editorial Limusa. Obtenido de [https://books.google.com.pe/books/about/El\\_proceso\\_de\\_la\\_entrevista.html?id=V-Wi4\\_aHmKAC](https://books.google.com.pe/books/about/El_proceso_de_la_entrevista.html?id=V-Wi4_aHmKAC)
- Aguiar, F. (2004). Teoría de la decisión e incertidumbre: modelos normativos y descriptivos. *EMPIRIA. Revista de metodología de ciencias sociales*, 0(8), 139-160. doi:<https://doi.org/10.5944/empiria.8.2004.982>
- Aladro , E. (2011). La Teoría de la Información ante las nuevas tecnologías de la comunicación. *CIC. Cuadernos De Información Y Comunicación*, 16, 83 - 89. doi:[https://doi.org/10.5209/rev\\_CIYC.2011.v16.4](https://doi.org/10.5209/rev_CIYC.2011.v16.4)
- Álvarez, K., Romero, B., Cadenas, J., Coronado, D., & Rodríguez, R. (2016). Arquitectura para la Gestión de Datos Imperfectos en la Era de Big Data. *ReVeCom*, 3(2), 47-56.
- Balaguer, C., & Arce, S. (2019). *Dashboard para la medición de la confiabilidad de los datos del Datamart de una entidad financiera aplicando principios de calidad*. (Tesis de Grado Bachiller), Universidad Científica del Sur, Lima, Perú.
- Bernal, C. (2010). *Metodología de la investigación administración, economía, humanidades* (Tercera Edicion ed.). Bogotá, Colombia: Pearson Educación de Colombia Ltda.
- Bordignon, F., & Tolosa, G. (2007). Recuperación de información: un área de investigación en crecimiento. *Ciencias de la Información*, 38(1-2), 13-24. Obtenido de <https://www.redalyc.org/articulo.oa?id=181414865002>

- Cabrera, F. (2019). *Módulo de gestión de datos de biopsias para el registro de cáncer en el centro Oncológico de Holguín*. (Tesis de Licenciatura), Universidad de Holguín, Holguín, Cuba.
- Caro, A., Fuentes, A., & Soto, A. (2013). Desarrollando sistemas de información centrados en la calidad de datos. *Ingeniare. Revista chilena de ingeniería*, 21(1), 54-69. doi:<https://dx.doi.org/10.4067/S0718-33052013000100006>
- Chacón, J., & Rugel, S. (2018). Teorías, Modelos y Sistemas de Gestión de Calidad. *Revista ESPACIOS*, 39(50), 14.
- Chilón, W. (2018). *Implementación de un Sistema de Gestión de Calidad ISO 9001:2015 para el manejo de Información en la Empresa "Corporación El Dorado"*. (Tesis de Maestría), Universidad Nacional de Cajamarca, Cajamarca, Perú.
- DAMA International. (2009). *The DAMA Guide to the Data Management Body of Knowledge (DAMA-DMBOK) Spanish Edition* (Primera ed.). Technics Publications, LLC.
- Dueñas, D. (2018). *Optimización en el Proceso de Limpieza de Datos para la Construcción de un Datawarehouse usando Herramientas Open Source*. (Tesis de Maestría), Universidad Andina Néstor Cáceres Velásquez, Juliaca, Perú.
- Escolar, C. (2002). El Proceso de "Gestión de Datos". Construcción, medición y evaluación de los datos. *Cinta de Moebio*, 0(14), 196-199. Obtenido de <https://www.redalyc.org/articulo.oa?id=10101404>
- Fernández, W. (2008). Calidad de datos en levantamientos topográficos. *Revista de Topografía AZIMUT*, 2, 5-17. Obtenido de Revista de Topografía AZIMUT: <https://revistas.udistrital.edu.co/index.php/azimut/article/view/4044>
- González, C., & Hernández, C. (2017). *Diagnóstico de la calidad y el entendimiento de los datos para el análisis y toma de decisiones*. (Tesis de Especialidad), Institución Universitaria Politécnico Gran Colombiano, Bogotá, Colombia.
- Hernández, R., Fernández, C., & Baptista, P. (2014). *Metodología de la Investigación* (Sexta ed.). Ciudad de México: McGRAW-HILL / INTERAMERICANA EDITORES, S.A. DE C.V.

- Hurtado, J. (2000). *Metodología de la Investigación Holística* (Tercera ed.). Caracas, Venezuela: Fundación Sypal.
- Instituto Nacional de Ciberseguridad. (2016). *Almacenamiento seguro de la información. Una guía de aproximación para el empresario*. Obtenido de incibe: <https://www.incibe.es/protege-tu-empresa/guias>
- Jiménez, A. (2017). *Gestión de la calidad de los datos en la empresa*. (Tesis de Grado), Universidad Politécnica de Madrid, Madrid, España.
- Martínez, F. (2004). *Recuperación de información: modelos, sistemas y evaluación* (Cuarta Edición ed.). Murcia:KIOSKO JMC,2004. Obtenido de <http://libros.metabiblioteca.org/handle/001/227>
- Monsalve, A. (2003). *Teoría de Información y Comunicación Social* (Primera ed.). Quito, Quito, Ecuador: Producciones Digitales Abya-Yala. Obtenido de <https://books.google.com.pe/books?id=UIKOnsffiuwC&printsec=frontcover&hl=es#v=onepage&q&f=false>
- Naumann, F. (2014). Data Profiling Revisited. *ACM SIGMOD Record*, 42(4), 40-49. doi:10.1145/2590989.2590995
- Niño, V. (2011). *Metodología de la Investigación: diseño y ejecución* (Primera ed.). Bogotá, Colombia: Ediciones de la U. Obtenido de [https://books.google.com.pe/books?id=CyejDwAAQBAJ&printsec=frontcover&dq=Metodolog%C3%ADa+de+la+Investigaci%C3%B3n+V%C3%ADctor+Miguel+N%C3%B1o+Rojas&hl=es&sa=X&ved=0ahUKEwjvoLmF\\_rnoAhXbIrkGHVZGA3QQ6AEIJAA#v=snippet&q=entrevista&f=false](https://books.google.com.pe/books?id=CyejDwAAQBAJ&printsec=frontcover&dq=Metodolog%C3%ADa+de+la+Investigaci%C3%B3n+V%C3%ADctor+Miguel+N%C3%B1o+Rojas&hl=es&sa=X&ved=0ahUKEwjvoLmF_rnoAhXbIrkGHVZGA3QQ6AEIJAA#v=snippet&q=entrevista&f=false)
- Osborne, J. (2013). *Best practices in data cleaning*. London, United Kingdom: SAGE Publications, Inc. doi:<https://dx.doi.org/10.4135/9781452269948>
- Osorio, M., Guerrero, C., & González, M. (2017). La gobernabilidad de datos como apoyo en la gestión de datos de instituciones de educación superior. *Revista ESPACIOS*, 38(51), 11-12.
- Pereira, T., Portilla, I., & Rodríguez, N. (2019). Big data y Relaciones Públicas. Una revisión bibliográfica del estado de la cuestión. *Revista de Comunicación*, 18(1), 151-165. doi:<https://doi.org/10.26441/RC18.1-2019-A8>

- Portilla, J. (2017). Gobierno de datos, un potenciador de los sistemas de gestión de calidad. *Signos*, 9(2), 159-172. doi:<https://doi.org/10.15332/s2145-1389.2017.0002.10>
- Rahm, E., & Hai Do, H. (2000). Data Cleaning: Problems and Current Approaches. *IEEE Computer Society Technical Committee on Data Engineering*, 23(4), 3-13. Obtenido de <http://dbs.uni-leipzig.de>
- Raman, V., & Hellerstein, J. (2001). Potter's Wheel: An Interactive Data Cleaning System. *The VLDB Journal*, 27(1), 381-390. Obtenido de <https://dl.acm.org/doi/10.5555/645927.672045>
- Salvador, J., & Arquero, R. (2006). Una aproximación al concepto de recuperación de información en el marco de la ciencia de la documentación. *Investigación bibliotecológica*, 20(41), 13-43. Obtenido de [http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S0187-358X2006000200002](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-358X2006000200002)
- Sandoval, A. (2018). *Análisis de métodos y técnicas de Limpieza de Datos existentes*. (Tesis de Licenciatura), Pontificia Universidad Católica del Perú, Lima, Perú.
- Santana, M., Muñoz, M., O'farrill, M., Martínez, D., & Martínez, M. (2017). Sistema informático para la gestión de datos del docente. *Educación Médica Superior*, 31(1), 89-98. Recuperado el 10 de Marzo de 2020, de [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S0864-21412017000100009](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412017000100009)
- Valverde, C., Marotta, A., & Vallespir, D. (2012). *Análisis de la Calidad de Datos en Experimentos en Ingeniería de Software*. Obtenido de SEDICI Repositorio Institucional de la Universidad Nacional de La Plata: <http://sedici.unlp.edu.ar/handle/10915/23711>
- Van den Broeck, J., Argeseanu, S., Eeckels, R., & Herbst, K. (2005). Data Cleaning: Detecting, Diagnosing, and Editing Data Abnormalities. *PLoS Medicine*, 2(10), 0966 - 0970. doi:<https://doi.org/10.1371/journal.pmed.0020267>
- Vázquez, S. (2015). Tecnologías de almacenamiento de información en el ambiente digital. *Revista e-Ciencias de la Información*, 5(2), 1-18. doi:<http://dx.doi.org/10.15517/eci.v5i2.19762>

Vázquez, S., De León, M., Velasco, P., Villa, J., & Briceño, S. (2017). Recuperación de información a través del índice invertido en Be Intelligent. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*(21), 85-98. doi:<http://dx.doi.org/10.17013/risti.21.85-98>

## **ANEXOS**

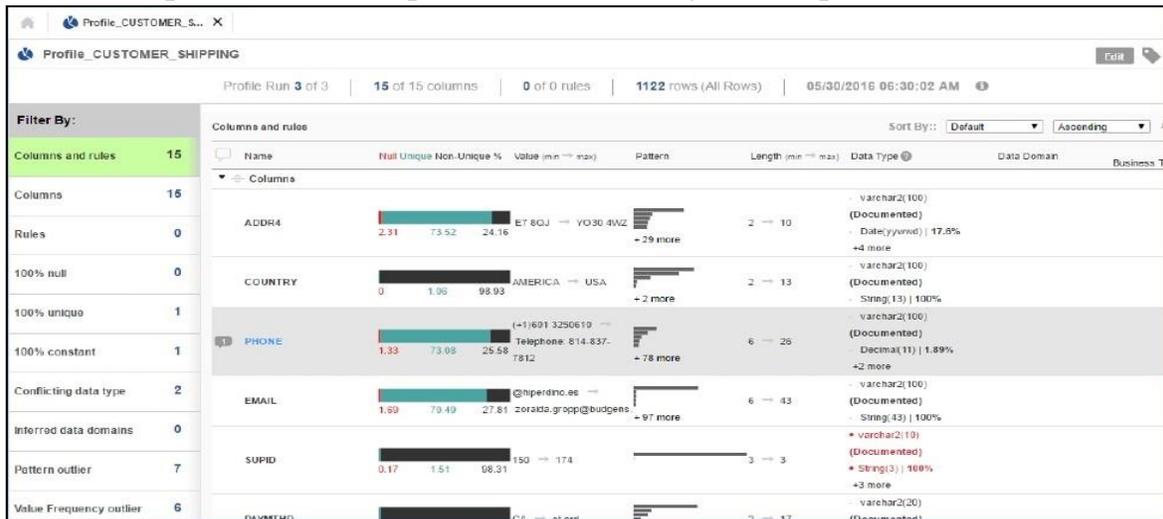
## Anexo 1: Matriz de la investigación

### Título:

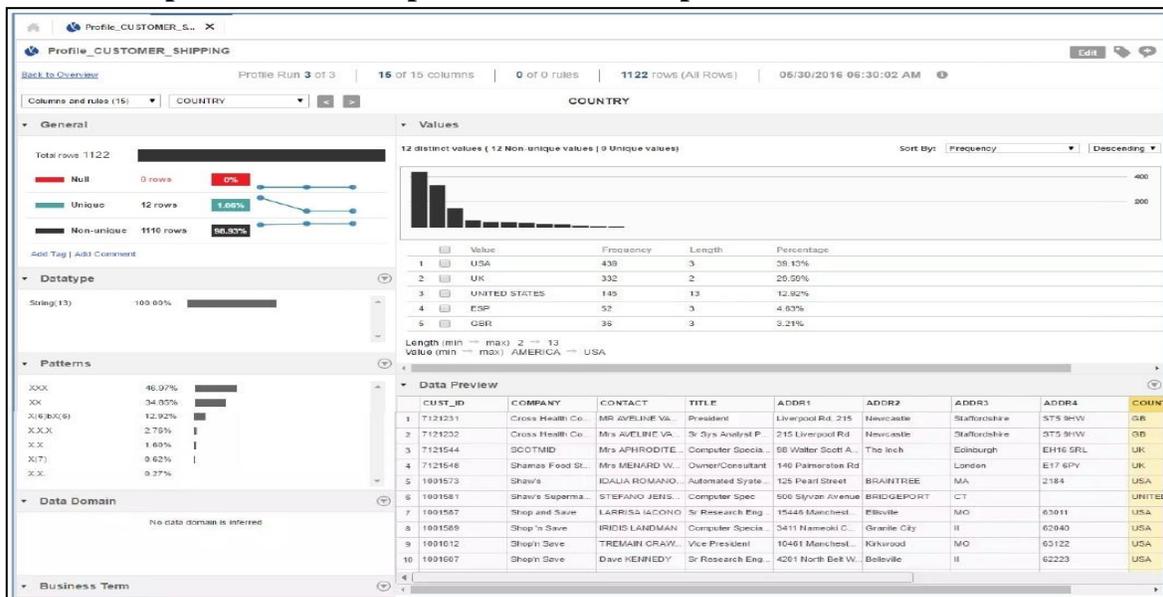
Problema general	Objetivo general	Categoría problema: Gestión de Datos	
		Sub categorías	Indicadores
¿Cómo mejorar la gestión de datos de la plataforma web de una empresa de ventas?	Proponer la optimización de la gestión de datos de la plataforma web de una empresa de ventas, Lima 2020.	Calidad de Datos	1. Consistencia de datos 2. Integridad de datos 3. Conformidad con datos
		Almacenamiento de Datos	4. Capacidad de Almacenamiento
		Recuperación de Información	5. Accesibilidad
			6. Disponibilidad
Problemas específicos	Objetivos específicos	Categoría solución: Perfilamiento y limpieza de datos	
¿Cómo validar la presencia de anomalías en los datos en la plataforma web de una empresa de ventas?	Analizar la calidad de los datos en la plataforma web de una empresa de ventas.	<b>Sub categorías emergentes</b>	
¿Cuál es la efectividad del almacenamiento de datos en la plataforma web de una empresa de ventas?	Diagnosticar la efectividad del almacenamiento de datos en la plataforma web de una empresa de ventas.		
¿Cuál es la efectividad de la recuperación de datos en la plataforma web de una empresa de ventas?	Diagnosticar la efectividad de la recuperación de datos en la plataforma web de una empresa de ventas.		
¿Qué factores causan las incidencias que perjudican la eficiencia en la plataforma web de una empresa de ventas?	Identificar si la calidad de los datos, la capacidad de almacenamiento y la recuperación de los datos perjudica la eficiencia en la plataforma web de una empresa de ventas.		
Tipo, nivel y método	Población, muestra y unidad informante	Técnicas e instrumentos	Procedimiento y análisis de datos
Sintagma: Holístico Tipo: Proyectiva Nivel: Comprensivo Método: Inductivo y deductivo	Población: Área de sistemas Muestra: 3 colaboradores. Unidad informante: 5.	Técnicas: Entrevista Instrumentos: Guía de Entrevista	Procedimiento: Autorización, revisión, aplicación de instrumentos, análisis, triangulación. Análisis de datos: Excel y ATLAS ti 8

## Anexo 2: Evidencias de la propuesta

### Plataforma para visualizar el perfilado de la tabla y sus campos.



### Plataforma para visualizar el perfilado de un campo.



## Plataforma para visualizar las métricas para la construcción del Scorecard

**Metrics**

Metric Name	Column Name
Accurate_Country	COUNTRY
Accurate_PayMthd	PAYMTHD
Accurate_PayTerm	PAYTERM
Completeness_Addr1	Completeness_Addr1
Completeness_Addr2	Completeness_Addr2
Completeness_Addr3	Completeness_Addr3
Completeness_Addr4	Completeness_Addr4
Completeness_Company	Completeness_Comp...

**Metric Thresholds: COUNTRY**

95	to	100	%	■	Good
85	to	94	%	■	Acceptable
0	to	84	%	■	Unacceptable

**Score using: Values**

**Available Values:9**

Value	Frequency
UK	333
UNITED ...	146
GB	35
U.S.A	31
ES	22
U.S	18
AMERICA	7

**Valid Values:3(25.00%)**

Value	Frequency
USA	440
ESP	52
GBR	36

**Cost of invalid data** Change Cost Unit

None  
 Fixed cost \$  
 Variable cost

Select Column

Back
Next
Save
Save & Run
Cancel

## Plataforma para visualizar el Scorecard.

Scorcard		Properties	
CustomerOrders - metrics			
Name	Total Rows	Invalid Rows	Score
Default			-
▼ Completeness			88.55
Completeness_Contact	1125	0	100
Completeness_Company	1125	0	100
Completeness_Email	1125	19	98.31
Completeness_Addr4	1125	26	97.69
Completeness_Title	1125	14	98.76
Completeness_Addr1	1125	17	98.49
Completeness_Addr2	1125	38	96.62
Completeness_Country	1125	0	100
Completeness_Addr3	1125	59	94.76
Completeness_Phone	1125	16	98.58

### Anexo 3: Instrumento cuantitativo



Universidad  
Norbert Wiener

#### Ficha de Registro Documental

<b>Título del documento:</b>	<b>Registros de atención a los puntos de venta</b>	
<b>Periodo o año:</b>	<b>2018</b>	
<b>Objetivo del documento:</b>	<b>Descripción del documento:</b>	<b>El documento responde al área de:</b>
Brindar información resaltante por medio de registros almacenados en la base de datos de la plataforma web.	Los documentos obtenidos se basan en datos registrados en la base de datos y documentos Excel.	Tecnología y transformación digital. (Área de Sistema)

Nro.	Documentos	Descripción	Registro Documental	Indicadores	Análisis
1	Base de Datos de la plataforma web.	Proporciona información que se encuentra almacenada directamente en la base de datos.	Registros de Atención ingresados de Enero a Junio del año 2019	Atención por usuario	Contabilizar el total de registros ingresados para determinar la capacidad de almacenamiento.
			Acceso a la Plataforma Web de Enero a Junio del año 2019	Ingreso por usuario	Analizar la recuperación de datos, determinando la accesibilidad la disponibilidad de los mismos, para los usuarios.
			Consultas de usuarios pdv de Enero a Junio del año 2019.	Consultas por usuario.	
2	Tabla de Empleados de la Base de Datos de la Empresa.	Lista de colaboradores asignados a los puntos de venta.	Patrones de DNI de usuario pdv Enero, Febrero y Marzo del año 2019	Documento Nacional de Identidad del usuario	Identificar la calidad de los datos en campos específicos.
3	Exportable de Registro de atención.	El archivo, obtenido del aplicativo web, presenta los registros extraídos de la base de datos del servicio de atención al cliente, en formato csv, logrando visualizar los campos que pueden presentar inconsistencias.	Patrones del campo Número de Consulta Enero, Febrero y Marzo del año 2019	Número de Consulta	
			Patrones en el campo Número de Orden Enero, Febrero y Marzo del año 2019	Número de Orden	

Fecha:.....Lugar:.....

## Anexo 4: Instrumento cualitativo



### Entrevista

Concepto de entrevista	Se puede considerar que la entrevista es una técnica, que apoya en la recolección de información relevante en la investigación. Se emplea través de una conversación entre el investigador y el participante, a base de preguntas y respuestas previamente seleccionadas, que sean capaces de obtener puntos de vista, opiniones o ideas de los participantes, sin perder el rumbo durante la conversación. (Niño, 2011).	
Entrevistados		
<b>Entrevistado (Entv.1)</b>	<b>Entrevistado (Entv.2)</b>	<b>Entrevistado (Entv.3)</b>
<b>Analista Programador:</b>	<b>Programador:</b>	<b>Analista y desarrollador:</b>

### Ficha de entrevista

Datos básicos:

Cargo o puesto en que se desempeña	
Nombres y apellidos	
Código de la entrevista	
Fecha	
Lugar de la entrevista	

Nro.	Preguntas de la entrevista
1	¿Considera que las tablas se encuentran integradas correctamente?
2	Cuando envían el requerimiento de crear un nuevo campo en la plataforma web ¿Detallan las características del campo?
3	¿El contenido de su base de datos presenta replicas? ¿Dónde se almacena? En su opinión ¿Sus datos presenten inconsistencia?
4	¿La capacidad de almacenamiento actual es factible frente al constante aumento de registros? ¿Toman medidas frente a este hecho?
5	¿Recomendaría aumentar la capacidad de almacenamiento? En caso contrario ¿Qué otra opción recomendaría?
6	¿Presento algún problema por sobrecarga de información que afectara los procesos?
7	¿La réplica de su base de datos afecta la capacidad de almacenamiento?
8	¿La plataforma web permite a los usuarios acceder a la información que ellos registraron? ¿Pueden acceder a los registros de otros usuarios? ¿Pueden editarla los datos? ¿Qué otras facilidades brinda la plataforma para que los usuarios puedan acceder y disponer de la información?

