Propuesta de modelo integrado de datos para la administración de garantías y toma de decisiones en Superbid.

Hernán Darío Garzón Duarte, Juan David Felipe López Rodríguez, Leandro Adriano Casas Torres, Luis Alfredo Narváez Jaramillo, Luisa Catherine Serrano Chaves, Sandra Milena Díaz Siza

Especialización en Big Data, Fundación Universitaria Compensar

Proyecto fin de grado

Wilson Hernando Soto Urrea

19 de mayo de 2023

Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional.

Propuesta de modelo integrado de datos para la administración de garantías y toma de decisiones en Superbid.

Hernán Darío Garzón Duarte, Juan David Felipe López Rodríguez, Leandro Adriano Casas Torres, Luis Alfredo Narváez Jaramillo, Luisa Catherine Serrano Chaves, Sandra Milena Díaz Siza

Trabajo de grado presentado como requisito parcial para optar al título de:

Especialista en Big Data

Ingeniero:

Wilson Hernando Soto Urrea

Fundación Universitaria Unipanamericana – Compensar

Facultad de Ingeniería, Especialización en Big Data

Bogotá, Colombia

2023

Resumen

La empresa Superbid dentro de sus funciones contempla la custodia de las garantías (dinero) de los ofertantes durante el proceso de subastas, sin embargo, la administración de estas garantías presenta algunas oportunidades de mejora debido a la naturaleza, altamente manual, del proceso y a la descentralización de los datos. Con la propuesta del modelo de integración se busca optimizar la gestión y oportunidad en la devolución de las garantías, trabajando desde un esquema de seguridad de la información para asegurar sus principales criterios como son la integridad, confidencialidad y disponibilidad.

El modelo seleccionado para el desarrollo de la solución fue CRISP-DM, puesto que permite ajustarse a las necesidades de la empresa. Por lo tanto, se logrará la integración a través de un modelo relacional que ofrecerá una estructura ordenada de los datos, con el fin de generar informes confiables para el proceso de administración de garantías. Esto, en última instancia, se traduce en una optimización en la gestión y en la toma de decisiones acertadas.

Palabras clave: Modelo CRISP-DM, modelo relacional, integridad, confidencialidad, disponibilidad, garantías.

Abstract

Superbid Colombia is a Brazilian company established in 1999, specialized in providing an auction platform for companies to showcase their assets and customers to place bids. One of the key requirements for participating in an auction and making offers is the payment of a warranty, which is refunded if the customer does not win any assets.

It has been identified areas for improvement in the warranty process, particularly in streamlining manual procedures and integrating relevant databases. This article proposes a model designed to enhance response times, ensure data integrity, maintain confidentiality, and improve overall data availability.

The selected model for the solution development is CRISP-DM, as it allows for adaptation to the company's needs. Therefore, integration will be achieved through a relational model that will provide an organized structure of the data, aiming to generate reliable reports for the warranty management process. Ultimately, this translates into optimization in management and making informed decisions.

Keywords: CRISP-DM Model, relational model, integrity, confidentiality, availability, warranties.

Tabla de Contenido

RESUMEN **3**

ABSTRACT **4**

TABLA DE CONTENIDO **5**

TABLAS **7**

INTRODUCCIÓN **8**

ANTECEDENTES Y JUSTIFICACIÓN 9

FORMULACIÓN DEL PROBLEMA 10

OBJETIVO GENERAL 11

OBJETIVOS ESPECÍFICOS 11

ALCANCES Y LIMITACIONES 11

MARCO TEÓRICO **13**

DISEÑO METODOLÓGICO **16**

COMPRENSIÓN DEL NEGOCIO. 17

COMPRENSIÓN DE LOS DATOS 17

DEFINICIÓN DE REQUISITOS 18

PREPARACIÓN DE LOS DATOS 18

RESULTADOS **25**

INDICADOR DE GESTIÓN DE DEVOLUCIONES DE GARANTÍAS 26

INDICADOR DE COMPORTAMIENTO DE CLIENTES ATÍPICOS 28

INDICADOR DE POTENCIAL DEL CLIENTE 29

DISCUSIÓN **32**

CONCLUSIONES Y RECOMENDACIONES 34

CONCLUSIONES 34

RECOMENDACIONES 35

BIBLIOGRAFÍA 36

Tablas

Tabla 1. Fuentes originales de garantías. 17

Tabla 2. Requerimientos funcionales y no funcionales..... 18

Tabla 3. Tabla Compradores 19

Tabla 4. Tabla Garantías 20

Tabla 5. Tabla Lotes 20

Tabla 6. Tabla Subastas. 21

Tabla 7. Tabla Vendedores..... 21

Tabla 8. Tablas Complementarias 22

Tabla 9. Indicador de gestión de devoluciones de garantías..... 27

Tabla 10. Rangos de Desempeño..... 28

Tabla 11. Ejemplo indicador Rangos de Desempeño..... 28

Tabla 12. Indicador de comportamiento de clientes atípicos..... 29

Tabla 13. Indicador de potencial del cliente 30

Tabla 14. Rangos de Desempeño..... 31

Tabla 15. Ejemplo de aplicación del indicador de potencial del cliente 31

Introducción

En el mundo actual se observa una creciente dependencia de las empresas con el análisis de datos. Según Forbes 2019 cerca del 59% de las empresas usan la analítica en alguna medida, ayudando a tomar mejores decisiones, adoptando mejores estrategias, y fortaleciendo las relaciones con clientes y socios comerciales. Superbid, tratándose de una compañía multinacional no es ajena a esta tendencia.

