



Universidad CENFOTEC

Maestría en Tecnología de Bases de Datos

Documento final de Proyecto de Investigación Aplicada 2

Tema: Esquema de integración y análisis de datos basado en metodologías ágiles para el fortalecimiento de la gestión en una empresa de insumos industriales SumIndustrial S.A.

Céspedes Torres, Juan Pablo.

Fecha: Julio, 2026.

Índice

Capítulo 1. Introducción	5
Generalidades	5
Antecedentes del Problema	5
Definición y Descripción del Problema	5
Justificación	6
Viabilidad	6
Objetivos	7
Alcances y Limitaciones	7
Marco de Referencia Organizacional y Socioeconómico	8
Revisión de la Literatura y Estado de la Cuestión	9
Capítulo 2. Marco Conceptual	10
Bases de datos relacionales y sistemas transaccionales (OLTP)	11
Data Warehouse y arquitectura analítica	11
Procesos ETL (Extract, Transform, Load)	12
Modelado dimensional	12
Procesamiento analítico en línea (OLAP)	13
Inteligencia de negocios (BI)	13
Segmentación de clientes y minería de datos	14
Análisis multidimensional y visualización analítica	14
Capítulo 3. Marco Metodológico	14
Tipo de investigación	14
Alcance investigativo	15
Enfoque	15
Diseño	15
Población y muestreo	16
Capítulo 4. Análisis de la Situación	17
Análisis de la gestión de datos	17
Análisis tecnológico	17
Análisis de procesos	17
Análisis organizacional y cultural	18
Análisis del entorno y necesidades del negocio	18

Identificación de oportunidades	18
Síntesis del diagnóstico	18
Capítulo 5. Propuesta de Solución	19
Arquitectura general propuesta	19
Diseño del modelo de datos analítico	20
Proceso de integración ETL propuesto	21
Herramientas tecnológicas	21
Aplicación de metodologías ágiles	21
Beneficios esperados.....	22
Hoja de ruta de implementación	22
Matriz de riesgos.....	22
Consideraciones finales de la propuesta.....	23
Capítulo 6. Implementación y Resultados	23
Descripción del sistema implementado y del conjunto de datos	23
Análisis OLAP multidimensional	24
Embudo de conversión de cotizaciones	26
Clasificación ABC/XYZ del inventario.....	26
Segmentación de clientes mediante K-means	27
Pronóstico de demanda con ARIMA.....	28
Modelo de regresión	29
Asistente conversacional y verificación	30
Conclusiones.....	31
Referencias	31

Lista de Figuras

Figura 1.1. Mapa esquemático de sedes de la empresa	9
Figura 2.1. Nube de conceptos del marco de referencia	11
Figura 2.2. Cubo OLAP multidimensional.....	13
Figura 3.1. Estructura jerárquica del enfoque metodológico	16
Figura 5.1. Arquitectura general de la solución propuesta	20
Figura 6.1. Ventas y margen por provincia del cliente, 2022–2025	24
Figura 6.2. Ventas por categoría industrial, 2022–2025	25
Figura 6.3. Evolución mensual de ventas, 2022–2025.....	26
Figura 6.4. Curva de Pareto de la clasificación ABC de inventario	27
Figura 6.5. Perfil comparado de los segmentos de cliente (K-means, $k = 2$)	28
Figura 6.6. Pronóstico de ventas a doce meses con intervalos de confianza al 95 %.....	29
Figura 6.7. Interacción con el asistente: consulta en lenguaje natural, respuesta y visualización	30

Lista de Tablas

Tabla 5.1. Tablas de hechos del modelo dimensional	20
Tabla 5.2. Dimensiones compartidas del modelo	21
Tabla 5.3. Herramientas tecnológicas seleccionadas	21
Tabla 5.4. Cronograma de implementación por fases	22
Tabla 5.5. Matriz de riesgos del proyecto	22
Tabla 6.1. Evolución anual de ventas, órdenes y margen.....	25
Tabla 6.2. Embudo de conversión de cotizaciones	26
Tabla 6.3. Clasificación XYZ por variabilidad de la demanda.....	27
Tabla 6.4. Caracterización de los segmentos de cliente	28
Tabla 6.5. Importancia relativa de variables en el modelo de regresión (sin fuga)	29
Tabla 6.6. Verificación de las respuestas del asistente contra los datos del estudio	30

Capítulo 1. Introducción

Generalidades

La alta exigencia que enfrentan las empresas de insumos industriales genera la necesidad de adoptar soluciones avanzadas de análisis de datos para comprender en profundidad el comportamiento de sus clientes, optimizar los procesos internos y fortalecer sus estrategias comerciales. Mediante arquitecturas adecuadas es posible aplicar inteligencia de negocios y evitar que las organizaciones sigan operando con información fragmentada, condición que compromete la identificación de oportunidades de crecimiento y la mejora de la gestión de inventarios sobre la base de datos confiables.