Observaciones

.....
.....

## Anexo 5: Base de datos (instrumento cuantitativo)

```

1 select * from lh_ticket where date_creation_ticket between "2019-01-01 00:00:00" AND "2019-02-28
23:59:59";

```

id	state_ticket_id	users_id	motive_id	module_id	type_id	chat_id	etiqueta	number_order	dat
400965	4	45947796	125		1	2	353850 (Null)	527256	201
400966	4	45962582	112		1	2	353851 (Null)	13205470	201
400967	4	48306108	125		1	1	353852 (Null)	948930537	201
400968	4	45947796	175		2	2	353853 (Null)	987765103	201
400969	4	73206899	170		2	2	353854 (Null)	935905869	201
400970	4	46651857	169		2	2	353855 (Null)	917131311	201
400971	4	45947796	9		1	1	353856 (Null)	936675633	201
400972	4	48327704	9		2	1	353857 (Null)	942894635	201
400973	4	74541257	68		1	1	353858 (Null)	963380812	201
400974	4	41147847	111		1	2	353859 (Null)	51307950	201
400975	4	47537690	171		2	2	353860 (Null)	989843557	201
400976	4	73206899	171		2	2	353861 (Null)	989843557	201
400977	4	46651857	171		2	2	353862 (Null)	931232740	201
400978	4	73206899	111		1	2	353863 (Null)	13633271	201
400979	4	46651857	111		1	2	353864 (Null)	15392197	201
400980	4	44984155	169		2	2	353865 (Null)	927691513	201
400981	4	41147847	37		1	1	353866 (Null)	984007107	201
400982	4	73206899	111		1	2	353867 (Null)	12409195	201
400983	4	45947796	125		1	2	353868 (Null)	949531119	201
400984	4	47537690	169		2	2	353869 (Null)	951760479	201
400985	4	47537690	169		2	2	353870 (Null)	953360439	201
400986	4	44761964	9		1	1	353871 (Null)	996291430	201
400987	4	71987454	171		2	2	353872 (Null)	929851484	201
400988	4	73206899	169		2	2	353873 (Null)	922196303	201
400989	4	44177871	170		2	2	353874 (Null)	951775302	201
400990	4	47537690	111		1	2	353875 (Null)	13363616	201

## Anexo 6: Transcripción de las entrevistas o informe del análisis documental



### Ficha de entrevista

Datos básicos:

Cargo o puesto en que se desempeña	Analista Programador
Nombres y apellidos	Edgar Alvarado
Código de la entrevista	Entrevistado 1 (Entv.1)
Fecha	09-05-2020
Lugar de la entrevista	Entrevista Virtual por Zoom

Nro.	Preguntas de la entrevista
1	¿Considera que las tablas se encuentran integradas correctamente?
2	Cuando envían el requerimiento de crear un nuevo campo en la plataforma web ¿Detallan las características del campo?
3	¿El contenido de su base de datos presenta replicas? ¿Dónde se almacena? En su opinión ¿Sus datos presenten inconsistencia?
4	¿La capacidad de almacenamiento actual es factible frente al constante aumento de registros? ¿Toman medidas frente a este hecho?
5	¿Recomendaría aumentar la capacidad de almacenamiento? En caso contrario ¿Qué otra opción recomendaría?
6	¿Presento algún problema por sobrecarga de información que afectara los procesos?
7	¿La réplica de su base de datos afecta la capacidad de almacenamiento?
8	¿La plataforma web permite a los usuarios acceder a la información que ellos registraron? ¿Pueden acceder a los registros de otros usuarios? ¿Pueden editarla los datos ¿Qué otras facilidades brinda la plataforma para que los usuarios puedan acceder y disponer de la información?

Observaciones

.....
.....
.....

### Entrevistado1 (Entv.1)

Nro.	Preguntas de la entrevista	Respuestas
1	¿Considera que las tablas se encuentran integradas correctamente?	Desde mi punto de vista, se puede considerar que la integración entre las tablas esta correcta. Sin embargo, si un especialista revisa la data, puede que encuentre algún error.
2	Cuando envían el requerimiento de crear un nuevo campo en la plataforma web ¿Detallan las características del campo?	Si, las características del campo son especificadas en el documento de requerimiento; sin embargo, en ocasiones esto no ocurre, y se programa reuniones para detallar las características.
3	¿El contenido de su base de datos presenta replicas? ¿Dónde se almacena? En su opinión ¿Sus datos presenten inconsistencia?	Si, para respaldar la información en caso de caída del servidor principal. Las réplicas se almacenan en servidores virtuales y físicos. Considero que no; sin embargo, entre sistemas que comparten información se puede presentar inconsistencias. Por ejemplo, el registro de nuevos usuarios, se almacena en la base de datos general de la empresa y en la base de datos de la plataforma web. Se menciona que presentan inconsistencias, porque a pesar de que registran la misma información, no se almacena de igual forma, puesto que algunos campos de las bases de datos difieren por el tipo de dato que se registra
4	¿La capacidad de almacenamiento actual es factible frente al constante aumento de registros? ¿Toman medidas frente a este hecho?	Actualmente, la capacidad de almacenamiento del servidor es alta; sin embargo, desconozco cuanto espacio se encuentra ocupado hasta el momento. Además, por política de la empresa, se tiene un responsable que se encuentra pendiente de la capacidad de almacenamiento de los servidores.
5	¿Recomendaría aumentar la capacidad de almacenamiento? En caso contrario ¿Qué otra opción recomendaría?	No puedo opinar, porque no me encuentro informado en relación a ese tema. Depende de la situación, podría ser factible habilitar un servidor virtual.
6	¿Presento algún problema por sobrecarga de información que afectara los procesos?	Bueno, se ha tenido problemas, pero no afectan directamente a la base de datos; sino a las consultas que se hacen, porque pueden ser muchas consultas que buscan información con un rango de tiempo muy amplio. Esto ocurre cuando los campos de las tablas no se encuentran optimizados. Por ejemplo, a causa de la ausencia de índices.
7	¿La réplica de su base de datos afecta la capacidad de almacenamiento?	Se ha presentado inconsistencia en algunas réplicas; sin embargo, fueron atendidas inmediatamente y hasta el momento no hemos presentado ningún problema que interfiera con los procesos ni que sature los servidores.
8	¿La plataforma web permite a los usuarios acceder a la información que ellos registraron? ¿Pueden acceder a los registros de otros usuarios? ¿Pueden editarla los datos? ¿Qué otras facilidades brinda la plataforma para que los usuarios puedan acceder y disponer de la información?	Si, la plataforma web presenta un módulo en donde pueden visualizar sus tickets. Este módulo se encuentra habilitado para los distintos tipos de usuarios. Si, los usuarios pueden visualizar los tickets propios y de otros usuarios. Pero solo los usuarios agente, asesor y administrador tienen permisos para editar la información. Los usuarios agente, asesor y administrador tienen permisos para acceder a los módulos de informes, en donde pueden visualizar la calidad de la atención por medio de gráficos, y permisos para poder extraer información por medio de una opción que les permite exportar la información, de la bandeja de ticket, en un documento Excel.

### Ficha de entrevista

Datos básicos:

Cargo o puesto en que se desempeña	Programador
Nombres y apellidos	Junior Huaranca
Código de la entrevista	Entrevistado 2 (Entv.2)
Fecha	09-05-2020
Lugar de la entrevista	Entrevista Virtual por Zoom

Nro.	Preguntas de la entrevista
1	¿Considera que las tablas se encuentran integradas correctamente?
2	Cuando envían el requerimiento de crear un nuevo campo en la plataforma web ¿Detallan las características del campo?
3	¿El contenido de su base de datos presenta replicas? ¿Dónde se almacena? En su opinión ¿Sus datos presenten inconsistencia?
4	¿La capacidad de almacenamiento actual es factible frente al constante aumento de registros? ¿Toman medidas frente a este hecho?
5	¿Recomendaría aumentar la capacidad de almacenamiento? En caso contrario ¿Qué otra opción recomendaría?
6	¿Presento algún problema por sobrecarga de información que afectara los procesos?
7	¿La réplica de su base de datos afecta la capacidad de almacenamiento?
8	¿La plataforma web permite a los usuarios acceder a la información que ellos registraron? ¿Pueden acceder a los registros de otros usuarios? ¿Pueden editarla los datos ¿Qué otras facilidades brinda la plataforma para que los usuarios puedan acceder y disponer de la información?

Observaciones

.....
.....
.....

## Entrevistado2 (Entv.2)

Nro.	Preguntas de la entrevista	Respuestas
1	¿Considera que las tablas se encuentran integradas correctamente?	Considero que las tablas iniciales, como las tablas de configuración y seguridad para los permisos de usuario, están bien relacionadas. Sin embargo, las tablas principales donde se relaciona el chat, el ticket y otras entidades que han ido aumentándose posteriormente, y que son tablas importantes porque de ellas depende la correcta comunicación entre los usuarios agente y punto de venta, son empleado como indicadores en los reportes, han sido agregadas de una manera incorrecta en su relación con otras tablas.
2	Cuando envían el requerimiento de crear un nuevo campo en la plataforma web ¿Detallan las características del campo?	Inicialmente, los requerimientos para crear nuevos campos, en la plataforma web y en la base de datos, se especifican detalladamente. Sin embargo, actualmente esto no ocurre, y los requerimientos enviados no son especificados. A pesar de que lo más recomendable es analizar el nuevo campo y luego enviar la documentación.
3	¿El contenido de su base de datos presenta replicas? ¿Dónde se almacena? En su opinión ¿Sus datos presenten inconsistencia?	Claro, es una necesidad. Manejamos 8 réplicas de la base de datos. Se observó que algunos campos se encuentran mal relacionados, causado por un mal registro de los datos y afectando las consultas y a la base de datos. Además, también pueden existir inconsistencias a causa de la discrepancia entre las características asignadas al campo y los datos que se registran. Por ejemplo, al asignar un máximo de caracteres a un campo tipo cadena, y la información que se registra sobrepasa ese límite, entonces se registra información incompleta y, a menos que se valide el registro completo, nunca se detectara ese error.
4	¿La capacidad de almacenamiento actual es factible frente al constante aumento de registros? ¿Toman medidas frente a este hecho?	Sí, la capacidad de almacenamiento actual es factible frente al incremento constante de datos. Sin embargo, se tiene que pensar que ha futuro esta situación puede cambiar. Razón por la cual, debe considerar algún método para salvaguardar su información en caso no logren tener la capacidad suficiente para almacenar sus datos. Actualmente no, pero se tienen diferentes opciones que pueden ser evaluadas para determinar la mejor solución. Por ejemplo separa la información muy antigua, a una base de datos histórica, o aumentar la capacidad de almacenamiento.
5	¿Recomendaría aumentar la capacidad de almacenamiento? En caso contrario ¿Qué otra opción recomendaría?	Aumentar la capacidad de almacenamiento, se considera como una propuesta de solución vertical. Sin embargo, no lo considero recomendable, porque sería demasiada inversión la compra de otro servidor para guardar información antigua y el aumento de un disco y procesador, necesarios para tener más capacidad y efectuar las consultas de forma rápida a la base de datos.
6	¿Presento algún problema por sobrecarga de información que afectara los procesos?	Sí, se presentaron problemas en donde el servidor no resistía la gran cantidad de consultas y tomaban demasiado tiempo para finalizarlas. Razón por la cual, se dividió la base de datos en una para consulta u otra para ingresar y actualizar datos. Sin embargo, no se presentaron incidencias en donde la base de datos se haya saturado.
7	¿La réplica de su base de datos afecta la capacidad de almacenamiento?	Es inevitable que las réplicas afecten el espacio de almacenamiento; sin embargo, no hemos presentado problemas que puedan afectar el registro de datos o saturar los servidores.
8	¿La plataforma web permite a los usuarios acceder a la información que ellos registraron? ¿Pueden acceder a los registros de otros usuarios? ¿Pueden editarla los datos ¿Qué otras facilidades brinda la plataforma para que los usuarios puedan acceder y disponer de la información?	Sí, el acceso a la plataforma web, presenta una interfaz amigable para el usuario. Y para acceder a la información de sus registros, disponen de un módulo de consultas para buscar el registro o ticket que deseen consultar. Se tienen diferentes tipos de usuarios, y cada uno tiene diferentes permisos. La mayoría son del tipo de usuario punto de venta y tienen permiso para visualizar los registros, y otros usuarios de tipo administrador tienen permiso para visualizar y modificar los registros y sus datos. Disponen de una bandeja de búsqueda para todos los usuarios, y una opción para extraer la información de la bandeja de ticket en un documento Excel, y por último un módulo de reporteria en donde se visualizan graficas que muestran la calidad de la atención.

### Ficha de entrevista

Datos básicos:

Cargo o puesto en que se desempeña	Analista y desarrollador
Nombres y apellidos	Erick Chavez
Código de la entrevista	Entrevistado 3 (Entv.3)
Fecha	09-05-2020
Lugar de la entrevista	Entrevista Virtual por Zoom

Nro.	Preguntas de la entrevista
1	¿Considera que las tablas se encuentran integradas correctamente?
2	Cuando envían el requerimiento de crear un nuevo campo en la plataforma web ¿Detallan las características del campo?
3	¿El contenido de su base de datos presenta replicas? ¿Dónde se almacena? En su opinión ¿Sus datos presenten inconsistencia?
4	¿La capacidad de almacenamiento actual es factible frente al constante aumento de registros? ¿Toman medidas frente a este hecho?
5	¿Recomendaría aumentar la capacidad de almacenamiento? En caso contrario ¿Qué otra opción recomendaría?
6	¿Presento algún problema por sobrecarga de información que afectara los procesos?
7	¿La réplica de su base de datos afecta la capacidad de almacenamiento?
8	¿La plataforma web permite a los usuarios acceder a la información que ellos registraron? ¿Pueden acceder a los registros de otros usuarios? ¿Pueden editarla los datos ¿Qué otras facilidades brinda la plataforma para que los usuarios puedan acceder y disponer de la información?

Observaciones

.....
.....
.....

### Entrevistado3 (Entv.3)

Nro.	Preguntas de la entrevista	Respuestas
1	¿Considera que las tablas se encuentran integradas correctamente?	El esquema que se maneja en el negocio, no tiene una estructura normalizada, sino relacional, básicamente estructurada por datos y relacionada por textos. La relación de datos es de acuerdo a la vinculación de información que se gestiona. La integridad que tenemos sobre esta data, es esencial, porque manejamos data de gran valor; razón por la cual manejamos estándares para almacenarla. Además, para salvaguardar los datos, manejamos estándares y normas internas. También nos aseguramos que la data se encuentre disponible y seguimos pautas técnicas para una correcta integración de datos. Opino que sería favorable que un experto nos asesore respecto a la mejora de la integridad, a la gestión de datos y como esquemizarla.
2	Cuando envían el requerimiento de crear un nuevo campo en la plataforma web ¿Detallan las características del campo?	Claro, todo requerimiento solicitado es detallado en el documento, sin embargo, si esto no ocurre, se programa reuniones con la persona que solicito el requerimiento para definir características y disipar cualquier duda.
3	¿El contenido de su base de datos presenta replicas? ¿Dónde se almacena? En su opinión ¿Sus datos presentan inconsistencia?	Actualmente manejamos 8 réplicas, sin embargo se redujeron porque inicialmente eran 16. Estas réplicas se encuentran sincronizadas por un balanceador, encargado de replicar cualquier cambio en la base de datos principal a las otras bases de datos de réplica. Sin embargo, en algunos casos se ha presentado una desincronización en algunas réplicas, siendo reflejadas a nivel de la aplicación; por ejemplo, en un servidor la data se encuentra actualizada y en otro servidor no ocurre la actualización. Son diferentes los motivos por los que ocurre. Pero somos conscientes de que hemos presentado algunos casos por desincronización, pero fueron atendidos inmediatamente.
4	¿La capacidad de almacenamiento actual es factible frente al constante aumento de registros? ¿Toman medidas frente a este hecho?	Actualmente la empresa posee 2 servidores físicos dedicados al almacenamiento de datos del negocio, a causa del incremento de volumen de datos, se está proyectando a incrementar el espacio. Se proyecta la compra de un servidor físico; sin embargo, esta compra debe ser justificada. Por ejemplo, si el servicio que se ofrece en lima, se quiere expandir a provincia, se proyecta mayor incremento de datos y por lo tanto se requiere mayor espacio de almacenamiento. Por otro lado, también se proyecta a no emplear servidores físicos, sino en almacenar los datos en la nube.
5	¿Recomendaría aumentar la capacidad de almacenamiento? En caso contrario ¿Qué otra opción recomendaría?	Si, se considera una necesidad aumentar la capacidad de almacenamiento frente al constante aumento de datos. Además, el aumento de volumen de los datos se debe a que se implementaron nuevas peticiones, requerimientos y otras cosas. Hasta el momento todo se encuentra bien, y lo normal sería un incremento de peticiones y canales de atención. Considero que la data histórica se debe guardar, y se incrementa la infraestructura para el almacenamiento. Es normal la compra de un store de almacenamiento, pero actualmente se está proyectando llevar toda la infraestructura de datos a la nube, porque puede beneficiar a la empresa mitigando gastos. Sin embargo, al no tener expertos, se está buscando soluciones para poder realizar esta implementación.
6	¿Presento algún problema por sobrecarga de información que afectara los procesos?	No hemos tenido una saturación a nivel de base de datos, pero se ha presentado problemas por saturación en transacciones a nivel de aplicativo. Por otro lado, a nivel de base de datos, no hemos presentado problema por saturación; como por ejemplo, no se puede acceder a la plataforma porque el disco está lleno.
7	¿La réplica de su base de datos afecta la capacidad de almacenamiento?	Cualquier efecto que pueda tener las réplicas en el espacio de almacenamiento, son previamente evaluadas. Por otro lado, es inevitable que la capacidad de un servidor físico se vea afectado por las réplicas; sin embargo, es importante que esto no impacte en la plataforma web, y evitar cualquier inconveniente durante su aplicación. Además, se tiene una persona responsable de revisar el espacio en los servidores.
8	¿La plataforma web permite a los usuarios acceder a la información que ellos registraron? ¿Pueden acceder a los registros de otros usuarios? ¿Pueden editarla los datos ¿Qué otras facilidades brinda la plataforma para que los usuarios puedan acceder y disponer de la información?	Si, la plataforma presenta módulos en donde el usuario puede buscar información de los registros realizados, tanto los propios como los de otros usuarios. Todos los usuarios tienen permiso para visualizar los registros, pero se tiene diferentes tipos de usuarios con diferentes permisos; tales como permisos para modificar y actualizar los datos del registro. Los usuarios con más privilegios, tienen acceso a la sección de informes, en donde pueden visualizar gráficos estadísticos, respecto a la atención brindada, y permiso para descargar la información, de los tickets registrados, en un documento Excel.

## Anexo 7: Pantallazos del Atlas. Ti

The screenshot displays the Atlas.ti software interface. The top menu bar includes 'Archivo', 'Inicio', 'Buscar', 'Analizar', 'Importar & Exportar', 'Herramientas & Soporte Técnico', 'Documento', 'Herramientas', and 'Vista'. Below the menu is a toolbar with various icons for document management and analysis. The left sidebar, titled 'Explorador del proyecto', shows a hierarchical view of the project structure, including 'Mixto\_Gonzales', 'Documentos (2)', 'Códigos (10)', 'Memos (0)', 'Redes (3)', and various groups of documents, codes, memos, and networks. The main workspace shows a document titled 'D 1: CAPÍTULO RESULTADOS MODELO 2020 \_ATLAS\_'. The document content is titled 'CAPÍTULO III: RESULTADOS' and contains the following text:

En el siguiente capítulo, presentaremos los resultados cuantitativos y cualitativos del trabajo de investigación.

**Resultados cuantitativos**

Lo primero que se analizará será el campo Número de Consulta, del registro que se obtiene luego de brindar la atención al punto de venta (en adelante: pdv) por medio de su plataforma web. Presentamos la Tabla 1, donde se observa que el campo Número de Consulta presenta Patrones de 8, 9 y 10 dígitos numéricos en sus datos para los meses de Enero, Febrero y Marzo del año 2019. Se contabiliza 30 000 registros de atención al pdv en cada mes, visualizando la frecuencia (f) y el porcentaje (%) de los patrones de los datos obtenidos del campo Número de Consulta. Siendo la frecuencia, el total de registros con datos que tienen los patrones indicados. Este campo permite revisar las incidencias registradas que presentan los puntos de venta durante la atención al cliente.