El Big Data contribuye a que las compañías puedan tomar decisiones de forma más rápida, eficiente y sin complicaciones, además de identificar nuevas oportunidades de negocio. Ahora, si bien el Big Data es un término que describe un gran volumen de datos, tanto estructurados como no estructurados que inundan los negocios hoy en día, es necesario mencionar que no es la cantidad de datos lo importante, si no lo que las organizaciones realizan con ellos, es decir, qué tanto análisis se puede llegar a realizar para obtener ideas que conduzcan a mejores resultados en los negocios.

De esta manera, la recopilación e integración de los datos, así como la búsqueda de tendencias dentro de los mismos, permiten que las empresas avancen de manera eficiente, haciendo que los datos sean confiables, consistentes y que sean gestionados en tiempo real. De la misma manera permite a las organizaciones eliminar las actividades problemáticas antes de que éstas acaben con sus beneficios o su reputación, llevando con esto, a movimientos de negocios más inteligentes, operaciones más eficientes, mayores ganancias y clientes más satisfechos.

Teniendo en cuenta estas premisas, este grupo de seminario de grado investigó sobre los procesos operativos y comerciales de la entidad Superbid[1], cuya casa matriz se encuentra en Brasil con sedes en algunos países de Latinoamérica, incluyendo a Colombia, y la cual, tiene como objetivo económico servir como intermediario para llevar a cabo subastas a través de una plataforma virtual, beneficiándose de las comisiones resultantes de esta actividad de prestación de servicios a los clientes.

El proceso principal de Superbid se asocia con la administración comercial de las subastas en la plataforma Web, con base en unas garantías previas que deben ser depositadas por los clientes. De esta actividad se generan los datos sobre las diferentes acciones que componen el proceso para llevar a cabo la gestión al interior de la entidad, siendo un procedimiento fundamental para la gestión de las garantías, toda vez que, representan un flujo de caja importante dentro de la actividad diaria de la entidad.

Antecedentes y Justificación

Superbid es una empresa fundada en Brasil en 1999, la cual, ofrece un ambiente digital de transacciones de compra y venta de bienes de consumo, bajo el modelo de subastas on-line. La entidad conecta a través de la plataforma a más de 8 mil empresas vendedoras y más de 1 millón de compradores, brindando credibilidad, seguridad y transparencia en el proceso de venta (Superbid, s.f).

[1] https://www.superbid.com.co/

Ahora bien, respecto a la operación implementada actualmente en la sucursal de Colombia para la entidad, fue posible identificar que se cuenta con diferentes fuentes de datos no relacionadas u homologadas en el proceso de garantías. De igual manera, se observaron oportunidades de mejoramiento en las actividades que componen el proceso asociado con su almacenamiento y procesamiento de los datos.

Contar con un modelo de datos estructurados, con base en las fuentes actuales, permitiría a Superbid optimizar el proceso de las garantías, mitigando los riesgos asociados al error humano. De igual manera, teniendo en cuenta que la implementación del proceso ha sido empírica, es importante contar con un estudio estratégico que permita optimizar la administración de los datos para la generación de información confiable, y poder finalmente, establecer indicadores clave (KPI) del proceso, mejorando la atención al cliente.

Formulación del Problema

La entidad Superbid no cuenta con la integración de las diferentes fuentes de datos asociadas al proceso financiero de garantías de los clientes. Dichas fuentes se basan en hojas de cálculo independientes y no cuentan con un procedimiento de homologación, lo cual, genera operatividad manual que conlleva a debilidades en la gestión de las garantías.

Por otra parte, al carecer de indicadores frente al comportamiento del proceso y gestión en los datos se observa un impacto en los criterios de seguridad de la

información como son la integridad, disponibilidad y confiabilidad para la toma de decisiones acertadas en la administración.

Objetivo General

Proponer un modelo de integración asociado a las operaciones de Garantías de los clientes de Superbid.

Objetivos Específicos

- Identificar las situaciones adversas que surgen de la falta de integración de los sistemas de información en Superbid.
- Estudiar las fuentes actuales de datos con los que cuenta el proceso de devolución de garantías.
- Clasificar las fuentes con el fin de establecer sus relaciones interdependientes.
- Diseñar un modelo de datos que integre las diferentes fuentes.
- Proponer indicadores claves de desempeño que permitirán llevar a cabo una medición sobre el proceso de las devoluciones de las garantías.

Alcances y Limitaciones

El alcance abarcado para llevar a cabo los objetivos específicos fue:

- La investigación se llevó a cabo con base en el proceso de gestión de garantías de la empresa Superbid, sucursal Colombia.

- Se tuvo en cuenta las fuentes de datos y reportes que se generan con base en la actividad de la plataforma Web y que, la entidad almacena en hojas de cálculo Google.
- Para la clasificación de las fuentes se tuvo en cuenta el conocimiento y asesoramiento del personal de la entidad, con quienes se llevaron a cabo entrevistas encaminadas a conocer las actividades y protocolos que componen el proceso.
- Para el diseño de la propuesta del modelo de datos, el grupo de proyecto se basó en el esquema de entidad - relación.
- En el diseño y propuesta de los indicadores se tuvo en cuenta la eficiencia y desempeño que se espera mejorar en el proceso de garantías.

Una limitación Importante que se debe mencionar, de acuerdo con los objetivos establecidos para este proyecto, consiste en que no se cuenta con una fase de implementación para el modelo que sea propuesto.