Antecedentes del Problema

El crecimiento acelerado del dato, sin una estrategia analítica que lo acompañe, produce grandes volúmenes de información provenientes de fuentes diversas (transacciones, sistemas de inventario, gestión comercial y actividad digital). Ese crecimiento no siempre viene acompañado de estrategias de gestión y análisis, lo que deriva en repositorios de datos sin valor, ecosistemas que no optimizan su información y procesos manuales que dificultan la integridad de los datos.

La suma de estas carencias evidencia la ausencia de modelos analíticos como esquemas estrella, copos de nieve o cubos OLAP, lo que interrumpe la capacidad de analizar la información desde varias perspectivas. Como consecuencia, se generan deficiencias en la calidad y la gobernanza de los datos, en un mercado de insumos industriales cada vez más competitivo donde la experiencia del cliente resulta determinante.

Definición y Descripción del Problema

El problema principal que aborda este proyecto radica en la carencia de una arquitectura integral de inteligencia de negocios que permita a las organizaciones de insumos industriales realizar análisis avanzados sobre sus datos operativos y transaccionales. Se requiere una plataforma analítica que facilite la exploración de patrones de compra, la segmentación de clientes según su perfil, el rendimiento de productos a nivel geográfico y el análisis del comportamiento de la operación comercial.

La ausencia de modelos analíticos multidimensionales y de esquemas dimensionales crea en las organizaciones datos dispersos y no integrados, lo que limita la capacidad de

generar conocimiento estratégico e impacta de forma negativa la toma de decisiones basada en datos. Ello afecta, además, la identificación de oportunidades de negocio, la optimización de inventarios y el seguimiento de los indicadores de desempeño.

Justificación

De forma general, el proyecto busca la necesidad de centralizar y estructurar la información proveniente de múltiples fuentes (ventas, inventarios, clientes y actividad comercial) con el fin de habilitar procesos de análisis que apoyen la toma de decisiones estratégicas. La investigación abarca el uso de principios de arquitectura de datos, analítica empresarial y metodologías ágiles, lo que permite una implementación incremental, adaptable y alineada con las necesidades reales del negocio.

En cuanto al aporte al área de estudio, el proyecto contribuye al campo de la tecnología de bases de datos mediante la aplicación práctica de conceptos como la integración de datos, la arquitectura analítica, el modelado dimensional y la explotación de la información, reforzando el enfoque de investigación aplicada propio de la maestría. Asimismo, incorpora metodologías ágiles, lo que añade valor al demostrar cómo pueden emplearse en proyectos de gestión de datos de forma incremental y adaptable.

Viabilidad

La propuesta es viable desde el punto de vista técnico, metodológico y académico, considerando el alcance de una investigación aplicada y el acompañamiento docente que garantiza su correcta ejecución.

Viabilidad técnica

El proyecto es técnicamente viable porque existen múltiples tecnologías accesibles que permiten desarrollar una arquitectura de inteligencia de negocios: herramientas de integración de datos (ETL), procesamiento analítico (OLAP), minería de datos y visualizaciones interactivas sobre plataformas que incorporan inteligencia artificial.

Viabilidad operativa

El proyecto es operativamente viable porque la solución está orientada a mejorar la toma de decisiones en puntos clave (ventas, marketing, logística e inventarios), integrando tableros interactivos con acceso en tiempo real y reduciendo la dependencia de los procesos manuales diarios.

Viabilidad económica

El proyecto es económicamente viable porque la inversión inicial se compensa con los beneficios de una mejor gestión de la información y de decisiones más acertadas, sumado a la disposición de la organización para invertir en el cambio.

Objetivos

Objetivo general

Proponer un esquema de integración y análisis de datos, basado en la aplicación de metodologías ágiles, que permita centralizar, estructurar y analizar la información generada por la empresa de insumos industriales, con el fin de apoyar la toma de decisiones estratégicas y mejorar la gestión operativa del negocio.