*Tabla 1.* Frecuencia (f) y porcentaje (%) de los Patrones del campo Número de Consulta de Enero.

On the right side of the document, there is a vertical toolbar with icons for 'Desvincular', 'Invertir dirección de vínculo', 'Relación de vínculo', 'Comentario', 'Nube de palabras', 'Lista de palabras', and 'Buscar en documento'. A small tooltip is visible over the 'Nube de palabras' icon, showing 'C.1.1.1. Integridad de datos' and 'C.1.1.2. Conformidad con datos'.

## Anexo 8: Fichas de validación de la propuesta



### Anexo 8. Ficha de validez de la propuesta

## CERTIFICADO DE VALIDEZ DE LA PROPUESTA

**Título de la investigación:** Perfilamiento y limpieza para optimizar la gestión de datos la plataforma web de una empresa de ventas, Lima 2020  
**Nombre de la propuesta:** Perfilamiento y limpieza de datos.

Yo, ...Walter Amador Chavez Alvarado..... identificado con DNI Nro .....09731774..... Especialista en...Ingeniería de Sistemas.....  
 Actualmente laboro en...Univ. Norbert Wiener.. Ubicado en...Av. Petit Thouars.....Procedo a revisar la correspondencia entre la categoría, sub categoría e ítem bajo los criterios:

**Pertinencia:** La propuesta es coherente entre el problema y la solución.

**Relevancia:** Lo planteado en la propuesta aporta a los objetivos.

**Construcción gramatical:** se entiende sin dificultad alguna los enunciados de la propuesta.

N°	INDICADORES DE EVALUACIÓN	Pertinencia		Relevancia		Construcción gramatical		Observaciones	Sugerencias
		SI	NO	SI	NO	SI	NO		
1	La propuesta se fundamenta en las ciencias administrativas/ Ingeniería.	X		X		X			
2	La propuesta está contextualizada a la realidad en estudio.	X		X		X			
3	La propuesta se sustenta en un diagnóstico previo.	X		X		X			
4	Se justifica la propuesta como base importante de la investigación holística- mixta -proyectiva	X		X		X			
5	La propuesta presenta objetivos claros, coherentes y posibles de alcanzar.	X		X		X			
6	La propuesta guarda relación con el diagnóstico y responde a la problemática	X		X		X			
7	La propuesta presenta estrategias, tácticas y KPI explícitos y transversales a los objetivos	X		X		X			

8	Dentro del plan de intervención existe un cronograma detallado y responsables de las diversas actividades	X		X		X			
9	La propuesta es factible y tiene viabilidad	X		X		X			
10	Es posible de aplicar la propuesta al contexto descrito	X		X		X			

Y después de la revisión opino que:

1. **Aprobar la propuesta**.....
2. ....
3. ....

Es todo cuanto informo;

Firma

## Anexo 9: Matrices de trabajo

### 1. Planteamiento del Problema

Problema de investigación a nivel internacional	<b>Informe mundial #1</b>	<b>Esencia del problema</b>	<b>Consolidación del problema</b>
	Riesgo en la gestión de datos.	<p>1. Actualmente, el 34% de organizaciones dirigen la gestión de datos como un riesgo competitivo, dado que si la información se maneja de forma incorrecta, puede afectar la reputación de la empresa con sus clientes y proveedores y el estado económico de la misma.</p> <p>2. Un aspecto importante a considerar, es que las organizaciones manejan información de sus colaboradores. Estos datos se obtienen de redes sociales, encuestas, sistemas de RH y finanzas. Por consiguiente, es inevitable el incremento de los datos al igual que los riesgos y las dudas en seguridad, transparencia y el uso correcto de la información con los permisos adecuados. De modo que, las organizaciones deben informar a los colaboradores, respecto al manejo que llevan en su gestión de datos, para no desatar preocupación.</p> <p>3. Es importante mencionar, que un porcentaje (54%) de las organizaciones en México incluyen analistas de Recursos Humano para la toma de decisiones; sin embargo, la diferencia (46%) no considera a los analistas como impulsores del negocio.</p>	<b>Internacional</b>
	<b>Título del informe</b>		<p>La presencia de datos o información incorrecta puede afectar su reputación, dañando la relación con los clientes y proveedores, y el sustento económico de la misma. Simultáneamente, la carencia en mejoras de gestión de datos, dificulta la obtención de información confiable para la toma de decisiones y mejora en las operaciones. Sobre todo, en las Instituciones de Educación Superior, que a pesar de ser organizaciones que genera conocimiento, no manejan una adecuada gestión y control de calidad de datos, dejando a los mismos propensos a encontrarse dispersos, duplicados, incompletos o en diferentes formatos. Por otra parte la ausencia de medidas y controles adecuados para manejar la gran cantidad de información disponible, puede abrumar al usuario que la maneja, perdiendo la oportunidad de obtener el conocimiento que los datos y la información pueden brindar. (Deloitte Consulting Group S.C., 2018; Osorio y otros, 2017; Álvarez y otros, 2016).</p>
	Tendencias en Capital Humano 2018 México		
	<b>Referencia</b>	Deloitte Consulting Group S.C. (2018). Tendencias en Capital Humano 2018 México. <i>Deloitte Insights</i> , 15-18.	
	<b>Informe mundial #2</b>	<b>Esencia del problema</b>	
	Confiabilidad en los datos. Gestión y control de datos.	<p>1. Actualmente, las organizaciones deben mejorar la gestión a sus datos para generar información más confiable. Puesto que, para la toma de decisiones y la mejora de las operaciones en una organización, se debe aprovechar la información que generan, utilizando fuentes de datos confiables, racionales y disponibles para su uso. Principalmente, debido al aumento de datos y las múltiples posibilidades de accesos a los mismos.</p> <p>2. Por otra parte, las Instituciones de Educación Superior (IES) como organizaciones que generan conocimiento; se espera que pongan en práctica, como otras organizaciones, la adecuada gestión, intercambio, flexibilidad y control de calidad de datos; no obstante estos datos son propensos a encontrarse dispersos, duplicados, incompletos o en diferentes formatos, dado que son procesados en múltiples aplicaciones. Esta condición, puede generar error e incongruencia en los reportes entregados a las comunidades, quienes consideran que la información que les proporcionan es legítima y confiable.</p>	
	<b>Título del informe</b>		
	La gobernabilidad de datos como apoyo en la gestión de datos de instituciones de educación superior.		
	<b>Referencia</b>	Osorio, M., Guerrero, C., & González-Zabala, M. (2017). La gobernabilidad de datos como apoyo en la gestión de datos de instituciones de educación superior. <i>Revista ESPACIOS</i> , 38(51), 11.	
<b>Informe mundial #3</b>	<b>Esencia del problema</b>		
Oportunidad de conocimiento	1. Debido a los retos empresariales y el aumento de la cantidad de datos		

		disponible, provenientes de fuentes diferentes y distinta naturaleza, el tratamiento que reciben los datos ha cambiado.	
	<b>Título del informe</b> ARQUITECTURA PARA LA GESTIÓN DE DATOS IMPERFECTOS EN LA ERA DE BIG DATA	2. Sin embargo, a pesar de los avances en las TIC (Tecnologías de la Información y Comunicación); la ausencia de medidas y controles adecuados, la gran cantidad de información disponible puede abrumar a los usuarios que la manejan. Por consiguiente, se pueden perder la oportunidad de obtener el conocimiento que puede brindar.	
	<b>Referencia</b>		
	Álvarez, K., Romero, B., Cadenas, J., Coronado, D., & Rodríguez, R. (2016). ARQUITECTURA PARA LA GESTIÓN DE DATOS IMPERFECTOS EN LA ERA DE BIG DATA. <i>ReVeCom</i> , 3(2), 47-56.		
<b>Problema de investigación a nivel nacional</b>	<b>Informe nacional #1</b>	<b>Esencia del problema</b>	<b>Consolidación del problema</b>
	Importancia y cambio en la gestión de datos.	1. La gestión de datos no es nueva en el campo de la comunicación ni en el de relaciones públicas, y con el incremento de las nuevas tecnologías y el entorno de comunicación actual, ha cambiado la forma en que se relacionan las organizaciones y su público que, cada vez más, producen y consumen contenidos en las redes y plataformas sociales; causando el crecimiento constante de datos y por consiguiente, los cambios en la gestión de datos.	<b>Nacional</b>
	<b>Título del informe</b>		En nuestro país, la gestión de datos está presente en diferentes empresas públicas y privadas, porque saben que los datos son importantes; sin embargo, reciben el valor que la empresa les brinda. Lamentablemente, muchas de ellas son indiferentes y no le dan el valor que corresponde, a pesar del cambio en la relación de las empresas y su público, debido a las nuevas tecnologías, la ausencia de buenas prácticas para la gestión de datos puede perjudicar la comunicación entre ambos. De la misma forma, la calidad y el contenido en los sistemas de información no cumplen el desempeño esperado; a causa de la falta de interés por fortalecer la información en los repositorios; en consecuencia, la posibilidad de explotar la enorme cantidad de datos para generar información que pueda ser utilizada para la toma de decisiones, se reduce. Además, para confirmar la seguridad y confiabilidad de un nuevo sistema informático, se debe verificar la información obtenida, para poder determinar la calidad de los datos (Pereira y otros, 2019; Curioso y otros, 2018; Vargas y otros, 2018).
	Big data y Relaciones Públicas. Una revisión bibliográfica del estado de la cuestión.	2. Por tal motivo, es necesario implementar nuevas prácticas en relaciones públicas, con las cuales se podrán gestionar los datos masivos y datos únicos de usuarios para lograr una comunicación más efectiva con ellos.	
	<b>Referencia</b>	3. Por otro lado, la inmensa cantidad de información generada, que puede ser negativa en muchas ocasiones y que pueda ocasionar una crisis comunicacional, incrementa el interés de los investigadores a la aplicación de gestión de datos masivos en comunicación de crisis.	
		Pereira Villazón, T., Portilla Manjón, I., & Rodríguez Salcedo, N. (2019). Big data y Relaciones Públicas. Una revisión bibliográfica del estado de la cuestión. <i>Revista de Comunicación</i> , 18(1), 151-165.	
	<b>Informe nacional #2</b>	<b>Esencia del problema</b>	
Explotación de Datos.		1. La atención primaria en salud permanecen con numerosos retos, de los cuales se van a resaltar sus procesos y su información. Esto se debe a la ausencia de organización en los niveles de atención y la demora para encontrar la información requerida y confiabilidad en la misma. Lamentablemente, para explotar al máximo la enorme cantidad de datos que maneja el Ministerio de Salud, es clave fortalecer el Repositorio Único Nacional de Información en Salud, aprobado mediante Resolución Ministerial N° 350-2015/MINSA; de lo contrario, las diversas estrategias como virtualización de datos, «data mining», algoritmos de predicción, etc. no podrán implementarse.	
<b>Título del informe</b>			
DESDE ALMA-ATA AL CIUDADANO DIGITAL: HACIA UNA ATENCIÓN PRIMARIA EN SALUD DIGITALIZADA EN EL PERÚ			
<b>Referencia</b>			

Curioso, W., Henríquez, M., & Espinoza, E. (2018). Desde Alma-Ata al ciudadano digital: hacia una atención primaria en salud digitalizada en el Perú. <i>Revista Peruana de Medicina Experimental y Salud Publica</i> , 35(4), 678-683.	
<b>Informe nacional #3</b>	<b>Esencia del problema</b>
Confiabilidad y calidad de datos.	<p>1. Debido a la falta de conocimiento en el manejo de información para la toma de decisiones, entre otros factores, los sistemas de información en salud, principalmente en lo que corresponde a calidad y uso de la información, no tienen el desempeño esperado. Además, la información registrada se encuentra dispersa y demora meses en consolidarse.</p> <p>2. Por ese motivo, se propuso el desarrollo e implementación del SINADEF o Sistema Informático Nacional de Defunciones, que permitirá a los médicos la elaboración del Certificado de Defunción o CEDEF en tiempo real y actualizando el sistema de Registro Civil y Estadísticas Vitales (RCEV), cubriendo toda la información necesaria de causas de difusión y mejorando la calidad de los datos.</p> <p>3. Además, el Ministerio de Salud junto a otras agencias gubernamentales busca asegurar la mejora en los sistemas de información de mortalidad y generar información para poder utilizarla en los procesos de toma de decisiones. Sin embargo, primero se debe consolidar la información obtenida hasta el momento para poder observar el registro completo de datos importante a considerar y la calidad de lo mismo. Para que la información sea empleada adecuadamente y con seguridad.</p>
<b>Título del informe</b>	
Resultados preliminares del fortalecimiento del sistema informático nacional de defunciones.	
<b>Referencia</b>	
Vargas, J., Pardo, K., Garro, G., Miki, J., Pérez, J., Valdez, W., . . . Cortez, J. (2018). Resultados preliminares del fortalecimiento del sistema informático nacional de defunciones. <i>Revista Peruana de Medicina Experimental y Salud Publica</i> , 35(3), 505-514.	

## 2. Problema y objetivo

Formulación del problema	Objetivo general
¿De qué manera se puede mejorar la gestión de datos de la plataforma web de una empresa de ventas?	Proponer la optimización de la gestión de datos de la plataforma web de una empresa de ventas, Lima 2020.
	<b>Objetivos específicos</b> Analizar la calidad de los datos en la plataforma web de una empresa de ventas. Diagnosticar la efectividad del almacenamiento de datos en la plataforma web de una empresa de ventas. Diagnosticar la efectividad de la recuperación de datos en la plataforma web de una empresa de ventas. Identificar si la calidad de los datos, el almacenamiento o la recuperación de los datos perjudica la eficiencia en la plataforma web de una empresa de ventas.

## 3. Justificación

<b>Justificación teórica</b>		
<b>Cuestiones</b>	<b>Respuesta</b>	<b>Redacción final</b>
¿Qué teorías sustentan la investigación?	En el presente trabajo se utiliza la Teoría de la Información y la Teoría de Decisión.	Para sustentar la investigación, se utilizarán las Teorías de la Información y de Decisión; puesto que ambas teorías aportan al trabajo de investigación. La primera teoría, nos indica que la información es un factor importante para la organización, puesto que si se emplea de forma correcta, puede proveer valiosa información del rendimiento de la misma. La segunda teoría expone, que en base a la información obtenida pueden ver opciones, no perceptibles hasta el momento, y tomar decisiones que pueden favorecer a la organización.
¿Cómo estas teorías aportan a su investigación?	La información es un factor importante para la organización, puesto que si se emplea de forma correcta, puede proveer valiosa información del rendimiento de la misma; además, en base a la información obtenida, pueden ver opciones, no perceptibles hasta el momento, y tomar decisiones que pueden favorecer a la organización.	
<b>Justificación práctica</b>		
¿Por qué hacer el trabajo de investigación?	El presente trabajo le permitirá a la empresa examinar e identificar anomalías que puede presentar los datos. Adicionalmente, les permitirá localizar el origen de los problemas de calidad en entradas de datos o interfaces incorrectas.	Permitirá a la empresa examinar e identificar anomalías que puede presentar los datos. Adicionalmente, permitirá localizar el origen de los problemas de calidad en entradas de datos o interfaces incorrectas. Con el aporte de la investigación, se espera que las anomalías identificadas en los datos sean corregidas, y cuando sean procesados, proporcionen información segura para un óptimo análisis del rendimiento de la organización y del servicio que ofrecen, o para la implementación de mejoras.
¿Qué espera con la investigación?	Al culminar la investigación se espera que las anomalías identificadas en los datos sean corregidas, y cuando sean procesados proporcionen información confiable, para un óptimo análisis de la organización y del servicio que ofrecen o la implementación de mejoras.	
<b>Justificación metodológica</b>		
¿Por qué investiga bajo ese diseño?	Implementación de la investigación holística es empleada porque se va a analizar los datos de las tablas principales, para identificar las anomalías y proponer las soluciones correspondientes acorde al negocio.	La implementación de la investigación holística, permitirá el análisis de los datos de las tablas principales, obtenido del modelamiento de datos del servicio que ofrece la empresa, logrando identificar las anomalías y proponer las soluciones correspondientes al negocio. El problema que presenta el trabajo de investigación, es la confiabilidad de datos por parte de la empresa, en su servicio de atención al punto de venta. El resultado de la investigación, permitirá a la empresa identificar las irregularidades en los datos, para que sean corregidos y luego utilizados con mayor confianza por la empresa y sus organizadores.
¿El resultado de la investigación permitirá resolver algún problema?	El principal problema en la investigación es la confiabilidad en datos por parte de la organización. El presente trabajo, permitirá identificar las irregularidades en los datos, para que sean corregidos y luego utilizados con mayor confianza por la organización.	

#### 4. Matriz de Teoría

<b>Teoría 1: Teoría de la Información</b>					
<b>Autor/es de la teoría</b>	<b>Año</b>	<b>Cita</b>	<b>Parfraseo (1)</b>	<b>Aplicación en su tesis (2)</b>	<b>Redacción final (1+2+3+4)</b>
Felicísimo Valbuena de la Fuente	1997	Valbuena (citado en Aladro, 2011), indico que “El objetivo fundamental de la Teoría de la Información es orientar y situar el conocimiento en torno a la comunicación, con una dirección concreta específica para investigar la información” (p. 83).	Valbuena (citado en Aladro, 2011) nos explica que “En la Teoría de la Información, su objetivo principal, se centra en encaminar el conocimiento en un entorno en donde sea compartido y comunicado, siendo direccionado a investigar la información adquirida del conocimiento” (Aladro , 2011).	<p>Esta teoría nos indica que la información obtenida debe ser investigada; debido a que, si la información no es confiable, el conocimiento obtenido en lugar de beneficiar a la empresa, puede ser perjudicial para la misma; sin embargo, si se analiza la información, y los datos que abarca, el conocimiento obtenido sería beneficioso para empresa.</p> <p>Aplicando esta teoría el presente trabajo de investigación, para un buen análisis de la información, es necesario que los datos que abarca sean confiables y de calidad. Por tal motivo, es necesario inspeccionar los datos e identificar las anomalías que pueda presentar, para que puedan ser corregidas y proporciona información segura y confiable.</p>	<p>La Teoría de la Información expuesta por Campbell, nos orienta a identificar a la información como un elemento necesario, para elaborar y estructurar propuestas de soluciones y mejoras. Asimismo, Valbuena respalda la importancia de la información, orientando a incursionar en la investigación de la misma, para que el conocimiento obtenido pueda ser encaminado a un entorno en donde sea compartido.</p> <p>En consecuencia, esta teoría le permitirá a mi investigación, lograr que la empresa entienda la importancia de la información e investigación de la misma, puesto que proporciona conocimiento variado, para aportar propuestas de solución o/y optimización en beneficio de la empresa. Pero para ello, la información que manejan requiere seguridad y confiabilidad, y eso se obtiene manejando datos de calidad. Lamentablemente, a pesar que las empresas gestionan sus datos, estas no suele estar en constante revisión y pueden presentar anomalías. Por tal motivo, es necesario que los datos sean analizados para identificar cualquier anomalía que pueda presentar.</p> <p>(Aladro , 2011; Monsalve, 2003)</p>
<b>Referencia:</b>	Aladro, E. (2011). La Teoría de la Información ante las nuevas tecnologías de la comunicación. <i>CIC. Cuadernos De Información Y Comunicación</i> , 16, 83 - 89. doi: <a href="https://doi.org/10.5209/rev_CIYC.2011.v16.4">https://doi.org/10.5209/rev_CIYC.2011.v16.4</a>				
<b>Autor/es de la teoría</b>	<b>Año</b>	<b>Cita</b>	<b>Parfraseo (3)</b>	<b>Aplicación en su tesis (4)</b>	
Jeremy Campbell	1989	Campbell explora la teoría de la información (citado en Monsalve, 2014) y la propone como la entrada hacia un terreno del conocimiento tan vasto como la naturaleza y tan complejo como la mente del hombre. Destaca de modo importante que la información debe comprenderse en el marco de fuerzas complementarias que permiten explicar el mundo, esto es, la entropía como agente del caos y la información como elemento no	Campbell (citado en Monsalve, 2014) muestra que en la Teoría de la Información, las fuerzas de la casualidad como las no accidentales, son universales. Explicado en la relación entre la entropía (medida de la información, para reducir incertidumbres) e información; que se relacionan de forma complementaria, como las fuerzas del azar y anti-azar.	Esta teoría será aplicada, para reducir la incertidumbre que pueda estar presente en la información de la empresa. Por tal motivo, se requiere de información confiable obtenida de la identificación de anomalías en los datos, para luego ser corregido. Por consiguiente, la información de calidad, puede aportar propuestas y soluciones para beneficio de la empresa.	

		casual “que explota la incertidumbre inherente al principio de la entropía para generar nuevas estructuras, para conformar al mundo de nuevas maneras” (p.147).	La información, elemento no casual, emplea la incertidumbre propia de la entropía para poder elaborar nuevas propuestas y soluciones que puedan ser beneficiosas (Monsalve, 2003).		
<b>Referencia:</b>	Monsalve, A. (2003). <i>Teoría de Información y Comunicación Social</i> (Primera ed.). Quito, Quito, Ecuador: Producciones Digitales Abya-Yala.				