Marco Teórico

La integración de datos es un proceso fundamental en el ámbito de la gestión de la información y se refiere a la combinación de datos provenientes de diferentes fuentes y sistemas en una única fuente coherente y unificada. Este proceso tiene como objetivo principal garantizar la disponibilidad de los datos correctos en el momento adecuado para su análisis y toma de decisiones.

Teniendo en cuenta los conceptos clave y los enfoques utilizados en la propuesta de integración de datos, este grupo de proyecto ha trazado una ruta para el desarrollo de la consultoría, la cual, con base en las fuentes de datos del proceso de garantías, se identificaron aquellas variables más relevantes para ser unificadas y almacenadas en un repositorio principal. Estas actividades se llevaron a cabo con base en lo sugerido por Cevallos Zhunio & Velásquez, quienes en su artículo de la revista científica ReciMundo, expone un proceso de transformación basado en la unificación, filtrado y clasificación de los datos más apreciables, para finalmente, llevar a cabo un proceso de carga que incluya técnicas propias de la carga de datos y actualización. (Cevallos y Velásquez, 2020).

Adicionalmente, para del presente proyecto se realizó una investigación sobre los diferentes trabajos relacionados con integración de datos, con el fin de retroalimentar los objetivos y actividades asociadas con el desarrollo del mismo. Algunos de los más importantes fueron:

El trabajo de grado titulado “Sistema de información para el monitoreo del proceso de compensación salarial en la empresa de servicios temporales Listos S.A.S. a través de tableros de control”, nos muestra el desarrollo de tableros de control por medio de

Power BI, para el monitoreo de nóminas, desempeño, compensación salarial, entre otros, con el fin de recolectar información valiosa de los datos para la toma de decisiones (Sánchez, 2020).

Igualmente, se revisó el trabajo “Energy Analytics”, una propuesta de un prototipo mediante el uso de Power BI, el cual, entrega un informe de consumo y genera una predicción de los gastos generados, además de brindar información para minimizar el costo y consumo de los mismos (Vargas, 2019).

El trabajo de grado “Unificar la información de las bases de datos de la UNIMINUTO mediante minería de datos en el sistema Banner 9.0” propuso unificar la información de las bases de datos del sistema académico de la Universidad UNIMINUTO, mediante el uso de minería de datos (Bohórquez, 2019).

Amézquita en su trabajo “Minería de datos para la identificación de factores de riesgo en pacientes con hipertensión arterial”, nos muestra el análisis de datos de pacientes que padecieron la enfermedad de hipertensión mediante la aplicación de ciertos algoritmos para encontrar patrones y variables puntuales que, pueden determinar riesgos y progresos del paciente (Amézquita, 2019).

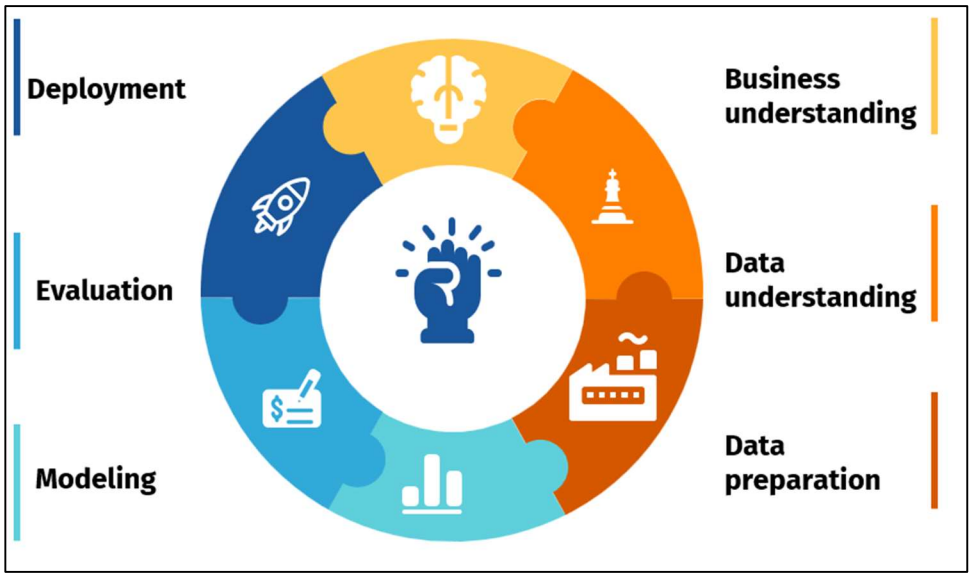
Finalmente, se revisó el trabajo “Modelo predictivo para analizar los datos de la calidad del aire urbano, capturados por el RMCAB en la localidad de Chapinero”, el cual, se enfoca en la captura de datos mediante la Red de Monitoreo de Calidad del Aire de Bogotá – RMCAB, en la localidad de Chapinero, para la creación de un modelo de machine learning, que pueda llegar a predecir el estado de la calidad del aire en meses futuros (Gonzales, 2017).

Para llevar a cabo el Proyecto, se contemplaron los tres modelos para minería de datos contemplados durante la Especialización en Big Data, los cuales son: KDD (Knowledge Discovery in Databases), SEMMA (Sample, Explore, Modify,. Model, Assess) y CRISPDM (Cross-Industry Standard Process for Data Mining). En revisión de cada uno de estos modelos referenciados, se selección el denominado CRISP-DM, siendo uno de los que se ajustaba con mayor precisión a los objetivos del proyecto, basado principalmente en la limpieza y modelado de los datos fuentes de Superbid para el proceso de garantías. El modelo CRISP-DM contempla las fases de comprensión del negocio, comprensión de los datos, definición de requisitos y preparación de los mismos, modelado, evaluación y despliegue. Importante mencionar que, de acuerdo con los alcances definidos para este proyecto, se utiliza esta metodología hasta su fase de modelado.

Diseño Metodológico

El modelo Cross-Industry Standard Process for Data Mining (CRISP-DM) es ampliamente utilizado en la minería de datos y análisis de los mismos. Proporciona una estructura detallada, sistemática y adaptable para que los profesionales realicen un proceso de minería de datos, desde la comprensión del problema hasta la implementación de las soluciones (Wirth, R., 2000).

Gráfica 1. Modelo CRISP-DM



Elaboración propia

Teniendo en cuenta las fases que propone el modelo, se procede a desarrollar cada una de ellas para las fuentes de datos en las que, se apoya el proceso de garantías.

Comprensión del negocio.

Para comprender los objetivos del negocio fue necesario llevar a cabo entrevistas con el personal especializado en el proceso de garantías de Superbid Colombia, con el fin de realizar el recorrido del flujo de información inmerso en el proceso y de cuya observancia fue posible identificar las oportunidades de mejora.

Posteriormente, teniendo en cuenta que el grupo de proyecto estaba integrado por profesionales en las diferentes áreas o disciplinas, cada uno de ellos generó los cuestionamientos relacionados con el proceso de garantías, para después llegar a unificar los criterios desde los diferentes puntos de vista. Como resultado de este ejercicio, se llegó a un consenso para determinar el objetivo principal del proyecto y los alcances relacionados.

Comprensión de los datos

Se llevó a cabo recolección sobre los datos que apoyan el proceso de garantías para llegar a explorarlos. Al respecto, se realizó un primer análisis con el fin de comprender la calidad y relevancia que estos reportes tienen y, cómo aportan valor en la operación del negocio.

Las fuentes identificadas fueron las siguientes:

Tabla 1.

Fuentes originales de garantías.

Nombre archivo	Formato	Área Responsable Custodia	Frecuencia actualización	Ubicación
RELATORIO	Hoja de Cálculo Excel	Gerencia Comercial	diaria	Servidor local
CONTROL GARANTÍAS	Hoja de Cálculo Google	Gerencia Comercial	diaria	Google drive
CUENTAS BANCARIAS	Hoja de Cálculo Excel	Gerencia Comercial	diaria	Servidor local

Nombre archivo	Formato	Área Responsable Custodia	Frecuencia actualización	Ubicación
CONTOL DEVOLUCIONES FINANCIERA	Hoja de Cálculo Excel	Gerencia Financiera	diaria	Servidor local

Definición de requisitos

Como parte de la clasificación de los reportes y su valor agregado dentro del proceso, se establecieron los requerimientos de negocio observables, los cuales, fueron determinados como funcionales y no funcionales, como se describe a continuación.

Funcionales: aquellos que se centran en las funcionalidades y acciones específicas que el sistema o la solución de integración de datos debe realizar.

No funcionales. Son aquellos que se centran en aspectos como el rendimiento, la escalabilidad, la seguridad, la usabilidad, la disponibilidad y otros criterios de calidad.

Tabla 2.

Requerimientos funcionales y no funcionales

No.	Requisito	FUNCIONAL	NO FUNCIONAL
1	Dar respuesta al estado real de cualquier garantía.	✓	
2	Dar respuesta al estado de cuenta de los compradores.	✓	
3	Dar respuesta a la cantidad de lotes por subasta, su ganador y el estado de las garantías de los que no resultaron ganadores.	✓	
4	Reducir los tiempos requeridos para la generación del informe RPG01.		✓
5	Permitir la identificación de los clientes compradores que se registran dos veces con cuentas bancarias distintas.		✓

Preparación de los datos.

Se hizo uso de las fuentes de los datos, los cuales, si bien contaban con la con la debida autorización de acceso por parte de Superbid, se encontraban enmascarados

para proteger su confidencialidad. Con base en estas fuentes se llevó a cabo la selección de atributos relevantes, aplicando técnicas de normalización de datos para construir el modelo de integración propuesto, el cual incluyó, además de los datos ya existentes en las fuentes actuales de Superbid, estructuras nuevas para garantizar que se elimina la redundancia y la duplicación de datos, facilitar las operaciones de actualización.

A continuación, se muestran las estructuras que componen el modelo de datos:

Tabla 3. Tabla COMPRADORES.

Tabla 3.

Tabla Compradores

Tabla	Variable	Tipo de Dato	Descripción
COMPRADORES	banco	varchar	Entidad bancaria.
	ciudad	varchar	Ciudad del banco.
	CodDaneComprador	varchar	Código Dane del comprador.
	contacto	varchar	Nombre del contacto
	cuenta_bancaria	varchar	Número de cuenta bancaria.
	departamento_comprador	varchar	Departamento de ubicación del comprador.
	direccion	varchar	Dirección de la cuenta bancaria.
	email	varchar	Email
	FechaCreacionRegistro	datetime	
	id_comprador	int	Identificador único de esta tabla, generado automáticamente
	id_cuenta	int	
	identificacion_cuenta	varchar	
	IdentificacionComprador	varchar	
	login	varchar	
	movil	varchar	
	nombre	varchar	
	telefono	varchar	
	telefono_comercial	varchar	
	telefono_residencia	varchar	
	tipo_cuenta	varchar	

Tabla 4.