<b>Teoría 2: Teoría de la Decisión</b>					
<b>Autor/es de la teoría</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo (1)</b>	<b>Aplicación en su tesis (2)</b>	<b>Redacción final (1+2+3+4)</b>
Gary Becker	1986	Becker nos explica (citado en Aguiar, 2004) que “La teoría de la decisión se ocupa de analizar cómo elige una persona aquella acción que, de entre un conjunto de acciones posibles, le conduce al mejor resultado dadas sus preferencias. Si debo invertir o no en bienes de equipo, qué carrera voy a estudiar, qué coche me compraré o, incluso, con quién debería casarme, son problemas muy comunes que nos afectan en nuestra vida cotidiana y a los que —en términos formales— se enfrenta la teoría de la decisión” (p. 139).	Becker declara (citado por Aguiar, 2004) que La Teoría de la Decisión, analiza la acción de una persona por elegir, entre un conjunto de opciones posibles, aquella que proporcione el mejor resultado posible conforme sus preferencias. Además, esta teoría afronta problemas muy frecuentes que impactan en la vida diaria (Aguiar, 2004).	Esta teoría será aplicada una vez que el conocimiento obtenido, luego de identificar y corregir la anomalías en los datos, proceda de información de calidad, segura y confiable; para la mejor toma de decisiones, que pueden llegar a ser mejoras o soluciones a problemas, en beneficio de la empresa.	La Teoría de la Decisión, respaldada por Becker, Bell, Raiffa, Tversky y Selten, sustenta el estudio de acciones y nuevas perspectivas de decisión, para seleccionar entre un conjunto de opciones, aquella que proporcione el mejor resultado según sus preferencias, y aplicarlas en diferentes entornos, que pueden ser: afrontando problemas frecuentes que impactan la vida diaria y entornos informativos en el cual un individuo se pueda desplegar. Por esa razón, la teoría aportará en la investigación, para la empresa y sus organizadores, la toma de decisiones con base a los conocimientos que poseen, logrando analizar sus problemas y reconocer el estado en el que se encuentran, para identificar soluciones y nuevas propuestas que se puedan implementar, seleccionando entre todas las opciones posibles, aquella que beneficie y cumpla los requerimientos de la empresa. Por tal motivo, primero se debe validar la confiabilidad de la información; de lo contrario, las decisiones con base a información, cuyos datos presentan errores o anomalías, puede perjudicar a la empresa (Aguiar, 2004).
<b>Referencia:</b>	Aguiar González, F. (2004). Teoría de la decisión e incertidumbre: modelos normativos y descriptivos. <i>EMPIRIA. Revista de metodología de ciencias sociales</i> , 0(8), 139-160.				
<b>Autor/es de la teoría</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo (3)</b>	<b>Aplicación en su tesis (4)</b>	
David E. Bell, Howard Raiffa, Amos Tversky (1988) Reinhard Selten (1996)	1988 1996	La teoría paramétrica de la decisión (o teoría de la decisión, sin más) que nos explican Bell, Raiffa, Tversky y Selten (citado en Aguiar, 2004) “aborda la naturaleza formal de las decisiones individuales y analiza criterios diversos de	La Teoría de la Decisión o Teoría Paramétrica de la Decisión, explicada por Bell, Raiffa, Tversky y Selten (citado por Aguiar, 2004) plantea que “las decisiones individuales, se pueden segmentar de	La teoría de decisión será aplicada, para orientar a las empresas y sus organizadores a buscar diferentes maneras para tomar las óptimas decisiones para beneficio de la empresa Por tal motivo, la información que se utilice, en la que se	

		decisión a tenor del contexto informativo en que se desenvuelva el individuo. Dicha naturaleza formal se puede tratar de manera normativa, prescriptiva o descriptiva” (p. 142).	maneras normativa, prescriptiva y descriptiva. Además, analiza nuevas perspectivas de decisión, para ser aplicadas en el entorno informativo en el cual el individuo se pueda desplegar” (Aguiar, 2004).	pueda desenvolver el usuario, debe ser segura y confiable.	
<b>Referencia:</b>	Aguiar González, F. (2004). Teoría de la decisión e incertidumbre: modelos normativos y descriptivos. <i>EMPIRIA. Revista de metodología de ciencias sociales</i> , 0(8), 139-160.				

<b>Teoría 3: Teoría de Calidad</b>					
<b>Autor/es de la teoría</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo (1)</b>	<b>Aplicación en su tesis (2)</b>	<b>Redacción final (1+2+3+4)</b>
Philip B. Crosby.	1980's	Cultura Preventiva. Indica que los errores se producen por falta de conocimiento, de atención o ambos. Sugiere un cambio de mentalidad para lograr el estándar cero defectos(CHACÓN & RUGEL, 2018)	Según la cultura preventiva, la falta de conocimiento y/o atención produce errores. Por lo cual, se recomienda cambiar la mentalidad para lograr un estándar de “cero defectos” (Chacón & Rugel, 2018)	Para mejorar la calidad es necesario conocer los datos que manejamos y asignarles la importancia debida. Por lo tanto, es necesario cambiar la mentalidad para entender la importancia de los datos con “cero defectos”.	Se puede considerar que la teoría de calidad total busca realizar un cambio organizacional, enfocado en mejorar la organización y corregir los errores producidos por la falta de conocimiento y/o atención. Por lo cual, se recomienda cambiar la mentalidad para poder lograr un estándar “cero defectos”. Además, se debe tener en cuenta, que cualquier cambio relacionado con calidad, afecta a los costos y colaboradores de la organización. (CHACÓN & RUGEL, 2018). Logrando comprender, que para aplicar una mejora en la calidad es necesario conocer los datos que manejamos y asignarles la importancia correspondiente. Para lo cual, se debe buscar la posibilidad de aplicar un cambio organizacional, y un cambio de mentalidad por parte de los colaboradores, para que entiendan la importancia de un estándar con “cero defectos”. Además, las mejoras en calidad se debe aplicar de forma correcta, de lo contrario puede afectar a la organización, impactando financieramente a la organización y sus colaboradores (Chacón & Rugel, 2018).
<b>Referencia:</b>	Chacón, J., & Rugel, S. (2018). Teorías, Modelos y Sistemas de Gestión de Calidad. <i>Revista ESPACIOS</i> , 39(50), 14.				
<b>Autor/es de la teoría</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo (3)</b>	<b>Aplicación en su tesis (4)</b>	
Sashkin	1992	Teoría de la Calidad Total. No se considera una herramienta de medición estadística, ni un proceso de control de calidad, sino una cultura de calidad y cambio organizacional (CHACÓN & RUGEL, 2018)	Se considera a la Teoría de Calidad Total como una cultura que busca efectuar el cambio organizacional enfocado a la mejora (Chacón & Rugel, 2018)	Aplicando la teoría de calidad total, es posible efectuar un cambio organizacional que pueda mejorar la calidad los diferentes en el negocio.	
<b>Referencia:</b>	Chacón, J., & Rugel, S. (2018). Teorías, Modelos y Sistemas de Gestión de Calidad. <i>Revista ESPACIOS</i> , 39(50), 14.				
<b>Autor/es de la teoría</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo (3)</b>	<b>Aplicación en su tesis (4)</b>	
Kaizen	1960's	Teoría de la Calidad Total de los Japoneses Kaizen (citado en Baird,	CHACÓN & RUGEL (2018) consideran que la Teoría de Calidad Total sugiere que los cambios	Los cambios relacionados con la calidad en una organización, pueden afectar de forma positiva o negativa,	

		2015) sugiere que la variación de la calidad tiene impacto directo en los costos y en la gente (CHACÓN & RUGEL, 2018)	relacionados con la calidad, pueden afectar los costos y a las personas (Chacón & Rugel, 2018)	dependiendo que como se aplique. Además, sus resultados impactan en sus costos y sus colaboradores.	
<b>Referencia:</b>	Chacón, J., & Rugel, S. (2018). Teorías, Modelos y Sistemas de Gestión de Calidad. <i>Revista ESPACIOS</i> , 39(50), 14.				

## 5. Marco Conceptual

Categoría problema 1: Gestión de Datos					
Autor/es	Año	Cita	Parfraseo	Aplicación en su tesis	Redacción final
Mariutsi Osorio Sanabria, Carlos Guerrero Alarcón y Mayda González-Zabala.	2017	Osorio, Guerrero y González (2017) explican a la gestión de datos como el “ejercicio que orienta las actividades relacionadas con las funciones de los datos, implica el desarrollo, ejecución y supervisión de planes, políticas, programas y prácticas que controlan, protegen, entregan y mejoran el valor de los datos y la información como activos de las organizaciones” (p. 11).	La gestión de datos dirige las tareas que se relacionen de forma directa o indirectamente con los datos, tales como desarrollar, ejecutar y supervisar la variedad de planes, programas, prácticas y políticas cuya función principal es la de controlar, proteger, entregar y mejorar la validez de los datos y la información, considerados como los activos más importantes en la organización (Osorio, Guerrero, & González, 2017).	Nos ayudara a identificar y entender las múltiples tareas que maneja el responsable de los datos. Los cuales, junto a la información, son los activos más importantes en la empresa.	Se puede definir a la gestión de datos como encargado de administrar y aprovechar los datos, que son empleados en procedimientos, procesos y proyectos, relacionados directa o indirectamente con los datos, que buscan mejorar, entregar, controlar y proteger a los mismos; debido que junto a la información, son activos de valor para el negocio. Asimismo, la gestión de datos se emplea como una herramienta, que beneficia a la empresa en la gestión de sus distintas áreas, generando valor y beneficio (Osorio, Guerrero, & González-Zabala, 2017; Portilla Romero, 2017; DAMA International, 2009).
<b>Referencia:</b>	Osorio, M., Guerrero, C., & González, M. (2017). La gobernabilidad de datos como apoyo en la gestión de datos de instituciones de educación superior. <i>Revista ESPACIOS</i> , 38(51), 11.				
Autor/es	Año	Cita	Parfraseo	Aplicación en su tesis	Redacción final
Juan Darío Portilla Romero	2017	Portilla (2017) nos explica que “la gestión de datos es una herramienta eficaz para el desarrollo de los componentes de un sistema de gestión en las organizaciones. Además, permite la disposición de las fuerzas productivas en busca de generar un mayor valor y la distribución de los beneficios que trae la	La gestión de datos es una herramienta que beneficia a las organizaciones, en el progreso de los elementos en su sistema de gestión. Adicionalmente, facilita la disponibilidad de las fuerzas productivas para la generación de mayor valor y la repartición de los beneficios obtenidos por la globalización.	Resalta su importancia, en la empresa, como una herramienta que posibilita el progreso.	La gestión de datos también es encargado de la construcción de información, empleando datos medibles para obtener información valiosa que pueda ser aprovechada en beneficio de la empresa. Asimismo, se considera un factor clave, debido al efecto que puede

		globalización para superar las múltiples desigualdades en cumplimiento del objeto social de Positiva Compañía de Seguros S. A., como única aseguradora estatal del mercado de personas”. (p.160)	(Portilla, 2017)		tener en la creación o gestión del conocimiento y el cambio de mentalidad respecto a las nuevas tecnologías, por medio de la gestión de la información (Escolar, 2002; Pereira Villazón, Portilla Manjón, & Rodríguez Salcedo, 2019).
<b>Referencia:</b>	Portilla, J. (2017). Gobierno de datos, un potenciador de los sistemas de gestión de calidad. <i>Signos</i> , 9(2), 159-172.				
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo</b>	<b>Aplicación en su tesis</b>	
DAMA International	2009	Según la Data Management Association (DAMA International, 2009) La Gestión de datos o <i>Data Management</i> (DM) es “la función de planificación de negocio, controlar y entregar dato y activos de información”. Esta función incluye: - Las disciplinas de desarrollo, ejecución y supervisión. - Planes, políticas, programas, proyectos, procesos, prácticas y procedimientos. - Controlar, proteger, entregar y mejorar. - El valor de los datos y activos de información. (p.23)	Se define a la gestión de datos como el encargado de la planificación del negocio, debido a que se encarga de controlar los datos y activos de información, validar que sean entregados y el valor que tienen en el negocio. Adicionalmente, está encargado de taras de control, protección, entrega y mejora. Sin olvidar su presencia en planes, proyectos, procesos, reglas o políticas del negocio y procedimiento (DAMA International, 2009)	Encargado de dirigir múltiples funciones, por ser responsable de los datos y activos informáticos.	
<b>Referencia:</b>	DAMA International. (2009). <i>The DAMA Guide to the Data Management Body of Knowledge (DAMA-DMBOK) Spanish Edition</i> (Primera ed.). Technics Publications, LLC.				
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo</b>	<b>Aplicación en su tesis</b>	
Cora Escolar	2002	Escolar (2002) explica que “el proceso de "gestión" de datos como tarea político-administrativa y de investigación supone el reconocimiento previo de un complejo proceso de construcción de la información. Desde esta perspectiva resulta fundamental entender que los datos no están "dados en la realidad" y que solo resta	La gestión de datos en un complejo procesos de construcción de la información, el cual requiere una evolución previa si se planea utilizar como misión político-administrativa y de investigación. Los datos no solo son registrados y almacenados, sino que son construidos y condicionadas por las perspectivas teórico-metodológicas. Unida a esta definición, “la gestión de información”	Comprender que sus tareas no solo se basan en aquellas que se relacionadas con los datos, sino que también se encarga de emplear a los mismo, para obtener información valiosa.	

		recopilarlos, sino que son fruto de una acción creadora y por tanto condicionada por las perspectivas teórico-metodológicas desde las cuales se los construye. Aunada a esta afirmación la "gestión de información" debe recurrir a datos que deben ser susceptibles de ser contados, medidos, pesados y para ello se apoya en una metodología cuantitativa que se basa en muy diversas fuentes" (p.196).	precisa recurrir a datos medibles; por esa razón, se apoya en la metodología cuantitativa basada en diversas fuentes (Escolar, 2002).		
<b>Referencia:</b>	Escolar, C. (2002). El Proceso de "Gestión de Datos". Construcción, medición y evaluación de los datos. (D. F. González, Ed.) <i>Cinta de Moebio</i> , 0(14), 196-199.				
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo</b>	<b>Aplicación en su tesis</b>	
Tatiana Pereira Villazón; Idoia Portilla Manjón; Natalia Rodríguez Salcedo.	2019	Según Pereira, Portilla y Rodríguez (2019) "en el campo de la comunicación interna podemos estudiar cómo la gestión de datos puede afectar en la creación o generación de conocimiento a partir de la administración de información de calidad, y como a través de ella se puede contribuir al éxito en el cambio de mentalidad de las organizaciones, hacia la transformación digital" (p.163).	Se define a la Gestión de Datos como un factor clave, debido a que por medio de la información, puede afectar la creación o gestión de conocimiento. Además, puede contribuir al cambio de mentalidad de la organización, respecto a las nuevas tecnologías, por medio de la administración de información de calidad (Pereira, Portilla, & Rodríguez, 2019).	Se le considera un factor clave, porque afectando la información, puede afectar la creación o gestión del conocimiento. Y promover el uso de nueva tecnologías.	
<b>Referencia:</b>	Pereira, T., Portilla, I. & Rodríguez, N. (2019). Big data y Relaciones Públicas. Una revisión bibliográfica del estado de la cuestión. <i>Revista de Comunicación</i> , 18(1), 151-165.				

<b>Sub categoría apriorísticas 1: Calidad de Datos</b>					
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo</b>	<b>Aplicación en su tesis</b>	<b>Redacción final</b>
Mariutsi Osorio Sanabria, Carlos Guerrero Alarcón y Mayda González-Zabala.	2017	Alvarado (citado en Osorio, Guerrero, & González, 2017) explica que "Es la percepción de los datos de acuerdo al propósito para el cual se utilizan en un contexto determinado y de cuan bien representan el mundo real" (Alvarado, 2011). Por otra parte, Batini y Scannapieco (citados en Osorio, Guerrero, & González, 2017) explican	Es identificar y entender los datos, para un propósito y contexto específicos, y su semejanza con la realidad. Es por esto que se requiere actividades para adquirir y estandarizar datos, identificar fuentes confiables, integrar y perfilar datos para identificar errores, y poder optimizar costos. (Osorio, Guerrero, & González, 2017)	Es necesario que los datos a emplear presenten su semejanza con la realidad; sin embargo, a causa de irregularidades, esto se dificulta, desconfiando de la veracidad de los datos. Por lo cual, la implementación de actividades para estandarizar los datos, asegurar la confiabilidad	La calidad de datos se puede considerar como una característica, que no solo cumple normas y requerimientos, sino también las expectativas del cliente. Por lo cual, es necesaria la semejanza de los datos con la realidad, para garantizar la

		que “Este proceso requiere contar con actividades para la adquisición y estandarización de datos, identificación de fuentes confiables, integración, perfilamiento, identificación de objetos, localización y corrección de errores, y optimización de costos”.		en la fuente, integrar y perfilar datos, permitirá identificar los errores, para luego corregirlos y obtener datos de calidad. .	confiabilidad en los mismos, al momento de brindar conocimiento y en la toma de decisiones. Además, para poder identificar errores que afectan el contenido de los datos; y por consiguiente, la efectividad y eficiencia de la organización, por lo cual se propone la estandarización de los datos, utilizar fuentes confiables e integrar y perfilar los datos, para luego ser tratados y mejorar la calidad en los mismos. (Fernández , 2008; Valverde, Marotta & Vallespir, 2012; Caro, Fuentes & Soto, 2013; Osorio, Guerrero & González, 2017)
<b>Referencia:</b>	Osorio, M., Guerrero, C., & González, M. (2017). La gobernabilidad de datos como apoyo en la gestión de datos de instituciones de educación superior. <i>Revista ESPACIOS</i> , 38(51), 11-12.				
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parfraseo</b>	<b>Aplicación en su tesis</b>	
Angélica Caro, Alejandra Fuentes y Maria Antonieta Soto	2013	La norma ISO/IEC 25012 (citado en Caro, Fuentes y Soto, 2013) define calidad de datos como “Grado en que las características de los datos satisfacen necesidades implícitas y establecidas cuando son usados en condiciones específicas”.	El grado asignado por el cumplimiento de necesidades o requerimientos por parte de las características que presentan los datos, al ser empleados en condiciones específicas (Caro, Fuentes, & Soto, 2013).	Se emplea para asegurar que los datos y sus características puedan cumplir con los requerimientos o necesidades, sin presentar defectos o imperfecciones que puedan afectar el contenido.	
<b>Referencia:</b>	Caro, A., Fuentes, A., & Soto, A. (2013). Desarrollando sistemas de información centrados en la calidad de datos. <i>Ingeniare. Revista chilena de ingeniería</i> , 21(1), 54-69. <a href="https://dx.doi.org/10.4067/S0718-33052013000100006">https://dx.doi.org/10.4067/S0718-33052013000100006</a>				
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parfraseo</b>	<b>Aplicación en su tesis</b>	
Carolina Valverde, Adriana Marotta y Diego Vallespir	2012	Valverde, Marotta y Vallespir (2012) explica que la “Calidad de Datos es un área de investigación en sí misma, en la cual se ha generado un gran volumen de trabajo (sobre todo en los últimos años) enfocado principalmente a: definir los distintos aspectos de la calidad de los datos, y proponer técnicas, métodos y metodologías para la medición y para el tratamiento de la calidad de los datos (p.2).	Valverde, Marotta y Vallespir (2012) consideran a la calidad de datos como un área de investigación, generando en los últimos años un gran volumen de trabajo. Está orientado a definir los distintos puntos que posee la calidad de datos, medir y tratar, por medio de técnicas, métodos y metodologías, la calidad de los datos (Valverde, Marotta, & Vallespir, 2012)	La calidad de datos es un área de investigación, orientado a la medición de la calidad de datos registrados, para luego determinar el tratamiento adecuado.	
<b>Referencia:</b>	Valverde, C., Marotta, A., & Vallespir, D. (2012). <i>Análisis de la Calidad de Datos en Experimentos en Ingeniería de Software</i> . Obtenido de SEDICI Repositorio Institucional de la Universidad Nacional de La Plata: <a href="http://sedici.unlp.edu.ar/handle/10915/23711">http://sedici.unlp.edu.ar/handle/10915/23711</a>				
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parfraseo</b>	<b>Aplicación en su tesis</b>	
Carolina Valverde, Adriana Marotta y Diego Vallespir	2012	Valverde, Marotta y Vallespir (2012) explica que en la calidad de datos, “los datos constituyen un recurso muy valioso para las organizaciones al ser utilizados principalmente para la toma de decisiones. La mala calidad de los mismos influye de manera significativa y	Para la toma de decisiones, los datos son considerados como un recurso muy valioso. Sin embargo, la mala calidad de los mismos, puede afectar la efectividad y eficiencia de la organización, y por consiguiente, de todo el negocio. Por lo cual,	Se reconoce el valor que tienen los datos en la organización, debido a que son utilizados para brindar más conocimiento referente al estado actual de la organización y en la toma de decisiones. Por lo cual, la	