Tabla Garantías

Tabla	Variable	Tipo de Dato	Descripción
GARANTIAS	BancoConsignacionGarantia	varchar	
	FechaAutorizacion	datetime	
	FechaConsignacionGarantia	datetime	
	FechaCreacionRegistro	datetime	Se diligencia automáticamente con la fecha en la que se crea el registro
	FechaTrasladoGarantia	datetime	
	id_comprador	int	
	id_Garantia	int	Identificador único de esta tabla, generado automáticamente
	id_Subasta	int	
	id_TipoGarantia	int	
	IdGarantiaOrigen	int	
	NumeroConsignacionGarantia	varchar	
	OfertanteHabilitado	bit	
	UsuarioAutoriza	varchar	
	ValorGarantiaConsignado	money	

Tabla 5.

Tabla Lotes

Tabla	Variable	Tipo de Dato	Descripción
LOTES	cantidad_lote	int	
	cantidad_rematada	varchar	
	cantidad_visitas	money	
	categoria	int	
	DescripcionLote	varchar	
	FechaCreacionRegistro	datetime	Se diligencia automáticamente con la fecha en la que se crea el registro
	id_CcompradorGanador	int	
	id_lote	int	Identificador único de esta tabla, generado automáticamente
	id_vendedor	int	
	mayor_oferta	int	
	NombreLote	varchar	
	placa_vehiculos	varchar	

Tabla	Variable	Tipo de Dato	Descripción
	valor_inicial	money	
	vli	money	
	vml	money	

Tabla 6.

Tabla Subastas.

Tabla	Variable	Tipo de Dato	Descripción
SUBASTAS	fecha_ relatorio	datetime	
	fecha_aprobación	datetime	
	FechaCreacionRegistro	datetime	Se diligencia automáticamente con la fecha en la que se crea el registro
	id_Subasta	int	Identificador único de esta tabla, generado automáticamente
	id_vendedor	int	
	NombreSubasta	varchar	
	ValorGarantiaSubasta	money	

Tabla 7.

Tabla Vendedores

Tabla	Variable	Tipo de Dato	Descripción
VENDEDORES	ciudad	varchar	
	CodDaneVendedor	int	
	direccion	varchar	
	email	int	
	FechaCreacionRegistro	datetime	Se diligencia automáticamente con la fecha en la que se crea el registro
	id_vendedor	int	Identificador único de esta tabla, generado automáticamente
	IdentificacionVendedor	varchar	
	movil	varchar	
	NombreVendedor	varchar	
	telefono	varchar	

Otras tablas complementarias para el modelo son:

Tabla 8.

Tablas Complementarias

Tabla	Variable	Tipo de Dato	Descripción
ESTADOS_GARANTIAS	DescripcionEstadoGarantia	varchar	
	FechaCreacionRegistro	datetime	Se diligencia automáticamente con la fecha en la que se crea el registro
	id_EstadoGarantia	int	Identificador único de esta tabla, generado automáticamente
	NombreEstadoGarantia	varchar	
ESTADOS_LOTES	FechaCreacionRegistro	datetime	Se diligencia automáticamente con la fecha en la que se crea el registro
	id_EstadoLote	int	Identificador único de esta tabla, generado automáticamente
	NombreEstadoLote	varchar	
ESTADOS_SUBASTAS	FechaCreacionRegistro	datetime	Se diligencia automáticamente con la fecha en la que se crea el registro
	id_EstadoSubasta	int	Identificador único de esta tabla, generado automáticamente
	NombreEstadoSubasta	varchar	
HISTORICO_ESTADOS_GARANTIAS	FechaCreacionRegistro	datetime	Se diligencia automáticamente con la fecha en la que se crea el registro
	id_EstadoGarantia	int	Identificador único en la tabla ESTADOS GARANTIAS
	id_Garantia	int	Identificador único en la tabla GARANTIAS

Tabla	Variable	Tipo de Dato	Descripción
	id_HistoricoGarantia	int	Identificador único de esta tabla, generado automáticamente
HISTORICO_ESTADOS_LOTES	FechaCreacionRegistro	datetime	Se diligencia automáticamente con la fecha en la que se crea el registro
	id_EstadoLote	int	
	id_HistoricoLote	int	Identificador único de esta tabla, generado automáticamente
	id_lote	int	
HISTORICO_ESTADOS_SUBASTAS	FechaCreacionRegistro	datetime	Se diligencia automáticamente con la fecha en la que se crea el registro
	id_EstadoSubasta	int	
	id_HistoricoSubasta	int	Identificador único de esta tabla, generado automáticamente
	id_Subasta	int	
SUBASTAS_LOTES	FechaCreacionRegistro	datetime	Se diligencia automáticamente con la fecha en la que se crea el registro
	id_lote	int	
	id_Subasta	int	
	id_SubastaLote	int	Identificador único de esta tabla, generado automáticamente
TIPOS_GARANTIAS	DescripcionTipoGarantia	varchar	
	FechaCreacionRegistro	datetime	Se diligencia automáticamente con la fecha en la que se crea el registro

Tabla	Variable	Tipo de Dato	Descripción
	id_TipoGarantia	int	Identificador único de esta tabla, generado automáticamente
	NombreTipoGarantia	varchar	

Resultados

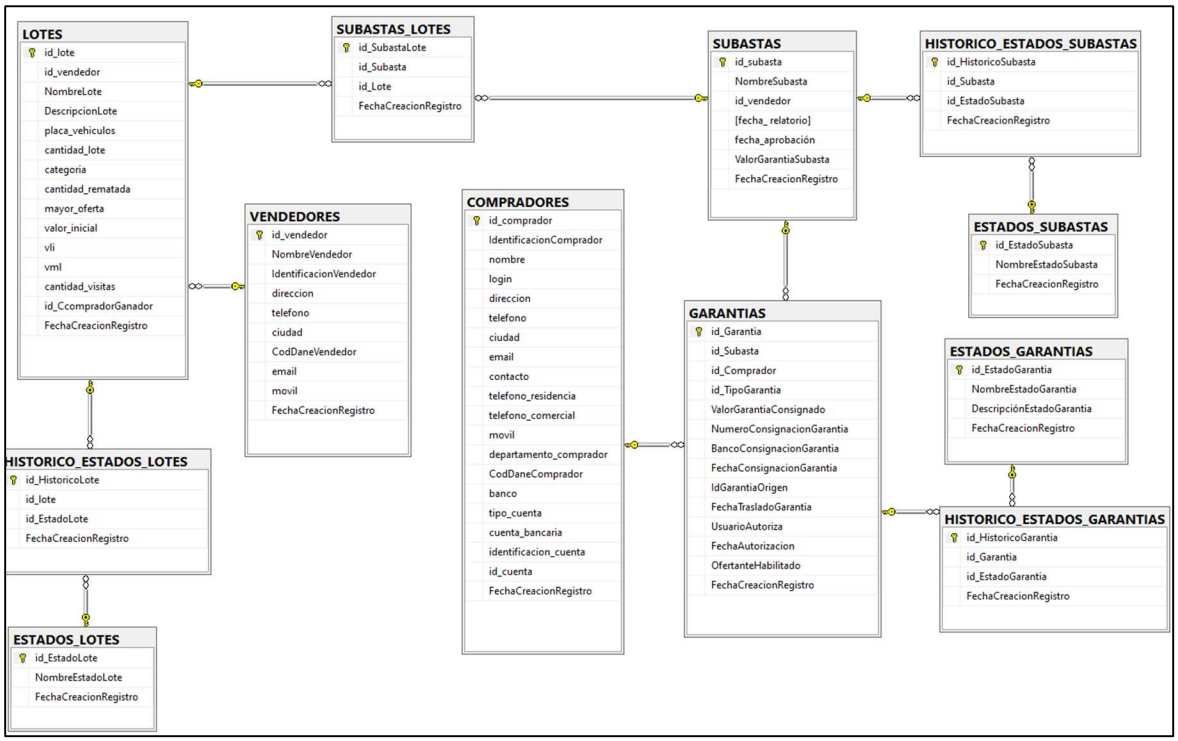
Con el objetivo de elaborar un modelo de datos integrado para la sucursal de Colombia de la empresa Superbid, el equipo de proyecto identificó, con el apoyo de los especialistas de la empresa, las oportunidades de mejora presentes en el proceso de garantías identificando como problemática general, desarticulación entre las diferentes fuentes de datos.

En consecuencia, fue necesario estudiar dichas fuentes en las que se basa el proceso establecido actualmente, llevando a un análisis y clasificación de las principales variables y sus relaciones interdependientes que impactan directamente sobre la actividad de las devoluciones dentro del proceso.

Posteriormente, se procedió a diseñar un modelo de datos basado en el esquema entidad relación usando el modelo CRISP-DM, el cual, se trata de uno de los modelos estándar de analítica, comúnmente usado para describir enfoques habituales en minería de datos. También se utilizaron herramientas conceptuales del Data Analytics y Big Data para inspeccionar, limpiar y proponer una transformación en los datos con el objetivo de resaltar información útil y sugerir resultados que apoye en la toma de decisiones.

El modelo entidad relación establecida se ilustra en la siguiente imagen.

Gráfica 2. Modelo entidad - relación



El modelo propuesto revela que existen datos que son cruciales para administración de las garantías. Algunos de estos datos pueden ser capturados automáticamente y son insumos primordiales para generar indicadores que brindan una visión del estado de las garantías.

Algunos indicadores que pueden derivarse de este modelo propuesto son los que se describen a continuación.

Indicador de gestión de devoluciones de garantías

Es una medida que permite evaluar el rendimiento del área de garantías en cuanto a la efectividad de sus solicitudes de devolución. Este indicador proporciona

información sobre el porcentaje de devoluciones que el área de garantías ha solicitado y no fueron efectivamente giradas.

Desde una perspectiva gerencial, este indicador es relevante porque refleja la efectividad del área de garantías en el proceso de devolución. Un alto porcentaje de pagos no devueltos podría indicar problemas en los procedimientos de seguimiento, comunicación o gestión de devoluciones.

Si el indicador muestra un porcentaje alto de pagos de garantía no devueltos, puede sugerir deficiencias en los procesos internos relacionados con la gestión de devoluciones. Esto podría incluir problemas con la coordinación entre el área de garantía y otros departamentos involucrados en el proceso de devolución, como el departamento financiero o el equipo de servicio al cliente.

Por otro lado, un bajo porcentaje de pagos de garantía no devueltos, indica que el área de garantía está logrando un alto nivel de éxito en la devolución de los pagos correspondientes a los clientes. Esto puede ser un reflejo positivo de la eficiencia y eficacia de los procesos internos, así como de una buena comunicación y coordinación con otros departamentos.

Tabla 9.