		profunda en la efectividad y eficiencia de las organizaciones así como en todo el negocio. Cada día se hace más notoria la importancia y necesidad en distintos contextos de un nivel de calidad adecuado para los datos (p.3).	diariamente se reconoce el valor de la calidad en los daos en distintas contextos (Valverde, Marotta, & Vallespir, 2012)	presencia de irregularidades en la data puede afectar la eficacia y eficiencia de la organización.	
<b>Referencia:</b>	Valverde, C., Marotta, A., & Vallespir, D. (2012). <i>Análisis de la Calidad de Datos en Experimentos en Ingeniería de Software</i> . Obtenido de SEDICI Repositorio Institucional de la Universidad Nacional de La Plata: <a href="http://sedici.unlp.edu.ar/handle/10915/23711">http://sedici.unlp.edu.ar/handle/10915/23711</a>				
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parfraseo</b>	<b>Aplicación en su tesis</b>	
Wilmar Darío Fernández Gómez	2008	Fernández (2008) expone que “La calidad es una característica cualitativa de algo que supone su excelencia y su certeza, cumpliendo una serie de requerimientos o normas, pero el hecho de que sean los requisitos mínimos y que el producto los cumpla no significa que éste tenga calidad cuando las expectativas del usuario son otras, es decir, aun cuando un producto o servicio cumpla con los requisitos mínimos para ser usado, dependiendo del usuario y de las expectativas con que éste lo use, lo puede calificar como de baja, media o alta calidad. La calidad de un dato topográfico supone la intervención de procesos y recursos, que deben ser óptimos para que “el dato” sea fi el a la realidad, dependiendo del uso que se quiere dar. Es necesario en topografía reconocer que la calidad se refiere no solo a la precisión, entendida como el grado de acercamiento a la realidad, en este caso a la determinación de las coordenadas de cada punto, también involucra todos los procesos y procedimientos que intervienen en su obtención” (p. 9).	Fernández (2008) considera la calidad una característica cualitativa referente a la excelencia, de un producto o servicio, cumpliendo los requerimientos, las normas, y las expectativas del usuario que pueden calificar la calidad como alta, media y baja. Por lo cual, se debe priorizar la optimización de procesos y recursos para obtener los datos, para que sean precisos con la realidad. (Fernández , 2008)	Según Fernández (2008) la calidad es como una característica, que además de cumplir normas o requerimientos, debe cumplir las expectativas del usuario. Por lo cual, es necesario asegurar la precisión de los datos con la realidad para afianzar la confiabilidad en los mismos por parte del cliente; sin embargo, las inconsistencias en la data, causan la desconfianza en la misma, asegurando su bajo nivel de calidad.	
<b>Referencia:</b>	Fernández , W. (2008). Calidad de datos en levantamientos topográficos. <i>Revista de Topografía AZIMUT</i> , 2, 5-17. Obtenido de Revista de Topografía AZIMUT: <a href="https://revistas.udistrital.edu.co/index.php/azimut/article/view/4044">https://revistas.udistrital.edu.co/index.php/azimut/article/view/4044</a>				

<b>Sub categoría apriorísticas 2: Almacenamiento de Datos</b>					
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parfraseo</b>	<b>Aplicación en su tesis</b>	<b>Redacción final</b>

Salvador E. Vázquez-Moctezuma	2015	Vázquez (2015) explica que “el almacenamiento de datos puede verse desde dos perspectivas. La primera observación se puede hacer desde el punto de las estructuras de sistemas de almacenamiento con opciones como DAS (Direct Attached Storage o Almacenamiento de Conexión Directa), NAS (Network Attached Storage o Almacenamiento Conectado en Red), SAN (Storage Area Network o Red de Área de Almacenamiento) y sistemas de almacenamiento en la nube, que incluye capacidades de espacio en unidades de discos duros tradicionales y sólidos, así como la tecnología de la Memoria de Cambio de Fase (PCM: Phase Change Memory). La segunda visión se enfoca en la naturaleza de los datos en una perspectiva más cercana a la administración de datos; probablemente se pueda tener la capacidad de espacio a través de los sistemas distribuidos de nube, pero surgen inconvenientes relacionados con la consistencia, disponibilidad y tolerancia de partición de los datos; es decir, se trata de una perspectiva más cercana a la administración de datos” (p.2).	Se puede considerar dos perspectivas de almacenamiento de datos. En primer lugar, el almacenamiento de datos con opciones respecto a las estructuras de almacenamiento. Por ejemplo tenemos DAS (Direct Attached Storage o Almacenamiento de Conexión Directa), NAS (Network Attached Storage o Almacenamiento Conectado en Red), SAN (Storage Area Network o Red de Área de Almacenamiento) y sistemas de almacenamiento en la nube, a los cuales se les puede atribuir capacidades de espacio, así como la tecnología de la Memoria de Cambio de Fase (PCM: Phase Change Memory). En segundo lugar, nos enfocamos en los datos, desde un ambiente relacionado con la administración de datos, considerando entre los sistemas de almacenamiento, el que puede brindar mayor capacidad de espacio, de igual forma, en la consistencia, disponibilidad y tolerancia en la repartición de datos (Vázquez, Tecnologías de almacenamiento de información en el ambiente digital, 2015).	Vázquez (2015) explica que en el almacenamiento de datos se consideran dos perspectivas. En primer lugar, determinar el sistema de almacenamiento a emplear (DAS, NAS, SAN y Cloud); en segundo lugar, desde un ambiente relacionado con la administración de datos, que el sistema seleccionado pueda cumplir con las características requeridas por parte del usuario. Tomando la segunda referencia como nuestro enfoque, porque a pesar que la empresa maneje un sistema de almacenamiento, se debe asegurar que cumpla con los requerimientos de espacio, consistencia, disponibilidad y tolerancia para una correcta administración de datos.	En el almacenamiento de datos, la estructura de los sistemas de almacenamiento (DAS, NAS, SAN y Cloud) y las características necesarias para una eficiente administración de datos, evitando inconvenientes, son puntos importantes a considerar.  Por otro lado, con los avances tecnológicos y el aumento de información, se recomienda el diseño de sistemas de almacenamiento más complejos con la capacidad de guardar y administrar grandes volúmenes de datos, para que pueda estar disponible posteriormente y ser compartido entre diferentes usuarios en la empresa. (Vázquez, 2015; Instituto Nacional de Ciberseguridad, 2016)
<b>Referencia:</b>	Vázquez, S. (2015). Tecnologías de almacenamiento de información en el ambiente digital. <i>Revista e-Ciencias de la Información</i> , 5(2), 1-18. <a href="http://dx.doi.org/10.15517/eci.v5i2.19762">http://dx.doi.org/10.15517/eci.v5i2.19762</a>				
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo</b>	<b>Aplicación en su tesis</b>	
Salvador E. Vázquez-Moctezuma	2015	Vázquez (2015) expone que “en la era digital, el almacenamiento de datos surge desde mediados del siglo XX con la aparición de las computadoras y unidades de disco duros (HDD): los datos digitales han cambiado nuestra forma de almacenar, considerando que se necesita mayor capacidad de espacio y velocidad en los dispositivos de entrada y	En la era digital, con la aparición de computadoras y unidades de disco, los datos han cambiado la forma de almacenamiento. Para poder acceder y recuperar información, es necesaria mayor capacidad de espacio y velocidad en dispositivos de entrada y salida (Vázquez, Tecnologías de almacenamiento de información en el	En la actualidad, con los avances de la tecnología, el almacenamiento de datos a cambiando, requiriendo mayor velocidad en la entrada y salida de datos.	

		salida para acceder y recuperar información” (p.3).	ambiente digital, 2015).	
<b>Referencia:</b>	Vázquez, S. (2015). Tecnologías de almacenamiento de información en el ambiente digital. <i>Revista e-Ciencias de la Información</i> , 5(2), 1-18. <a href="http://dx.doi.org/10.15517/eci.v5i2.19762">http://dx.doi.org/10.15517/eci.v5i2.19762</a>			
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo</b>	<b>Aplicación en su tesis</b>
Salvador E. Vázquez-Moctezuma	2015	Vázquez (2015) expone de forma general, que “las tecnologías de almacenamiento de información van cambiando constantemente de modo que el profesional de la información debe conocer la variedad que existe para seleccionar la solución de acuerdo a sus necesidades y entorno. Entre más información se tenga, será necesario crear diseños de sistemas de almacenamiento más complejos” (p.15).	El constante cambio en las tecnologías de almacenamiento de información ha llevado al profesional de la información a investigar y conoce las diferentes sistemas de almacenamiento para seleccionar el más adecuada según sus necesidades y entorno; considerando que será necesario el diseño de sistemas más complejos a causa del aumento de información (Vázquez, Tecnologías de almacenamiento de información en el ambiente digital, 2015).	Se debe conoce los diferentes sistemas de almacenamiento para seleccionar el más adecuado. Sin embargo, se debe considerar el diseño de sistemas más complejos, capaces de administrar el gran volumen de datos registrados.
<b>Referencia:</b>	Vázquez, S. (2015). Tecnologías de almacenamiento de información en el ambiente digital. <i>Revista e-Ciencias de la Información</i> , 5(2), 1-18. <a href="http://dx.doi.org/10.15517/eci.v5i2.19762">http://dx.doi.org/10.15517/eci.v5i2.19762</a>			
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo</b>	<b>Aplicación en su tesis</b>
Instituto Nacional de Ciberseguridad	2016	Los Servidores de almacenamiento en red, son utilizados para poder disponer de un lugar común de trabajo donde almacenar el resultado de los trabajos individuales y poder compartir información entre los diferentes usuarios de la empresa se dispone de servidores de almacenamiento en red (p.11)	Se puede disponer de diferentes sistemas de almacenamiento de información, siendo uno de ellos el sistema de almacenamiento en red. Este sistema, permite disponer de un lugar común de trabajo para almacenar resultados de trabajos y para compartir información entre diferentes usuarios de la empresa (Instituto Nacional de Ciberseguridad, 2016).	Se puede disponer de diferentes sistemas de almacenamiento, que permitan disponer de un lugar común de trabajo, para administrar correctamente los, para ser compartidos entre los diferentes usuarios de la empresa.
<b>Referencia:</b>	Instituto Nacional de Ciberseguridad. (2016). <i>Almacenamiento seguro de la información. Una guía de aproximación para el empresario</i> . Obtenido de incibe: <a href="https://www.incibe.es/protege-tu-empresa/guias">https://www.incibe.es/protege-tu-empresa/guias</a>			
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo</b>	<b>Aplicación en su tesis</b>
Instituto Nacional de Ciberseguridad	2016	El objetivo básico de un dispositivo de almacenamiento es guardar información para que esté disponible posteriormente. Para ello todos los componentes físicos del dispositivo tienen que funcionar correctamente y el sistema operativo debe encontrar la información que ha sido almacenada de un modo ordenado (p.14).	Existen diferentes sistemas de almacenamiento que pueden ser empleados y que tiene como objetivo guardar la información para que posteriormente pueda estar disponible. Pero primero, se debe asegurar su correcto funcionamiento; de igual forma, debe encontrar la información que ha sido almacenada de un modo	El objetivo principal de los sistemas de almacenamiento es guardar información para que pueda estar disponible posteriormente; asegurando su correcto funcionamiento y administración, para encontrar de un modo ordenado la información almacenada.

			ordenado (Instituto Nacional de Ciberseguridad, 2016).		
<b>Referencia:</b>	Instituto Nacional de Ciberseguridad. (2016). <i>Almacenamiento seguro de la información. Una guía de aproximación para el empresario</i> . Obtenido de incibe: <a href="https://www.incibe.es/protege-tu-empresa/guias">https://www.incibe.es/protege-tu-empresa/guias</a>				

<b>Sub categoría apriorísticas 3: Recuperación de Información</b>					
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo</b>	<b>Aplicación en su tesis</b>	<b>Redacción final</b>
José Antonio Salvador Oliván y Rosario Arquero Avilés	2006	Salvador y Arquero (2006) explica que “La recuperación de información abarca los aspectos intelectuales de la descripción de información y su especificación para la búsqueda, y también cualquier sistema, técnica o máquina que se utilice para llevar a cabo la operación (p.15).	La recuperación de información, engloba temas de conocimiento referente a descripciones que puedan facilitar la búsqueda de datos; además, también considera cualquier herramienta que se pueda emplear para llevar cabo la operación”. (Salvador & Arquero, 2006)	En la recuperación de información, se requiere tener conocimiento de su contenido y emplear las herramientas necesarias para facilitar su búsqueda.	En la actualidad, la recuperación de la información obtiene un rol de gran valor, puesto que ofrece a las empresas facilidades para obtener datos con atributos de calidad, en el menor tiempo y debidamente organizados, para ejecutar tareas y determinar la eficiencia o el fracaso de las operaciones.
<b>Referencia:</b>	Salvador , J., & Arquero, R. (2006). Una aproximación al concepto de recuperación de información en el marco de la ciencia de la documentación. <i>Investigación bibliotecológica</i> , 20(41), 13-43. Recuperado de: <a href="http://www.scielo.org.mx/scielo.php?script=sci_arttext&amp;pid=S0187-358X2006000200002">http://www.scielo.org.mx/scielo.php?script=sci_arttext&amp;pid=S0187-358X2006000200002</a>				
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo</b>	<b>Aplicación en su tesis</b>	
Sodel Vázquez-Reyes, María de León-Sigg, Perla Velasco-Elizondo, Juan Villa-Cisneros y Sandra Briceño-Muro	2017	Vázquez, De León, Velasco, Villa y Briceño (2017) consideran “la recuperación de información como un importante campo de investigación y desarrollo de sistemas computacionales que deben: a) ser capaces de procesar rápidamente grandes colecciones de documentos; b) permitir operaciones flexibles de búsqueda, y c) permitir la clasificación de la información recuperada. En este sentido, el propósito final de los sistemas de recuperación de información es ofrecer mecanismos que permitan a las empresas adquirir, producir y transmitir, al menor costo, datos e información con los atributos de calidad, precisión y validez, que sean útiles para la toma de decisiones (p.86).	La recuperación de información proporciona la capacidad de procesar rápidamente grandes cantidades de información, realizar operaciones flexibles de búsqueda y clasificar la información en los sistemas computacionales; logrando su propósito, de ofrecer a las empresas mecanismos para adquirir, producir y transmitir, al menor costo, datos e información con atributos de calidad. (Vázquez, De León, Velasco, Villa, & Briceño, 2017)	La recuperación de información, ofrece a las empresas mecanismos para procesar, buscar y clasificar grandes cantidades de información, facilitando las tareas con la presencia de datos con atributos de calidad.	Dicho lo anterior, se debe brindar los accesos necesarios al personal encargado de manejar la información, para que los mismos puedan tener conocimiento de la información que manejan; además, la información almacenada debe estar organizada, facilitando la localización de datos relevantes con características específicas, de un almacén de información o base de datos, capaces de brindar solución a un problema. (Martínez, 2004; Salvador & Arquero, 2006; Bordignon & Tolosa, 2007; Vázquez, De León, Velasco, Villa & Briceño, 2017)
<b>Referencia:</b>	Vázquez, S., De León, M., Velasco, P., Villa, J., & Briceño, S. (2017). Recuperación de información a través del índice invertido en Be Intelligent. <i>RISTI - Revista Ibérica de Sistemas e Tecnologías de Informação</i> (21), 85-98. <a href="http://dx.doi.org/10.17013/risti.21.85-98">http://dx.doi.org/10.17013/risti.21.85-98</a>				
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo</b>	<b>Aplicación en su tesis</b>	
Sodel Vázquez-Reyes, María de	2017	Este proceso de recuperación de información implicaba tiempo de	La recuperación de información depende de varios factores: tiempo de	La recuperación de información, requiere del conocimiento del	

León-Sigg, Perla Velasco-Elizondo, Juan Villa-Cisneros y Sandra Briceño-Muro		búsqueda y de envío, y dependía de la habilidad del personal que hacía el filtrado de manera manual de los documentos almacenados, de su conocimiento de los documentos existentes y de la organización del material almacenado (Vázquez, De León, Velasco, Villa, & Briceño, 2017, p.87)	búsqueda y envío, habilidad y conocimiento del personal, y organización de material almacenado (Vázquez, De León, Velasco, Villa, & Briceño, 2017).	personal y la organización en información almacenada, puesto que son factores importantes para facilitar la búsqueda de datos, que pueden ser requeridos en otras actividades.	
<b>Referencia:</b>	Vázquez, S., De León, M., Velasco, P., Villa, J., & Briceño, S. (2017). Recuperación de información a través del índice invertido en Be Intelligent. <i>RISTI - Revista Ibérica de Sistemas e Tecnologías de Informação</i> (21), 85-98. <a href="http://dx.doi.org/10.17013/risti.21.85-98">http://dx.doi.org/10.17013/risti.21.85-98</a>				
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo</b>	<b>Aplicación en su tesis</b>	
Fernando Bordignon y Gabriel Tolosa Chacón	2007	Croft (citado en Bordignon y Tolosa, 2007) estima que la recuperación de información es «el conjunto de tareas mediante las cuales el usuario localiza y accede a los recursos de información que son pertinentes para la resolución del problema planteado. En estas tareas desempeñan un papel fundamental los lenguajes documentales, las técnicas de resumen, la descripción del objeto documental, etc.» (Bordignon & Tolosa, 2007, p.14).	Croft (citado en Bordignon y Tolosa, 2007) considera la recuperación de información como la agrupación de actividades que permiten al usuario localizar y acceder a la información necesaria para resolver problemas. (Bordignon & Tolosa, 2007)	La recuperación de información engloba actividades para que el usuario tenga facilidad para acceder y localizar información para emplearla en la resolución de problemas.	
<b>Referencia:</b>	Bordignon, F., & Tolosa, G. (2007). Recuperación de información: un área de investigación en crecimiento. <i>Ciencias de la Información</i> , 38(1-2), 13-24. Disponible en: <a href="https://www.redalyc.org/articulo.oa?id=181414865002">https://www.redalyc.org/articulo.oa?id=181414865002</a>				
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo</b>	<b>Aplicación en su tesis</b>	
Fernando Bordignon y Gabriel Tolosa Chacón	2007	La Recuperación de Información (RI) no es un área nueva, sino que se viene desarrollando desde finales de la década de 1950. Sin embargo, en la actualidad adquiere un rol más importante debido al valor que tiene la misma. Se puede plantear que disponer o no de la información justa en tiempo y forma puede resultar en el éxito o fracaso de una operación (Bordignon & Tolosa, 2007, p.14).	La Recuperación de Información (RI) es un área desarrollada, que en la actualidad adquiere un rol mayor debido a su importancia. Por consiguiente, tener disponible o no de la información de valor, organizada y a tiempo, puede determinar el éxito o fracaso de una operación. (Bordignon & Tolosa, 2007)	En la actualidad, la recuperación de información es de suma importancia para la empresa; puesto que disponer de los datos necesarios en el momento preciso y debidamente organizados, puede determinar que la operación se efectúe de forma eficiente o termine siendo un fracaso.	
<b>Referencia:</b>	Bordignon, F., & Tolosa, G. (2007). Recuperación de información: un área de investigación en crecimiento. <i>Ciencias de la Información</i> , 38(1-2), 13-24. Disponible en: <a href="https://www.redalyc.org/articulo.oa?id=181414865002">https://www.redalyc.org/articulo.oa?id=181414865002</a>				
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo</b>	<b>Aplicación en su tesis</b>	
Francisco Javier Martínez Méndez	2004	Meadow (citado en Martínez, 2004) piensa que la recuperación de la	Meadow (citado en Martínez, 2004) opina que la recuperación de	La recuperación de información facilita obtener información	

	información es “una disciplina que involucra la localización de una determinada información dentro de un almacén de información o base de datos”. Este autor, implícitamente, establece que la recuperación de información se encuentra asociada con el concepto de selectividad, ya que la información específica ha de extraerse siguiendo algún tipo de criterio discriminatorio (selectivo por tanto). Pérez-Carballo y Strzalkowski (citado en Martínez, 2004) redundan que “una típica tarea de la recuperación de información es traer documentos relevantes desde una gran archivo en respuesta a una pregunta formulada y ordenarlos de acuerdo con su relevancia” (Martínez, 2004, p.2).	información es selectiva, porque busca determinada información que presenta algún tipo de características específicas dentro de un almacén de información o base de datos. Pérez-Carballo y Strzalkowski (citado en Martínez, 2004) afirman que la recuperación de información cumple la tarea de obtener información relevante de un conglomerado de respuestas a una problema y puedan ser ordenados según su importancia. (Martínez, 2004).	relevante y ordenada, según su importancia, de un almacén de información o base de datos; además, la información presenta algún tipo de característica que brinda respuesta a un problema.	
<b>Referencia:</b>	Martínez, F. (2004). <i>Recuperación de información: modelos, sistemas y evaluación</i> (Cuarta Edición ed.). Murcia:KIOSKO JMC,2004. Recuperado de: <a href="http://libros.metabiblioteca.org/handle/001/227">http://libros.metabiblioteca.org/handle/001/227</a>			