Indicador de gestión de devoluciones de garantías

Formula	Descripción
$IDG_{(a)} = 1 - \frac{\sum x}{\sum y}$	$a=$ rango periodo, subasta u otros $x=$ Solicitudes de devolución efectivamente giradas por el área financiera $y=$ Solicitudes enviadas por el área de garantías

Rangos de Desempeño del indicador de gestión de devoluciones de garantías.

El resultado del indicador se clasifica en un rango y para cada uno se detalla la novedad del no giro y su correspondiente participación dentro del rango.

Tabla 10.

Rangos de Desempeño

Rango (a)	Clasificación	Detalle novedad no girado	Resultado
0%-30%	Verde	• Pago no aplicado • Sin saldo • Cuenta rechazada • Pendiente documentación	Identifica detalle de las devoluciones solicitadas por el área de garantía que no fueron efectivamente giradas
31%-60%	Amarillo		
61% >	Rojo		

Ejemplo de uso del indicador.

Tabla 11.

Ejemplo indicador Rangos de Desempeño

Formula	Descripción
$IGDG_{(ene)} = 1 - \frac{80}{100} = 20\%$	Durante el mes de enero el área de garantías solicitó devolución de 100 casos al área financiera. El área financiera devolvió efectivamente 80. Como resultado, el indicador arroja un índice de gestión del 20%, alertando que del total de devoluciones solicitadas existe un 20% que no fue devuelto y sobre el cual se debe realizar seguimiento y gestión. Adicional permite detallar cual fue el estado de rechazo
Pago no aplicado: 50% Sin saldo: 10% Cuenta rechazada: 20% Pendiente de documentación: 20% **Total: 100%**	Dentro del 20% de giros no pagados efectivamente se encontró el siguiente detalle, lo que permite tomar acciones preventivas y tener un control eficiente de este proceso.

Indicador de comportamiento de clientes atípicos

El propósito de este indicador es identificar clientes que podrían beneficiarse de una gestión alternativa de sus garantías, evitando así procesos operativos adicionales de

devolución. Puede ser útil para optimizar el manejo de las garantías de participación en subastas. Algunos puntos clave a considerar para la aplicación de este indicador son:

- Identificación de oportunidades: permite identificar clientes que participan repetidamente en subastas sin lograr ganar, lo que sugiere la posibilidad de establecer un proceso alternativo para gestionar sus garantías.

- Reducción de costos y esfuerzos operativos: al evitar devoluciones frecuentes de garantías, se pueden reducir los costos y esfuerzos asociados con los procesos operativos de devolución, como el procesamiento de solicitudes y los tiempos de espera.

- Mejora de la eficiencia: al proponer una gestión alternativa de las garantías, se puede mejorar la eficiencia del proceso al evitar interrupciones y retrasos causados por devoluciones repetidas.

- Fomento de la participación continua: al retener las garantías no utilizadas para futuras subastas, se incentiva a los clientes a seguir participando en el proceso, lo que puede generar mayores oportunidades de negocio y fortalecer las relaciones comerciales.

Tabla 12.

Indicador de comportamiento de clientes atípicos

Formula	Descripción
$ICCA = \sum x$	a= Rango, por cliente x= Veces devolución no ganador

Indicador de potencial del cliente

Este indicador proporciona información sobre el éxito relativo del cliente en las subastas en las que ha participado. Un valor alto en este indicador indica que el cliente

tiene una buena capacidad para ganar subastas y aprovechar las oportunidades de negocio. Esto podría ser indicativo de su capacidad para ofrecer mejores precios, productos o servicios que la competencia, lo que lo convierte en un cliente valioso para la empresa.

Por otro lado, un valor bajo en este indicador podría sugerir que el cliente tiene dificultades para ganar subastas y puede tener menos influencia en el mercado. Esto podría ser una señal de que el cliente necesita más apoyo o que su potencial como cliente no es tan alto como se esperaba.

Este indicador brinda información valiosa para tomar decisiones gerenciales, como identificar a los clientes más valiosos y asignar recursos de manera estratégica. Además, permite evaluar el desempeño de los clientes en subastas y establecer estrategias para mejorar su participación y éxito en el futuro.

Tabla 13.

Indicador de potencial del cliente

Formula	Descripción
$IPC_{(a)} = \frac{\sum x}{\sum y}$	a= Periodo tiempo, subasta, cliente x= Participaciones ganador y= Participaciones totales

Creación del rango del indicador de potencial del cliente.

Tabla 14.

Rangos de Desempeño

Rango	Clasificación
0-30	Cliente no potencial o con falta de asesoría
31-59	Cliente recurrente
60-79	Cliente potencial básico
80-100	Cliente potencial premium

Ejemplo de aplicación para el indicador.

Tabla 15.

Ejemplo de aplicación del indicador de potencial del cliente

Formula	Descripción
$IPC = \frac{80}{100} = 80\%$	El cliente (a) de las 100 subasta que participó quedó ganador en 80, lo que significa que indicador de potencial es del 80%. Este cliente se ubica en el rango de cliente potencial premium
Carros 80% Chatarra 10% Maquinaria amarilla 10% **Total 100%** *Tabla 1: Ejemplo detalle categoría subasta*	Adicional al indicador de potencial es posible detallar el potencial por categoría de subasta, por periodos de tiempo u otra característica que pueda ser útil para la toma de decisiones

Discusión

Con el fin de dar respuesta al cuestionamiento de cómo aumentar la eficacia del proceso de las garantías de Superbid aplicando técnicas de Big Data y Data Analytics, los resultados generaron evidencia sobre los beneficios que implica la centralización, consolidación lógica y definición de indicadores claves para generar confiabilidad, disponibilidad e integridad en los reportes relevantes de la entidad para el proceso. Fue posible establecer una propuesta para optimizar y explorar la información que ofrecen los datos almacenados en la entidad, con el fin de obtener una oportunidad de mejora en la gestión de las garantías, haciendo mayor énfasis en su actividad de devolución; aspecto que se traduce en un mejor servicio para los clientes.