Sub categoría Emergentes 1: Limpieza de Datos					
Autor/es	Año	Cita	Parfraseo	Aplicación en su tesis	Redacción final
Erhard Rahm y Hong Hai Do	2000	Rahm and Do (2000) explain that “Data cleaning, also called data cleansing or scrubbing, deals with detecting and removing errors and inconsistencies from data in order to improve the quality of data. Data quality problems are present in single data collections, such as files and databases, e.g., due to misspellings during data entry, missing information or other invalid data. When multiple data sources need to be integrated, e.g., in data warehouses, federated database systems or global web-based information systems, the need for data cleaning increases significantly. This is because the sources often contain redundant data in different representations. In order to provide access to accurate and consistent data, consolidation of different data	Rahm and Do (2000) declara que “La limpieza de datos o <i>Data cleaning</i> , también llamada <i>data cleansing</i> o <i>scrubbing</i> , tiene como objetivo mejorar la calidad de datos, tratando de identificando los errores e inconsistencias que pueda presentar, para luego eliminarlas. Los problemas pueden estar presentes en archivos o base de datos, conocidos como colección de datos individuales, que pueden presentar errores ortográficos, pérdida de información o datos inválidos. También en integración de múltiples fuentes de datos, tales como un <i>warehouses</i> , la necesidad de limpieza de datos aumenta significativamente. Porque las fuentes a menudo contienen datos redundantes en diferentes representaciones. Por tal motivo, es necesario consolidar los datos y eliminar la información duplicada para proporcionar datos precisos y	Se puede considerar necesaria una limpieza de datos porque los datos o información pueden presentar errores ortográficos, campos vacíos, datos inválidos o información duplicada. Además, su objetivo de la limpieza de datos, es mejorar la calidad de los datos, para lo cual busca identificar errores o inconsistencias que puedan estar presentes en archivos o base de datos para luego corregirlos y proporcionar datos precisos u constantes.	Se puede considerar que el objetivo de la limpieza de datos ( <i>Data cleaning</i> o <i>scrubbing</i> ) es la mejora de la calidad de los mismos. Razón por la cual, se busca identificar errores o inconsistencias, por ejemplo campos vacíos, datos inválidos y/o datos duplicados que puedan estar presentes en archivos o base de datos, para minimizar el impacto negativo que pueda afectar a la información o para luego ser corregidos y brindar datos precisos y consistentes. Además, para ser aplicada con éxito, se debe cumplir su proceso, que

		representations and elimination of duplicate information become necessary". (p.3)	constantes. (Rahm & Hai Do, 2000)		<p>incluye 3 etapas, identificar el error, realizar un diagnóstico y corregir anomalías. Por otro lado, la ausencia de una correcta limpieza de datos, puede afectar a otros datos o los resultados de un análisis que emplearon datos que contienen errores o anomalías (Rahm &amp; Hai Do, 2000; Van den Broeck, Argeseanu, Eeckels, &amp; Herbst, 2005; Osborne, 2013).</p>
<b>Referencia:</b>	Rahm, E., & Hai Do, H. (2000). Data Cleaning: Problems and Current Approaches. <i>IEEE Computer Society Technical Committee on Data Engineering</i> , 23(4), 3-13. <a href="http://dbs.uni-leipzig.de">http://dbs.uni-leipzig.de</a>				
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo</b>	<b>Aplicación en su tesis</b>	
Vijayshankar Raman and Joseph M. Hellerstein	2001	Data cleaning has three components: auditing data to find discrepancies, choosing transformations to fix these, and applying the transformations on the dataset (p. 381).	Se identifican tres componentes en la limpieza de datos. En primer lugar, identificar disconformidad en los datos; en segundo lugar, elegir transformación para corregirlos; y por último, aplicar transformaciones al conjunto de datos. (Raman & Hellerstein, 2001)	Se puede determinar que la limpieza de datos posee una secuencia de actividades para ser aplicada con éxito.	
<b>Referencia:</b>	Raman, V., & Hellerstein, J. (2001). Potter's Wheel: An Interactive Data Cleaning System. <i>The VLDB Journal</i> , 27(1), 381-390. <a href="https://dl.acm.org/doi/10.5555/645927.672045">https://dl.acm.org/doi/10.5555/645927.672045</a>				
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo</b>	<b>Aplicación en su tesis</b>	
Jan Van den Broeck, Solveig Argeseanu Cunningham, Roger Eeckels y Kobus Herbst	2005	Data cleaning intends to identify and correct these errors or at least to minimize their impact on study results (p.0966). We present data cleaning as a threestage process, involving repeated cycles of screening, diagnosing, and editing of suspected data abnormalities (p.0967)	La limpieza de datos tiene la intención de identificar y corregir los errores que pueda presentar la data o al menos lograr minimizar el impacto que pueda tener en los resultados. Por otro lado, la limpieza de datos se presenta como un proceso con tres etapas, en primer lugar, detectar ciclos repetitivos; en segundo lugar, diagnosticar; y por último, la edición de probables anomalías. (Van den Broeck, Argeseanu, Eeckels, & Herbst, 2005)	La limpieza de datos busca identificar y corregir los errores presentes en los datos, o minimizar el impacto negativo que pueda tener en la información. Además, presenta 3 etapas para ser efectuado con éxito. Las etapas son: identificar el error, realizar un diagnóstico y corregir anomalías.	
<b>Referencia:</b>	Van den Broeck, J., Argeseanu, S., Eeckels, R., & Herbst, K. (2005). Data Cleaning: Detecting, Diagnosing, and Editing Data Abnormalities. <i>PLoS Medicine</i> , 2(10), 0966 - 0970. <a href="https://doi.org/10.1371/journal.pmed.0020267">https://doi.org/10.1371/journal.pmed.0020267</a>				
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo</b>	<b>Aplicación en su tesis</b>	

Jason W. Osborne	2013	Data cleaning is critical to the validity of quantitative methods. Not only can problematic data points lead to violation of other assumptions but can lead to misestimation of parameters and effects without causing severe violation of assumptions (p.9).	La limpieza de datos es fundamental para validar los métodos cuantitativos. Los distintos puntos de datos que presentan problemas, pueden afectar a otros grupos; además, también pueden causar una estimación errónea de los parámetros y resultados. (Osborne, 2013)	La ausencia de una correcta limpieza de datos, puede tener consecuencias, afectando a otros grupos de datos o los resultados de un análisis que emplearon datos que contienen errores o anomalías.	
<b>Referencia:</b>	Osborne, J. (2013). <i>Best practices in data cleaning</i> . London, United Kingdom: SAGE Publications, Inc.				

<b>Sub categoría emergentes 2: Perfilamiento de Datos</b>					
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo</b>	<b>Aplicación en su tesis</b>	<b>Redacción final</b>
DAMA International	2009	Antes de realizar las mejoras en los datos, primero hay que ser capaz de distinguir entre los datos buenos y malos. El intento de calificar la calidad de datos es un proceso de análisis y descubrimiento. Un analista de datos no necesariamente será capaz de identificar todas las instancias de datos defectuosos. Sin embargo, la capacidad de documentar situaciones en las que los valores de datos parece que no son correctos proporciona un medio para comunicar estos casos con expertos en la materia, cuyo conocimiento del negocio puede confirmar la existencia de problemas de datos.	El primer paso para mejorar los datos, es identificar los datos válidos y dañados. Para lo cual, se requiere un proceso de análisis. Por otro lado, un análisis de datos puede no ser capaz de identificar todos los datos defectuosos. Sin embargo, la documentación de casos donde se presentan problemas, ayuda a expertos a confirmar la presencia de anomalías o errores en los datos (DAMA International, 2009).	El perfilamiento de datos es el primer paso para realizar una mejora en los datos. Para lo cual, se requiere un análisis de los datos y un documento donde se tenga registrado los casos que presentan problemas, para validar la presencia de errores y anomalías en los datos.	Se puede considerar al perfilamiento de datos, como el primer paso para realizar la mejora en los datos. Se encarga de cumplir dos funciones primordiales, en primer lugar obtener resultados estadísticos de cada columna. Por ejemplo la cantidad de valores nulos, los valores distintos dentro de la columna, tipo de dato o patrones más frecuentes de los datos en la columna, para luego analizar y evaluar la calidad de los datos. En segundo lugar, se encarga de examinar los valores de datos que pueden presentar dificultades para ser calculados, porque involucran múltiples columnas, y presentan inclusión, dependencias funcionales y de orden. Contribuyendo para validar la presencia de errores y anomalías en los datos. Además, se considerada como una actividad importante para el profesional de IT (DAMA International, 2009;
<b>Referencia:</b>	DAMA International. (2009). <i>The DAMA Guide to the Data Management Body of Knowledge (DAMA-DMBOK) Spanish Edition</i> (Primera ed.). Technics Publications, LLC.				
DAMA International	2009	El perfilamiento de datos es un conjunto de algoritmos utilizados para dos propósitos: - Análisis y evaluación estadística de la calidad de los valores de los datos dentro de un conjunto de datos. - Exploración de las relaciones que existen entre las colecciones de valor dentro ya través de los conjuntos de datos.	El perfilamiento de datos es empleado con dos propósitos, El primero es analizar y evaluar mediante estadísticas, la calidad de los valores de datos; el último es explorar la relación que existe entre los valores dentro y alrededor de los conjuntos de datos (DAMA International, 2009)	El perfilamiento de datos cumple dos funciones primordiales. Emplear las estadísticas para analizar y evaluar la calidad de los datos, y examinar los valor de los datos de uno o varios conjuntos de datos.	
<b>Referencia:</b>	DAMA International. (2009). <i>The DAMA Guide to the Data Management Body of Knowledge (DAMA-DMBOK) Spanish Edition</i>				

	(Primera ed.). Technics Publications, LLC.				Naumann, 2014; Abedjan, Golab, & Naumann, 2017)
Autor/es	Año	Cita	Parfraseo	Aplicación en su tesis	
Felix Naumann	2014	“El perfilamiento de datos es un proceso de examinar datos disponibles en un origen de datos existentes [...] y recolectar estadísticas e información sobre los datos.” El perfilamiento de datos es una actividad importante y frecuente para cualquier profesional de IT o investigador.	El perfilamiento de datos es un proceso encargado de examinar los datos para luego obtener estadísticas e información de los mismos. Es considerado una actividad importante y frecuente para el profesional de IT o un investigador (Naumann, 2014)	El perfilamiento de datos es considerado como una actividad importante para el profesional de IT o un investigador, porque el perfilamiento de datos es un proceso encargado de examinar los datos, y luego obtener estadísticas e información de los mismos.	
<b>Referencia:</b>	Naumann, F. (2014). Data Profiling Revisited. <i>ACM SIGMOD Record</i> , 42(4), 40-49. 10.1145/2590989.2590995.				
Autor/es	Año	Cita	Parfraseo	Aplicación en su tesis	
Abedjan, Ziawasch & Golab, Lukasz & Naumann, Felix.	2017	El perfilamiento de datos es un conjunto de actividades y procesos para determinar los metadatos de un conjunto de datos dado. Entre los resultados más simples están las estadísticas por columna, como el número de valores nulos y valores distintos en una columna, su tipo de datos, o los patrones más frecuentes de sus datos. Los metadatos que son más difíciles de calcular son los que involucran múltiples columnas, como inclusión, dependencias funcionales y de orden. Los casos de uso tradicionales de perfiles de datos incluyen la exploración de datos, limpieza de datos e integración de datos. Las estadísticas sobre datos también son útiles en la optimización de consultas.	El perfilamiento de datos es un conjunto de actividades y procesos para poder analizar los metadatos, que posee un conjunto de datos, y obtener resultados estadísticos por columna, indicando el número de valores nulos, valores distintos en una columna, el tipo de datos o patrones más frecuentes en los datos. Por otro lado, los metadatos más difíciles de calcular involucran múltiples columnas, como inclusión, dependencias funcionales y de orden. En los casos más comunes, donde emplean el perfilamiento de datos, comprende la exploración de datos, limpieza de datos e integración de datos. Además, se considera que para optimizar consultas, las estadísticas sobre los datos también son útiles (Abedjan, Golab, & Naumann, 2017).	El perfilamiento de datos nos permitirá analizar un conjunto de datos pertenecientes a una tabla para obtener resultados estadísticos de cada columna, tales como la cantidad de valores nulos, los valores distintos dentro de la columna, tipo de dato o patrones más frecuentes de los datos en la columna. Por otro lado, los datos que pueden presentar dificultades para ser calculados, porque involucran múltiples columnas, y presentan inclusión, dependencias funcionales y de orden. Por otro lado, el perfilamiento de datos, comúnmente, suele comprender la exploración, limpieza e integración de datos.	
<b>Referencia:</b>	Abedjan, Z., Golab, L., & Naumann, F. (2017). Data Profiling: A Tutorial. <i>Conference: the 2017 ACM International Conference</i> (págs. 1747-1751). Chicago Illinois USA : Association for Computing MachineryNew YorkNYUnited States. 10.1145/3035918.3054772				

## 6. Matriz de Antecedentes

Datos del antecedente 1: Internacional		
<b>Título</b>	Sistema informático para la gestión de datos del docente.	<b>Metodología</b>

<b>Autor</b>	Santana Espinosa, María; Muñoz Morejón, Madelayne; O'Farril Fernández, María; Martínez Delgado, Daisy; Martínez Noa, Mariela.	<b>Tipo</b>	
<b>Año</b>	2017	<b>Enfoque</b>	
<b>Objetivo</b>	El objetivo de este trabajo es implementar una herramienta para la gestión de información de los jefes de departamentos docentes de la Escuela Nacional de Salud Pública.	<b>Método</b>	
<b>Resultados</b>	Se obtuvo una base de datos relacional con una tabla principal que recoge la información del docente y 14 tablas más relacionadas entre sí. Los datos recogidos fueron: profesión, especialidad, categoría docente, categoría investigativa, grado científico, dominio de idioma, resultado de evaluación docente, condecoraciones y títulos honoríficos y datos personales. Posee formularios de entrada, un sistema de informes que permite el acceso de forma rápida a los datos. La fase de prueba arrojó criterios positivos y propuestas de mejoras.	<b>Población</b>	
		<b>Muestra</b>	
		<b>Técnicas</b>	
		<b>Instrumentos</b>	
<b>Conclusiones</b>	La herramienta permitirá a los jefes de departamentos llevar a cabo un flujo de información completo de cada profesor para la elaboración de una estrategia departamental evidenciada en la fase de prueba.	<b>Método de análisis de datos</b>	
<b>Redacción final al estilo artículo</b>	Santana, Muñoz, O'Farril, Martínez & Martínez (2017) nos explican que el sistema estadístico informatizado (SAGECU), empleado por el departamento de cuadros de la Escuela Nacional de Salud Pública (ENSAP) para recoger datos de docentes, no cumple los requerimientos solicitados por los jefes del departamento de docentes; debido a que no pueden disponer de un flujo de información completo de cada profesor; en consecuencia, impide el desarrollo de estrategias de crecimiento individual. Además, no poseen una base de datos estructurada, en la cual puedan registrar los datos del profesor, tanto personales como relacionados con su labor.		

Fuente: (Santana, Muñoz, O'farrill, Martínez, & Martínez, 2017)

<b>Datos del antecedente 2: Internacional</b>			
<b>Título</b>	Módulo de gestión de datos de biopsias para el registro de cáncer en el centro Oncológico de Holguín.	<b>Metodología</b>	
<b>Autor</b>	Cabrera Bruzón, Fidel Antonio	<b>Tipo</b>	
<b>Año</b>	2019	<b>Enfoque</b>	Cualitativo
<b>Objetivo</b>	Se propone como objetivo general desarrollar una herramienta informática que permita gestionar los datos de biopsias para el registro de cáncer del Hospital General Universitario Vladimir Ilich Lenin.		
<b>Resultados</b>	<p>Capítulo 1</p> <p>Mediante los métodos de la investigación empleados se detectaron las principales deficiencias del proceso de la gestión de datos de biopsias oncológicas del departamento de Anatomía Patológica del Hospital General Universitario Vladimir Ilich Lenin. Esto posibilitó el planteamiento de una solución.</p> <p>Luego de realizar los fundamentos teóricos de la investigación, se logró un análisis de las principales tecnologías y herramientas. Este sirvió de base para la selección de estas siendo utilizadas en el proceso de la gestión de datos de biopsias del departamento de Anatomía Patológica del Centro Oncológico de Holguín. Finalmente se seleccionó la metodología XP para el desarrollo de la solución propuesta. Esta brinda múltiples facilidades al equipo de trabajo y es la que mejor se adapta al escenario planteado.</p> <p>Capítulo 2</p> <p>En el presente capítulo se completaron las fases del ciclo de vida de la solución propuesta: las historias de usuario, el plan de iteraciones y el de las entregas, permitiendo separar el proceso en iteraciones para tener un mayor control sobre el</p>	<b>Método</b>	Empírico analítico.
		<b>Población</b>	Información sobre las biopsias de los pacientes con cáncer en el Centro Oncológico de Holguín
		<b>Muestra</b>	Los datos de biopsias de mama in situ, biopsias de mama infiltrantes y biopsias de próstata
		<b>Técnicas</b>	Observación

	desarrollo del producto. También se abordaron todos los temas referentes al diseño, la implementación y la estrategia de pruebas de la solución. La efectividad del desarrollo dirigido por pruebas y la aplicación de las pruebas de aceptación demuestran ser muy altas en el proceso de desarrollo de software. Ambas juegan un papel fundamental en el proceso de construcción de la herramienta con una metodología ágil. Para determinar la sostenibilidad del producto se analizaron las dimensiones administrativas, socio-humanista, ambiental y tecnológica. Luego de realizar el estudio se arriba a la conclusión de que el producto es sostenible.		
<b>Conclusiones</b>	La utilización del módulo web propuesto facilita el manejo de la información de las biopsias en el Centro Oncológico de Holguín, permitiendo un aumento de la eficiencia del trabajo de los especialistas del centro y con ello el bienestar de los pacientes. Luego de realizar la valoración de sostenibilidad en las dimensiones administrativa, socio-humanista, ambiental y tecnológica, se arriba a la conclusión de que el producto es sostenible. Por todo lo anterior, la herramienta informática propuesta para la gestión de datos de biopsias para el registro de cáncer en el Centro Oncológico de Holguín cumple con el objetivo trazado al inicio de esta investigación y contribuye con la solución de la problemática planteada.	<b>Instrumentos</b>	
		<b>Método de análisis de datos</b>	
<b>Redacción final al estilo Tesis</b>	Cabrera(2019) realizó la tesis titulada <i>Módulo de gestión de datos de biopsias para el registro de cáncer en el centro Oncológico de Holguín</i> ; mediante un enfoque cualitativo y aplicando el método empírico-analítico, y como muestra, utilizó los datos de biopsias de mama in situ, biopsias de mama infiltrantes y biopsias de próstata de los pacientes con cáncer en el Centro Oncológico de Holguín; detectando las deficiencias en los procesos de gestión de datos, a causa de la dificultad de organizar y controlar el gran volumen de datos de biopsias oncológicas, del departamento de Anatomía Patológica del Hospital General Universitario Vladimir Ilich Lenin, que se generan cada año en el Centro Oncológico de Holguín. Además, la información se encuentra almacenada de forma física, con riesgo de pérdida; por consiguiente, los datos registrados y con posibles errores, dificulta el trabajo de los especialistas del centro y el afecta bienestar de los pacientes.		
<b>Referencia (tesis)</b>	Cabrera, F. (2019). <i>Módulo de gestión de datos de biopsias para el registro de cáncer en el centro Oncológico de Holguín</i> . (Tesis de Grado). Holguín, Cuba: Universidad de Holguín.		

<b>Datos del antecedente 3: Internacional</b>			
<b>Título</b>	Diagnóstico de la calidad y el entendimiento de los datos para el análisis y toma de decisiones en las áreas de negocio de la empresa de Telecomunicaciones XYZW	<b>Metodología</b>	
<b>Autor</b>	Carlos González y Claudia Hernández	<b>Tipo</b>	
<b>Año</b>	2017	<b>Enfoque</b>	Cualitativo
<b>Objetivo</b>	Realizar un diagnóstico de la calidad y el entendimiento de los datos para el análisis y toma de decisiones en las áreas de negocio de una empresa de telecomunicaciones con el fin de disminuir en un 30% las inconsistencias de calidad en los datos y reducir los costos de re-procesos en la generación de reportes de información en el área de BI debido a los defectos de calidad identificados por las áreas de negocio de la compañía que usan la información.		
<b>Resultados</b>	En el ecosistema empresarial, se puede observar que el entorno en el cual las organizaciones tienen soportados los ámbitos de negocio, llamasen personas, procesos, tecnología, entre otras, están soportados por diferentes sistemas de información y de acuerdo a la información procesada surgen distintas estrategias que le permiten ubicarse en un mercado muy competitivo. Particularmente en el desarrollo de este proyecto en el sector de las telecomunicaciones, hemos descubierto que actualmente los grandes volúmenes de información y la forma en que se acceden a ellos se ha convertido en un problema. Dado que, no todas las organizaciones interiorizan que la información es un activo vital y que basado en ella las estrategias corporativas podrían ser mucho más efectivas, y es aquí donde la Inteligencia de Negocios surge como una pieza clave para actuar proactivamente en el momento de tomar mejores decisiones y de conseguir mejor conocimiento y control del negocio, estudiar tendencias de consumo, descubrir las preferencias de nuestros clientes, identificando fortalezas que nos diferencien	<b>Método</b>	Explicativa
		<b>Población</b>	Áreas del negocio.
		<b>Muestra</b>	Usuarios de Reportes de Información Solicitados al BI.
		<b>Técnicas</b>	Encuesta de satisfacción

	de la competencia.		
<b>Conclusiones</b>	Lo que logramos interiorizar al desarrollar este proyecto es que la gran mayoría de empresas y la misma en la cual se desarrolló el presente proyecto, no utilizan sistemas de inteligencia de negocios para gestionar los mismos, o si lo hacen la forma de acceder a la información con los procesos establecidos no son los adecuados, si bien, esto se produce por varios factores es vital la revisión de los procesos internos de consecución de la información, transformación y cargue de la misma en los diferentes sistemas de consumo de información. Sin embargo, entendemos que, con un adecuado flujo de procesos, cultura corporativa y métodos eficaces de procesamiento de la información, las herramientas que se utilicen serán beneficiosas para la gestión actual y futura de la organización, teniendo en cuenta las ventajas que nos brinda la inteligencia de negocios.	<b>Instrumentos</b>	Talonario de Preguntas
		<b>Método de análisis de datos</b>	-
<b>Redacción final al estilo Tesis</b>	González & Hernández (2017) realizo la tesis titulada <i>Diagnóstico de la calidad y el entendimiento de los datos para el análisis y toma de decisiones en las áreas de negocio de la empresa de telecomunicaciones XYZW</i> . La investigación se realizó mediante un enfoque cualitativo y aplicando el método explicativo; utilizando como muestra a los usuarios de las áreas del negocio de la empresa que solicitan reportes de información a BI. Se empleó la entrevista para determinar que la empresa presenta problemas en sus grandes volúmenes de información, la forma de acceder a los mismos y por consiguiente, pérdidas económicas. Además, el gran número de inconsistencias que presentan sus datos dañan la confiabilidad en los reportes de información; y por consiguiente, realizan gastos para el re-procesos de los mismos en el área de BI. Lamentablemente, muchas organizaciones no consideran a la información como un activo vital, a pesar de que su implementación en las estrategias corporativas podría ser beneficiosa; gracias a la Inteligencia de Negocios, que puede surgir durante la toma de decisiones, aumentando el nivel de conocimiento y control en el negocio.		
<b>Referencia (tesis)</b>	González, C & Hernández, C. (2017). <i>DIAGNÓSTICO DE LA CALIDAD Y EL ENTENDIMIENTO DE LOS DATOS PARA EL ANÁLISIS Y TOMA DE DECISIONES EN LAS ÁREAS DE NEGOCIO DE LA EMPRESA DE TELECOMUNICACIONES XYZW</i> . (Tesis de Especialidad). Bogotá, Colombia: Institución Universitaria Politécnico Grancolombiano.		

<b>Datos del antecedente 4: Internacional</b>			
<b>Título</b>	Gestión de la calidad de los datos en la empresa.	<b>Metodología</b>	
<b>Autor</b>	Alejandro Jiménez Martín	<b>Tipo</b>	-
<b>Año</b>	2017	<b>Enfoque</b>	Cualitativo
<b>Objetivo</b>	En cuanto a los objetivos que se buscan con la realización de este proyecto son los siguientes: <ul style="list-style-type: none"> <li>✓ Análisis de los problemas de gestión de calidad de datos en una empresa.</li> <li>✓ Establecimiento de una posible solución a la problemática planteada.</li> <li>✓ Análisis y evaluación de los resultados obtenidos.</li> <li>✓ Establecimiento de conclusiones a partir del desarrollo del proyecto y explicación de líneas futuras que ayuden a mejorar la calidad del dato y, por consiguiente, el beneficio a una empresa.</li> </ul>		
<b>Resultados</b>	Cada una de las pruebas enunciadas han sido probadas secuencialmente y todas ellas han cumplido con su objetivo. Por motivos de confidencialidad, los resultados obtenidos no pueden mostrarse, pero si se puede comentar que la implantación de la ETL ha ayudado en gran medida a la limpieza de datos inconsistentes y, consecuentemente, ha conseguido la integridad de los mismos.	<b>Método</b>	
		<b>Población</b>	Base de datos de una empresa del Sector Financiero.
		<b>Muestra</b>	Tablas de clientes.
		<b>Técnicas</b>	
<b>Conclusiones</b>	En conclusión, esta empresa dispone de datos inconsistentes sobre sus clientes lo que provoca que los ejecutivos de la misma no puedan realizar un análisis exhaustivo en busca de soluciones y, por consiguiente, tomar decisiones de negocio correctas que ayuden a la economía de la organización. Este hecho da lugar a problemas potencialmente peligrosos que ponen en riesgo a la entidad. Las conclusiones alcanzadas tras el análisis, el estudio de la problemática de la gestión de datos en las empresas y el	<b>Instrumentos</b>	
		<b>Método de análisis de datos</b>	Herramienta ETL

	<p>desarrollo de la solución planteada me han ayudado en gran medida a comprender que la implantación de la calidad de datos en una organización no es necesaria sino fundamental. Las ventajas que supone tener datos de calidad vienen ligadas a los beneficios económicos, de rendimiento, eficiencia, eficacia y toma de decisiones que aporta a la empresa. Destacar también que los procesos de extracción, transformación y carga son una parte imprescindible en cualquier proyecto de calidad de datos ya que son el motor encargado de filtrar, limpiar, actualizar, codificar y extraer datos que ofrezcan valor al empleado. Otro de los puntos importantes ha sido el establecimiento de un control de datos (Data Governance) ya que permite conocer el origen de los mismos de forma que ante cualquier incidencia o búsqueda, el tiempo empleado en encontrar dicho dato sea mínimo.</p> <p>En conclusión, cualquier empresa que presente un volumen de datos elevado, necesita un proceso de calidad de datos para poder seguir creciendo y ser competitiva frente a otras empresas del sector ya que, en caso de no hacerlo, posiblemente quede en un segundo plano, las pérdidas económicas sean sustanciales y llegue incluso al cierre de la misma.</p>		
<b>Redacción final al estilo Tesis</b>	<p>Jiménez (2017) realizó la tesis titulada <i>Gestión de la calidad de los datos en la empresa</i>. La investigación se realizó con un enfoque cualitativo y aplicando el método explicativo. Se utilizó como muestra la tabla de clientes de la base de datos de una empresa del sector financiero, para determinar que la problemática en la gestión de datos corresponden a inconsistencia en los datos de clientes, puesto que la misma presenta ausencia de limpieza e integridad de datos, impidiendo el análisis para la búsqueda de soluciones y tomando decisiones que puede perjudicar y poner en riesgo a la empresa. Se debe agregar que, el manejo de un elevado volumen de datos, ausente de calidad de datos puede perjudicar a la empresa con pérdidas económicas y el cierre de la misma.</p>		
<b>Referencia (tesis)</b>	<p>Jiménez, A. (2017). <i>Gestión de la calidad de los datos en la empresa</i>. (Tesis de Grado). Madrid, España: Universidad Politécnica de Madrid.</p>		

<b>Datos del antecedente 5: Nacional</b>			
<b>Título</b>	Dashboard para la medición de la confiabilidad de los datos del Datamart de una entidad financiera aplicando principios de calidad.	<b>Metodología</b>	
<b>Autor</b>	César Balaguer García y Sandro Arce Villanueva	<b>Tipo</b>	Experimental
<b>Año</b>	2019	<b>Enfoque</b>	Cualitativo
<b>Objetivo</b>	Establecer la confiabilidad de los datos del Datamart aplicando indicadores de calidad.		
<b>Resultados</b>	Los resultados obtenidos en la encuesta dan a conocer la necesidad que se presenta, no solo la entidad en estudio sino toda entidad financiera o no, de sentirse seguro y confiado con la información recibida para una correcta toma de decisiones, de la misma manera, se pudo apreciar que se conoce muy poco sobre estándares de calidad de datos, sistemas de medición, normas, principios de calidad de datos y de información, que puedan de alguna manera garantizar que la información utilizada para la toma de decisiones es confiable, es segura y obedece a la realidad.	<b>Método</b>	Analítico sintético
		<b>Población</b>	Banco BBVA
		<b>Muestra</b>	Área de finanzas y riesgo
		<b>Técnicas</b>	Encuesta
<b>Conclusiones</b>	Mediante este estudio hemos querido demostrar que los responsables de garantizar la calidad de la información, no están cumpliendo su rol a cabalidad, al igual que aquellas personas encargadas de generar estándares de confiabilidad de la información, la presente investigación es el punto de partida para la generación de un proyecto integral que pueda medir cuan confiables son los datos utilizados para la toma de decisiones, mediante la aplicación de normas estandarizadas y reconocidas, de tal manera que puedan certificar cada uno de los principios que se	<b>Instrumentos</b>	Cuestionario
		<b>Método de análisis de datos</b>	Análisis Grupal

	<p>exigen para poder garantizar la información, estableciendo umbrales de calidad por cada uno de los principios que requiera y/o necesite la entidad y estableciendo semaforizaciones a los datos que serán medidos.</p> <p>Cabe señalar, que el presente estudio se basa únicamente en datos e información que es utilizada para la toma de decisiones, mas no para información de trabajo o información propia de los procesos de negocio; el presente estudio genera así una línea de trabajo para la aplicación de un Dashboard donde los usuarios encargados de las tomas de decisiones y generación de estrategias (nivel táctico y nivel estratégico) puedan tener la confianza y la garantía que los datos que utilizan para generar información útil para sus organizaciones se encuentra dentro de los umbrales de aceptabilidad y, de esta manera sus objetivos, metas y estrategias son ajustadas a una realidad vigente.</p>		
<b>Redacción final al estilo Tesis</b>	Balaguer & Arce (2019) realizaron la tesis titulada <i>Dashboard para la medición de la confiabilidad de los datos del Datamart de una entidad financiera aplicando principios de calidad</i> . La investigación se realizó con un enfoque cualitativo y aplicando el método analítico sintético. Se utilizó como muestra el área de finanzas y riesgos del banco BBVA, empleando la encuesta para dar a conocer que, en entidades financieras o no, presentan desconfianza en la información que manejan a causa de, la baja seguridad y confiabilidad en sus datos, puesto que los responsables de garantizar la calidad de información y estándares de confiabilidad no cumplen su rol. Así mismo, la ausencia de conocimiento en estándares y principios de calidad de datos, normas, sistemas de medición, etc. es otro factor que garantiza que la información que utilizan no es segura, confiable y auténtica, y puede perjudicar a la entidad financiera, en caso de que se utilice en la toma de decisiones.		
<b>Referencia (tesis)</b>	Balaguer, C; Arce, S. (2019). <i>Dashboard para la medición de la confiabilidad de los datos del Datamart de una entidad financiera aplicando principios de calidad</i> . (Tesis de Bachiller). Lima, Perú: Universidad Científica del Sur		

<b>Datos del antecedente 6: Nacional</b>			
<b>Título</b>	Análisis de métodos y técnicas de Limpieza de Datos existentes y aplicación en un Sistema CRM para una institución educativa limeña.	<b>Metodología</b>	
<b>Autor</b>	Sandoval Linares, Angel	<b>Tipo</b>	
<b>Año</b>	2018	<b>Enfoque</b>	Cuantitativo
<b>Objetivo</b>	El objetivo general del presente proyecto de tesis es proponer una nueva técnica para la limpieza de datos usando una combinación de técnicas que permitan mejorar la calidad de datos de los alumnos en una institución educativa limeña.		
<b>Resultados</b>		<b>Método</b>	
		<b>Población</b>	Institución Educativa Limeña
		<b>Muestra</b>	Dato de Alumnos
		<b>Técnicas</b>	
<b>Conclusiones</b>	Los datos son el recurso más importante de toda organización y tener este recurso con una buena calidad es una labor que debería ser prioritaria de toda organización. El estado del arte evidencia que existen varias técnicas y métodos para la limpieza de datos; sin embargo, el número de estudios primarios encontrados, utilizando el método de Revisión Sistemática, refleja que no es un campo de estudio muy estudiado.	<b>Instrumentos</b>	
		<b>Método de análisis de datos</b>	
<b>Redacción final al estilo Tesis</b>	Sandoval (2018) realizó la tesis titulada <i>Análisis de métodos y técnicas de Limpieza de Datos existentes y aplicación en un Sistema CRM para una institución educativa limeña</i> . La investigación se realizó con un enfoque cuantitativo y aplicando el método de explicación. Se utilizó como muestra los datos de alumnos de una Institución Educativa Limeña; logrando determinar que requiere de una adecuada limpieza de datos, porque presentan un bajo nivel de calidad de datos e información. Además, debe ser prioridad el mantener una buena calidad en información y datos, puesto que son recursos valiosos para la organización; sin embargo, algunas empresas ignoran el impacto y las consecuencias de una baja calidad de datos, que puede afectar el desempeño, la relación con sus clientes, la toma de decisiones y los planes estrategias en una empresa.		

<b>Referencia (tesis)</b>	Sandoval, A. (2018). <i>Análisis de métodos y técnicas de Limpieza de Datos existentes y aplicación en un Sistema CRM para una institución educativa limeña</i> . (Tesis de Grado). Lima, Perú: Pontificia Universidad Católica del Perú
---------------------------	--

<b>Datos del antecedente 7: Nacional</b>			
<b>Título</b>	Implementación de un Sistema de Gestión de Calidad ISO 9001:2015 para el manejo de información en la empresa Corporación “EL DORADO”	<b>Metodología</b>	
<b>Autor</b>	Wilmer Chilón Gonzales	<b>Tipo</b>	
<b>Año</b>	2018	<b>Enfoque</b>	Mixto
<b>Objetivo</b>	<p>GENERAL</p> <p>Determinar la efectividad de la implantación de un sistema de gestión de calidad basado en la ISO 9001:2015 para el manejo de la información en la Empresa Corporación El Dorado.</p> <p>ESPECÍFICO</p> <ul style="list-style-type: none"> <li>• Medir la eficiencia del manejo de la información en la Empresa Corporación El Dorado antes de la implantación de un sistema de gestión de calidad.</li> <li>• Analizar la eficiencia del manejo de la información de la Empresa Corporación el Dorado con la implantación del sistema de calidad propuesto.</li> </ul>		
<b>Resultados</b>	<p>Se empezó abordar la problemática que atravesaba la empresa Corporación el Dorado una empresa Cajamarquina con una amplia gama de materiales de construcción y vidriería, la cual, a raíz del mal manejo de información en todos sus niveles, es que tenía constantes quejas, devoluciones de productos y pérdida de clientes. Estos productos y servicios fueron el motivo de la tesis de investigación, cuyo principal objetivo fue la satisfacción total de sus clientes, la cual buscó la implantación de un SISTEMA DE GESTION DE CALIDAD (SGC), basada en la ISO 9001: 2015, en post de mejora.</p> <p>Durante el proceso de implementación del sistema de gestión de calidad (SGC), tomamos como referencia a lo aplicado por Ramírez, C y Sánchez. M (2006), ya que se determinó un incremento de la productividad en COPRORACION EL DORADO de manera muy considerable, así mismo se redujeron el número de quejas, así como el número de equipos devueltos.</p> <p>De igual manera coincidimos con lo aportado por Font, O (2013) ya que gracias a la implantación de un sistema de gestión de calidad es que la gestión documentación mejoro de manera muy considerable, evitando duplicidad en atenciones por cliente, así mismo en la gestión documentaria interna.</p> <p>También acotamos lo que asevera López K. y Roa A. (2016), que gracias al desarrollo de un sistema de gestión de calidad, se mejora la calidad de servicio y por ende la atención al cliente, algo que ha sido muy significativo en la empresa CORPORACION EL DORADO, en la cual aumento el número de clientes satisfechos.</p>	<b>Método</b>	Deductivo y Analítico
		<b>Población</b>	Personal de la empresa Corporación el Dorado.
		<b>Muestra</b>	50 personas.(gerente, administradores y trabajadores de área productiva y operativa)
		<b>Técnicas</b>	Encuesta Observación.
		<b>Instrumentos</b>	Cuestionario
<b>Conclusiones</b>	<p>1. En la empresa corporación el dorado, la toma de decisiones está a cargo únicamente de la alta gerencia, mientras que los niveles intermedios y productivos se dedican a cumplir funciones; del mismo modo, se conoce que el 90% de los trabajadores manejaba poca información para realizar sus funciones y solo un 10% manifestó que la información era suficiente. Por otro lado, se conoce la calidad de la información que manejaban era deficiente, siendo el caso que la totalidad de los trabajadores afirmo estar disconforme al respecto, siendo los más afectados los encargados de operaciones y producción. Respecto a las acciones de control, el 55.3% manifestó que se realizaban pocas veces y el 38% manifestó</p>	<b>Método de análisis de datos</b>	Excel SPSS

	<p>que nunca las ponían en práctica. Por ende, a raíz de la implantación del SGC, la efectividad mejoró significativamente en un 60%, esto a su vez ayudo a las distintas áreas a fluir la información a todo nivel.</p> <p>2. Se implantó un sistema de control de calidad basado en la norma ISO 9001:2015 que permitió mejorar el manejo de la información, así mismo ayudó a reorganizar la empresa de acuerdo a sus necesidades por área y nivel organizacional, documentándose cada una de sus dependencias, obteniéndose resultados positivos en la reducción de costos y satisfacción del personal y los clientes, facilitando la producción y la entrega efectiva de los productos.</p> <p>3. Después de la implantación del sistema de gestión de calidad, la eficiencia del manejo de la información mejoró de manera muy considerable, al punto que la cantidad de la información es suficiente en un 68% para los trabajadores y mucha para el 32%.</p> <p>De igual forma, el 98.6% manifiesta estar conforme (30.8%) y totalmente conforme (67.8%) con la calidad de la información que se maneja. Las acciones de auditoria de la información se realizan siempre en un 68% de los casos y con frecuencia en 32% de ellos.</p>		
<b>Redacción final al estilo Tesis</b>	<p>Chilón (2018) realizó la tesis titulada <i>Implementación de un Sistema de Gestión de Calidad ISO 9001:2015 para el manejo de Información en la Empresa "Corporación El Dorado"</i>. La investigación se realizó con un enfoque mixto y aplicando el método deductivo y analítico. Se utilizó como muestra a 50 trabajadores del personal de la empresa Corporación el Dorado, empleando la encuesta y observación para recopilar datos y demostrar que la empresa, dedicada a la venta de materiales de construcción y vidriería, presenta un manejo deficiente de la información, razón por la cual presenta constantes quejas de sus clientes y sus trabajadores, causados por la insatisfacción al cliente y dando como resultado la devolución de equipos; además, los trabajadores de las áreas de operaciones y producción se ven afectado por la deficiente calidad y manejo de la información, impidiendo que puedan cumplir sus funciones. Así mismo, presentan obstáculos en el flujo de la información dentro de la empresa, y problemas en la gestión documentaria relacionados a duplicidad de datos externos e internos. Así, por ejemplo, la duplicidad en atención al cliente.</p>		
<b>Referencia (tesis)</b>	<p>Chilón, W. (2018). <i>Implementación de un Sistema de Gestión de Calidad ISO 9001:2015 para el manejo de Información en la Empresa "Corporación El Dorado"</i>. (Tesis de Maestría). Cajamarca, Perú Universidad Nacional de Cajamarca.</p>		