Como parte del diseño metodológico se tomó el modelo CRISP-DM por ser uno estándar y usualmente aceptado para analizar datos y describir enfoques comunes. Este modelo fue utilizado durante la etapa de identificación y análisis de las fuentes de datos, corroborando su utilidad al presentar unas fases lógicas y consecutivas, así:

- Comprensión del negocio.
- Comprensión de los datos.
- Preparación de los datos.
- Modelado.
- Evaluación.
- Despliegue.

Es importante mencionar que esta metodología fue aplicada hasta su fase de modelado, considerando que la evaluación y despliegue hace parte de una segunda fase

que implique poner en marcha la propuesta dada en este documento, siendo el modelado, exclusivamente, el alcance dado en la fase de planeación del proyecto.

Una vez identificados los resultados generales del presente trabajo de tesis, en el que se logró identificar las debilidades en el proceso de garantías de Superbid para luego, llegar a proponer un modelo integrado y estructurado de las fuentes y datos actuales de la entidad, es posible identificar similitud con los resultados publicados por Cevallos Zhunio & Velásquez (2020), dado que expusieron el proceso de transformación basado en la unificación, en el filtrado y la clasificación de los datos más relevantes; sin poder llegar a comparar los resultados asociados al proceso de implementación con base en técnicas propias de la carga de datos, dado que, no fue parte del alcance en nuestro trabajo.

Finalmente, fue posible establecer un modelo de datos basado en un esquema entidad relación, el cual, en su primera fase contó con actividades propias de la Data Analytics para realizar procedimientos de limpieza y transformación, y poder así, generar una propuesta de valor para Superbid en su proceso de garantías. Igualmente, se proponen indicadores relevantes que llevarán a medir de una manera objetiva y ajustada el desempeño del proceso, en pro de optimizarlo y gestionarlo de una manera más eficiente, como apoyo a los objetivos de la compañía.

Conclusiones y recomendaciones

Conclusiones

En virtud de lo descrito a lo largo de este documento, fue posible sistematizar los datos desarticulados de un proceso mediante un modelo sistemático, específicamente, el proceso de garantías en Superbid, proponiendo una estructura integrada, en la cual, se configuró un orden sobre las fuentes actuales para presentar una opción de optimización en la eficiencia y calidad del proceso. No obstante lo anterior, todavía queda mucho camino por recorrer frente a la implementación de dicho modelo, el cual, requerirá recursos adicionales para que pueda llegar a generar los frutos esperados. Es por esto que, resulta imprescindible que la organización considere darle continuidad al proyecto en su segunda fase de desarrollo e implementación.

Por otra parte, se logró demostrar con la ejecución de este proyecto, la gran importancia que tiene el Data Analytics para apoyar en el perfeccionamiento de los resultados relevantes en las empresas, para lo cual, los fundamentos del Big Data son de gran utilidad considerando que, su principal objetivo es el de convertir los datos sin procesar o dispersos en conjuntos de datos significativos que, luego puedan ser usados para obtener información de valor o, que permitan resolver problemas comerciales complejos.

Recomendaciones

Una vez concluido el presente proyecto, se pone a consideración del lector y de la comunidad educativa investigar sobre otros aspectos como son:

- Profundizar en otros estudios relacionados con los beneficios que ofrece el Data Analytics y su respectiva implementación, asociada con la sistematización de los datos.
- Trabajar en mejorar el modelo de integración propuesto, para alinearlo con un esquema de modelado orientado a objetos, con el fin de hacerlo escalable en el tiempo.
- Analizar con mayor detenimiento y, con el apoyo del personal especializado de la entidad, sobre otras razones subyacentes que afectan los resultados en el proceso de garantías, establecido actualmente en Superbid.

Bibliografía

Amézquita Moreno, G. I. (2019). Minería de datos para la identificación de factores de riesgo en pacientes con hipertensión arterial. Fundación Universitaria Unipanamericana - Compensar.

Bohórquez Jején, P. A. (2019). Unificar la información de las bases de datos de la UNIMINUTO mediante minería de datos en el sistema Banner 9.0

Cevallos y Velásquez (2020). Virtualización de datos una alternativa moderna de integración de datos. Revista científica Mundo de la investigación y el conocimiento.

Gonzales Peña, S. N. (2017). Modelo predictivo para analizar los datos de la calidad del aire urbano, capturados por el RMCAB en la localidad de Chapinero. Bogotá

Sánchez Lasprilla, A. F. (2020). Sistema de información para el monitoreo del proceso de compensación salarial en la empresa de servicios temporales Listos S.A.S. a través de tableros de control. Fundación Universitaria Compensar.

Superbid. (s.f). Obtenido de https://www.Superbid.com.co/quem-somos

Vargas Landinez, J. L. (2019). Energy Analytics. Fundación Universitaria Compensar.

Wirth, R. (2000). CRISP-DM : Towards a Standard Process Model for Data Mining. Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, 24959, 29–39.