<b>Datos del antecedente 8: Nacional</b>			
<b>Título</b>	Optimización en el proceso de limpieza de datos para la construcción de un Datawarehouse usando herramientas Open Source.	<b>Metodología</b>	
<b>Autor</b>	Darío Dueñas Bustinza	<b>Tipo</b>	Proyectivo
<b>Año</b>	2018	<b>Enfoque</b>	Mixto
<b>Objetivo</b>	<p>OBJETIVO GENERAL</p> <p>Efectuar la optimización que posee la herramienta de limpieza de datos respecto a la inteligencia de negocios Pentaho, en la transformación de datos de un repositorio a otro.</p> <p>OBJETIVOS ESPECIFICOS</p> <p>Analizar las bases de datos, para encontrar homogenización respecto a la información a unificarse</p>		
<b>Resultados</b>	Como resultado se logró construir el almacén de datos empleando PDI (Pentaho Data Integration), SPoon como herramientas de la suite Pentaho que a pesar de no ser herramientas de limpieza de datos ofrecen la suficiente funcionalidad para ello.	<b>Método</b>	
		<b>Población</b>	Instituto de Sistemas Cusco de la Universidad Nacional San Antonio Abad del Cusco (UNSAAC)
		<b>Muestra</b>	Base de datos del instituto de informática de la

			universidad nacional San Antonio de Abad del cuzco.
		<b>Técnicas</b>	Observación
<b>Conclusiones</b>	<p>En la fase de limpieza da datos la suite de Inteligencia de Negocios Pentaho posee una herramienta llamada Pentaho Data Integration antes llamada Kettle, la cual posee varias herramientas que permiten realizar un fase de limpieza de datos adecuado estas son Spoon, Pan, Kitchen y Carte, en nuestra investigación se trabajó con las herramientas que provee Spoon con ello pudimos llevar a cabo un fase de limpieza de datos adecuado, lo cual es fundamental en la fase ETC, con ello se logró cargar la data a la Suite de BI de una manera adecuada sin tener muchos problemas en la carga de datos, las migraciones de los mismos fueron adecuadas tal como se puede ver en los anexos al presente trabajo.</p> <p>Se realizó un análisis dela base de datos del instituto de informática de la universidad nacional San Antonio de Abad del cuzco, en la cual se encontró varias omisiones, campos en blanco y otros que tuvo que corregirse para así poder homogenizar la data que se procedió a cargar obviamente la herramienta PDI nos brindó muchas ventajas a la hora de realizar esta operación, además se tuvo que realizar una limpieza da datos adecuada, luego de esto se logró homogenizar la información para que pueda ser analizada de una manera correcta.</p>	<b>Instrumentos</b>	Guías de Observación
		<b>Método de análisis de datos</b>	
<b>Redacción final al estilo Tesis</b>	<p>Dueñas (2018) realizó la tesis titulada <i>Optimización en el Proceso de Limpieza de Datos para la Construcción de un Datawarehouse usando Herramientas Open Source</i>. La investigación se realizó con un enfoque mixto. Se utilizó como muestra la Base de Datos del Instituto de Informática de la Universidad Nacional San Antonio de Abad del Cusco (UNSAAC), empleando la observación para confirmar la presencia de inconsistencias en la base de datos, causadas por campos en blanco, datos incorrectos, duplicados, datos con formatos diferentes, etc, y dañando la unificación y migración de los datos de un repositorio a otro, por la ausencia de una adecuada limpieza de datos.</p>		
<b>Referencia (tesis)</b>	<p>Dueñas, D. (2018). <i>Optimización en el Proceso de Limpieza de Datos para la Construcción de un Datawarehouse usando Herramientas Open Source</i>. (Tesis de Magister). Juliaca, Perú: Universidad Andina Néstor Cáceres Velásquez.</p>		

## 7. Matriz de Método

<b>Sintagma: Holístico</b>				
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parfraseo</b>	<b>Redacción final</b>
Jacqueline Hurtado de Barrera	2000	La investigación holística surge como una necesidad de proporcionar criterios de apertura y una metodología más completa y efectiva a las personas que realiza investigación en las diversas áreas del conocimiento. Es una propuesta que presenta la investigación como un proceso global, evolutivo, integrado, concatenado y organizado. La investigación holística trabaja los procesos que tienen que ver con la invención, con la formulación de propuestas novedosas, con la descripción y la clasificación, considera la creación de teorías y modelos, la indagación acerca del futuro, la aplicación práctica de soluciones, y la evaluación de proyectos, programas y acciones sociales, entre otras cosas (p.14).	La investigación Holística orienta al investigador a ampliar su percepción, direccionándola a una metodología más completa y efectiva. Considera a la investigación como un proceso global que puede evolucionar, integrar, concatenar y ser organizado. Además, también maneja los procesos relacionados con la fabricación de propuestas, creación de modelos, futuras propuestas, prototipos de soluciones, evaluar proyectos, etc. (Hurtado, 2000).	La investigación Holística orienta al investigador a una metodología más completa y efectiva, ampliando su percepción, manejando diferentes procesos orientados al crecimiento de la empresa.  La metodología holística proporciona a la investigación un enfoque global del problema que presentan en su gestión de datos, determinando los procesos o personas que se encuentran afectados, para luego proponer soluciones capaces de optimizar los

				datos, la información que proporcionan, los procesos en los que son empleado y la confianza por parte del personal asignado a su aplicación (Hurtado, 2000).
<b>Referencia:</b>	Hurtado, J. (2000). <i>Metodología de la Investigación Holística</i> (Tercera ed.). Caracas, Venezuela: Fundación Sypal.			

<b>Enfoque mixto</b>				
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo</b>	<b>Redacción final</b>
Roberto Hernández Sampieri, Carlos Fernández Collado y María del Pilar Baptista Lucio	2014	Chen, (citado por Hernández, Fernández y Baptista, 2014) los define como “la integración sistemática de los métodos cuantitativo y cualitativo en un solo estudio con el fin de obtener una “fotografía” más completa del fenómeno, y señala que éstos pueden ser conjuntados de tal manera que las aproximaciones cuantitativa y cualitativa conserven sus estructuras y procedimientos originales (“forma pura de los métodos mixtos”); o bien, que dichos métodos pueden ser adaptados, alterados o sintetizados para efectuar la investigación y lidiar con los costos del estudio (“forma modificada de los métodos mixtos”)(p. 534).	El enfoque mixto integra los métodos cualitativo y cuantitativos, clasificándolos en un solo estudio, capaz de brindar una visualización completa del fenómeno, y señala que las aproximaciones cualitativa y cuantitativas se pueden unificar de dos formas: en primer lugar, conservando sus estructuras y procedimientos originales (“forma pura de los métodos mixtos”); y en segundo lugar, que ambos métodos pueden ser adaptados, alterados o sintetizados para realizar la investigación y lidiar con los costos. (“forma modificada de los métodos mixtos”) (Hernández, Fernández, & Baptista, 2014)	El método mixto proporciona a la investigación una visualización más amplia de la misma. Logrando la unificación e implementación de los puntos positivos, correspondientes a los enfoques cuantitativos y cualitativos. Por consiguiente, con una visión subjetiva de la investigación, se puede determinar la confiabilidad en la gestión de datos por parte de los colaboradores; por otro lado, con una visión objetiva de la investigación, se puede obtener la solución al problema de estudio que viene afectando a la empresa y a sus colaboradores (Hernández, Fernández, & Baptista, 2014).
<b>Referencia:</b>	Hernández, R., Fernández, C., & Baptista, P. (2014). <i>Metodología de la Investigación</i> (Sexta ed.). Ciudad de México: McGRAW-HILL / INTERAMERICANA EDITORES, S.A. DE C.V.			

Tipo proyectiva				
Autor/es	Año	Cita	Parafraseo	Redacción final
Jacqueline Hurtado de Barrera	2000	<p>Hurtado (2000) explica que “La investigación proyectiva hace explícitos escenarios alternativos de futuros posibles (estado predictivo), permite identificar riesgos y oportunidades de ciertas situaciones futuras (estado predictivo). Proporciona orientaciones para la acción (estado proyectivo), y establece criterios de decisión para alcanzar el mejor futuro posible (estado proyectivo).</p> <p>También involucra creación, diseño, elaboración de planes, o de proyectos. Para que un proyecto se considere investigación proyectiva, la propuesta debe estar fundamentada en un proceso sistemático de búsqueda e indagación que recorre los estados descriptivo, comparativo, analítico, explicativo y predictivo de la espiral logística” (p.328).</p>	<p>La investigación proyectiva está orientada a posibles escenarios futuros, identificando riesgos y oportunidades, acciones a realizar y criterios de decisión. También incluye elaboración de planes o proyectos. Por lo tanto, para considerar un proyecto como investigación proyectiva, de antemano debe haber realizado la investigación previa, recorriendo los estados de la espiral logística (descripción, comparación, análisis, explicación y predicción) (Hurtado, 2000).</p>	<p>En la investigación proyectiva, es fundamental haber realizado la indagación previa por los estados de la espiral logística (descripción, comparación, análisis, explicación y predicción), dado que está orientada a plantear escenarios futuros, para identificar oportunidades y definir acciones. (Hurtado, 2000) Por tal motivo, su aplicación en la investigación nos permitirá otorgar propuestas de solución, para optimizar la calidad y confiabilidad en la gestión de datos en la empresa, con previa búsqueda de fundamentos para la propuesta.</p>
<b>Referencia:</b>	Hurtado, J. (2000). <i>Metodología de la Investigación Holística</i> (Tercera ed.). Caracas, Venezuela: Fundación Sypal.			
Método: Deductivo				
Autor/es	Año	Cita	Parafraseo	Redacción final
César Augusto Bernal Torres	2010	<p>Según Bernal (2010) “Este método de razonamiento consiste en tomar conclusiones generales para obtener explicaciones particulares. El método se inicia con el análisis de los postulados, teoremas, leyes, principios, etcétera, de aplicación universal y de comprobada validez, para aplicarlos a soluciones o hechos particulares” (p. 59).</p>	<p>El método deductivo es un razonamiento que parte de conclusiones generales a explicaciones particulares y válidas. Así, por ejemplo, si una conclusión o principio general es verdadera, entonces las conclusiones particulares también lo son (Bernal, 2010).</p>	<p>Ambos métodos, el deductivo y el inductivo, serán empleados en la investigación, aportando de distintas formas. La metodología inductiva proporcionando a la investigación nueva información obtenida del análisis de los datos. Cumpliendo su principio, partiendo de lo particular a lo general. De igual manera, la metodología deductiva permite la búsqueda de situaciones o hechos reales similares al problema por el que pasa la empresa, para aplicar sus teorías, principio, leyes, etc. para ser adaptados en soluciones o situaciones particulares (Bernal, 2010).</p>
<b>Referencia:</b>	Bernal, C. (2010). <i>Metodología de la investigación administración, economía, humanidades</i> (Tercera ed.). Bogotá, Colombia: Pearson Educación de Colombia Ltda.			
Método: Inductivo				
Autor/es	Año	Cita	Parafraseo	Redacción final
César Augusto Bernal Torres	2010	<p>Según Bernal (2010) “Este método utiliza el razonamiento para obtener conclusiones que parten de hechos particulares aceptados como válidos, para llegar a conclusiones cuya aplicación sea de carácter general. El método se inicia con un estudio individual de los</p>	<p>El método inductivo es un razonamiento que parte del estudio individual a formular conclusiones generales o universales que pueden ser empleadas como principios.</p>	

		hechos y se formulan conclusiones universales que se postulan como leyes, principios o fundamentos de una teoría (p. 60)	Así, por ejemplo, con situaciones o datos particulares se obtienen conclusiones generales (Bernal, 2010).	
<b>Referencia:</b>	Bernal, C. (2010). <i>Metodología de la investigación administración, economía, humanidades</i> (Tercera ed.). Bogotá, Colombia: Pearson Educación de Colombia Ltda.			

## 8. Población, muestra y unidades informantes

<b>Población</b>					
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo</b>	<b>Aplicación en su tesis</b>	<b>Redacción final</b>
César Augusto Bernal Torres	2010	De acuerdo con Fracica (citada en Bernal, 2010), la población es “el conjunto de todos los elementos a los cuales se refiere la investigación. Se puede definir también como el conjunto de todas las unidades de muestreo” (p. 36). Según Jany (citada en Bernal, 2010) la población es “la totalidad de elementos o individuos que tienen ciertas características similares y sobre las cuales se desea hacer inferencia” (p. 48).	Fracica (citada en Bernal, 2010) define a la población como “el conjunto de conjunto de elementos o unidades de muestreo a los cuales está enfocada la investigación”. Además, Jany (citada en Bernal, 2010) explica que la población es “la agrupación total de elementos o individuos con ciertas características semejantes, requeridos para obtener conclusiones de la investigación” (Bernal, 2010).	El contexto anterior define el grupo que será empleado para el trabajo de investigación, con el propósito de obtener información relacionada a un problema. Por lo cual, se asignó a los colaboradores, que pertenecen al área de Tecnología y Transformación Digital, dentro del área de sistemas, en la empresa encargada de ventas como población; además; ellos serán sujetos para la entrevista y aportaran los datos para obtener información valida de una fuente confiable.	La población es el conjunto o agrupación de elementos o individuos que presentan características semejantes a las cuales está enfocada la investigación. (Bernal, 2010) Dicho lo anterior, se puede determinar, que el conjunto o agrupación asignado en el trabajo, es la fuente requerida que posee las características necesarias para la investigación y para proporcionar una respuesta a la problemática principal en la investigación; por lo cual, se asignó como la población, a los colaboradores, que pertenecen al área de Tecnología y Transformación Digital, dentro del área de sistemas, de la empresa encargada de ventas; además; ellos serán sujetos para la entrevista y aportaran los datos para obtener información valida de una fuente confiable.
<b>Número de colaboradores:</b>	Área de Sistemas				
<b>Referencia:</b>	Bernal, C. (2010). <i>Metodología de la investigación administración, economía, humanidades</i> (Tercera ed.). Bogotá, Colombia: Pearson Educación de Colombia Ltda.				
<b>Muestra</b>					
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parafraseo</b>	<b>Aplicación en su tesis</b>	
César Augusto Bernal Torres	2010	Bernal (2010) explica que “Es la parte de la población que se selecciona, de la cual realmente se obtiene la información para el desarrollo del estudio y sobre la cual se efectuarán la medición y la observación de las variables	Un grupo de personas u objetos que forman parte de la población, y que brindaran información a la investigación; efectuando la medición y observación en base a los objetivos	La muestra seleccionada en el presente trabajo de investigación serán tres colaboradores confirmados de la empresa, encargados de la base de datos, el diseño de la plataforma web y cualquier	Por otra parte, la muestra es la agrupación de personas u objetos que forman parte de la población, siendo una cantidad determinada, que brinda información a la investigación; efectuando la medición y observación en base a los objetivos planteados. (Bernal, 2010)  Por consiguiente, en el presente trabajo de investigación se seleccionó como muestra a tres colaboradores confirmados de la empresa, encargados de la base de datos, el diseño de la

		objeto de estudio” (p. 161).	planteados. (Bernal, 2010)	cambio necesario a la misma. También se puede contar con los datos registrados en la base de datos de la plataforma web, con datos verídicos, del servicio de atención al punto de venta.	plataforma web y cualquier cambio necesario a la misma. De igual modo, se ha tomado como muestra, los registros de la atención a los puntos de venta.	
<b>Número de colaboradores:</b>	3 colaboradores					
<b>Referencia:</b>	Bernal, C. (2010). <i>Metodología de la investigación administración, economía, humanidades</i> (Tercera ed.). Bogotá, Colombia: Pearson Educación de Colombia Ltda.					
<b>Unidades informantes</b>						
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parfraseo</b>	<b>Descripción de cada unidad informante</b>	<b>Redacción final</b>	
Jacqueline Hurtado de Barrera	2000	Hurtado (2000) indica que “Las unidades de estudio se deben definir de tal modo que a través de ellas se pueda dar respuestas completas, y no parcial o desviada, a la interrogante de la investigación; por ello debe incluir a todos los involucrados en los eventos de estudio” (p.152).	Las unidades de estudio definen de forma práctica y sencilla, las características en común que presentan la muestra de estudio, para poder brindar una respuesta concreta al problema planteado en la investigación. (Hurtado, 2000)	a) Registro de atención brindada al punto de venta de la base de datos. b) Documento en Excel del Registro de atención brindada al punto de venta. c) Analista de Operaciones. d) Analista Programador. e) Programador	La unidad de estudio es la definición práctica y sencilla de las características que tengan en común la muestra de estudio, para poder brindar una respuesta más concreta y minuciosa al problema planteado en la investigación. Se detalla que en el presente trabajo de investigación se cuenta con tres colaboradores que son un Programador, el Analista Programador y el Analista de Operaciones y brindarán información necesaria respecto a la empresa y el proyecto en el que se encuentran. Por otro lado, también contaremos con dos objeto como una de las unidades informantes, que son el registro de atención brindada al punto de venta de la base de datos y el documento en Excel del registro de atención brindada al punto de venta. (Hurtado, 2000)	
<b>Número de Unidades Informantes:</b>	2 objeto. 3 sujetos.					
<b>Referencia:</b>	Hurtado, J. (2000). <i>Metodología de la Investigación Holística</i> (Tercera ed.). Caracas, Venezuela: Fundación Sypal.					

## 9. Técnicas e instrumento

<b>Técnica: Entrevista</b>					
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parfraseo</b>	<b>Aplicación en su tesis</b>	<b>Redacción final</b>
Victor Miguel Niño	2011	Niño (2011) señalo que “La entrevista es una técnica, fundamentalmente de tipo oral, basada en preguntas y respuestas entre investigador	La entrevista es una técnica que nos permitirá recolectar información relevante para la investigación, a través de una conversación entre el	La entrevista será empleada para obtener información, por parte de los trabajadores responsables de la gestión de datos y el uso de los mismos, para cumplir con los	La entrevista es una técnica empleada para recolectar información relevante para la investigación, por medio de la conversación, a base de preguntas y

Rojas		y participantes, que permite recoger las opiniones y puntos de vista de dichos participantes o, eventualmente, según objetivos, intercambiar con ellos en algún campo” (p.64)	investigador y el participante, a base de preguntas y respuestas, siendo capaces de obtener puntos de vista, opiniones o ideas de los participantes (Niño, 2011).	objetivos planteados en la investigación. Además, mediante el uso de la técnica, se puede obtener información adicional de los problemas que presentan, para exponer una conclusión más completa.	respuestas. (Niño, 2011) Además, se puede obtener información adicional por medio de puntos de vista, opiniones o ideas de los participantes respecto a los problemas que presentan, para exponer una conclusión más completa en el trabajo de investigación. Razón por la cual, se aplicará la encuesta a los trabajadores responsables de la gestión de datos y el uso de los mismos, para cumplir con los objetivos planteados en la investigación.
<b>Referencia:</b>	Niño, V. (2011). <i>METODOLOGÍA DE LA INVESTIGACIÓN. DISEÑO Y EJECUCIÓN</i> (Primera ed.). Bogotá, Colombia: Ediciones de la U.				
<b>Instrumento: Guía de entrevista</b>					
<b>Autor/es</b>	<b>Año</b>	<b>Cita</b>	<b>Parfraseo</b>	<b>Aplicación en su tesis</b>	<b>Redacción final</b>
Alejandro Acevedo Ibáñez y Alba Florencia López Martín	2004	Acevedo & López (2004) explico que “La importancia de la guía de entrevista es de recordarle cuales son las áreas principales que debe tocar con su indagación, la secuencia que llevara la conversación, y también que tipo de preguntas plantear. De esta forma, sabrá como movilizar la conversación y evitara que esta, más que una entrevista, parezca un interrogatorio (p.77).	La guía de entrevista es un documento que contiene temas y preguntas a analizar relacionadas al objetivo planteado en la investigación; para poder direccionar la conversación y evitar que el sujeto a encuestar, se aburra o incomode, y pueda brindar mayor información (Acevedo & López, 2004).	El instrumento explicado, será una guía para no salir de contexto. Además, permitirá desarrollar la entrevista de forma fluida y armoniosa; obteniendo mayor cantidad de información para la investigación con la variación y el aumento de preguntas. También puede contener notas con puntos importantes a tratar.	La guía de entrevista es un documento que contiene temas y preguntas a analizar, relacionadas al objetivo planteado en la investigación; siendo una guía para no salir del contexto, y direccionar la conversación, evitando que el sujeto a encuestar, se aburra o incomode, y pueda brindar mayor información. (Acevedo Ibáñez & López Martín, 2004). Además, también puede contener notas con puntos importantes a tratar. De esta forma, la entrevista se desarrolle de forma fluida y armoniosa; con la libertad de poder variar o aumentar preguntas para lograr recolectar mayor cantidad de información para la investigación.
<b>Referencia:</b>	Acevedo, A. & López, A. (2004). <i>El proceso de la entrevista: conceptos y modelos</i> . México: Editorial Limusa.				