Objetivos específicos

- Analizar las fuentes de datos existentes en la empresa de insumos industriales, identificando su estructura, tipo y relevancia para el análisis del negocio y sus sedes.
- Diseñar una arquitectura de integración de datos que permita centralizar la información proveniente de los sistemas operativos de la empresa.
- Proponer un modelo de datos orientado al análisis que facilite la exploración de la información desde múltiples dimensiones del negocio.
- Definir un esquema de análisis de datos que permita identificar patrones de ventas, comportamiento de clientes y desempeño de productos aplicando inteligencia artificial.
- Integrar principios de metodologías ágiles para el desarrollo incremental y adaptable del esquema propuesto.
- Evaluar la viabilidad técnica y organizacional de la propuesta dentro del contexto de la empresa y su distribución geográfica.

Alcances y Limitaciones

Alcances

El proyecto contempla el diseño conceptual y lógico de un esquema de integración y análisis de datos aplicado a la empresa de insumos industriales. El alcance incluye:

- Identificación y análisis de las principales fuentes de datos del negocio.

- Propuesta de una arquitectura de datos que contemple la integración de información operativa y analítica.
- Diseño de modelos de datos orientados al análisis (modelo dimensional).
- Definición de procesos generales de integración y transformación de datos.
- Propuesta de mecanismos de análisis y visualización de información relevante para el negocio.
- Incorporación de un enfoque de desarrollo basado en metodologías ágiles.

El proyecto se orienta al diseño y a la validación de la solución como investigación aplicada. En esta fase se construyó, además, un prototipo funcional sobre datos sintéticos que permite demostrar el comportamiento del esquema, sin constituir todavía un despliegue en un entorno productivo real.

Limitaciones

- Disponibilidad y calidad de los datos existentes en la empresa.
- Posibles restricciones de acceso a información sensible, atendidas mediante el uso de datos sintéticos.
- Alcance centrado en el diseño y la validación mediante prototipo, y no en una implementación completa en producción.
- Dependencia del contexto organizacional para la validación de la propuesta.

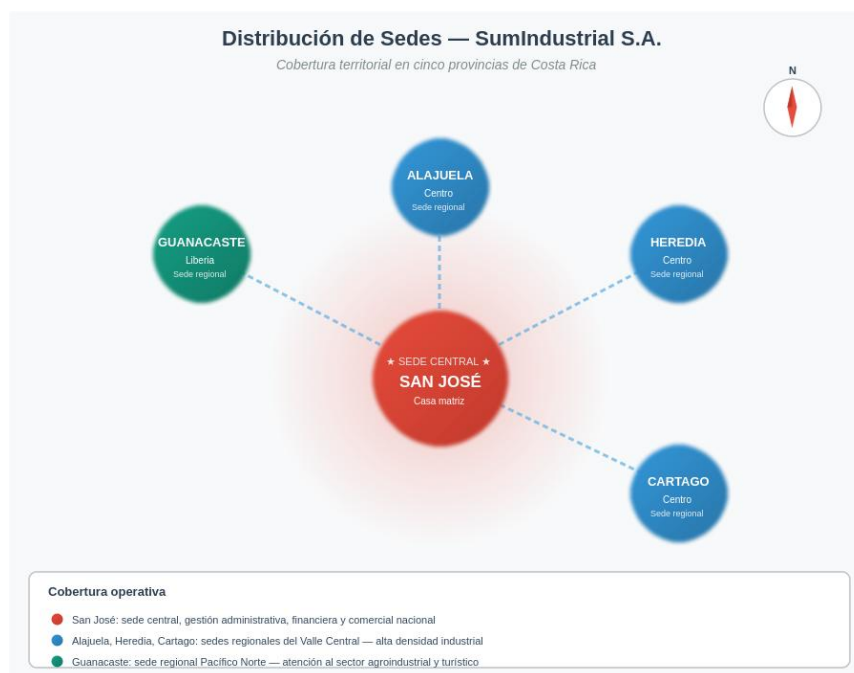
Marco de Referencia Organizacional y Socioeconómico

La empresa se dedica a la comercialización de insumos industriales orientados a la manufactura, el mantenimiento industrial y los servicios técnicos. Su portafolio incluye herramientas, repuestos, equipos de protección, componentes eléctricos y mecánicos, lo que incrementa la complejidad de la gestión de inventarios y de las operaciones comerciales. En el plano organizacional, la empresa dispone de sistemas transaccionales orientados a la gestión operativa (facturación, inventario y ventas), pero presenta limitaciones para el análisis estratégico debido a la falta de integración de los datos y a la ausencia de herramientas analíticas avanzadas.

En el ámbito socioeconómico, la empresa opera en un entorno altamente competitivo donde se busca eficiencia en la cadena de suministro, reducción de procesos y costos operativos, y una respuesta ágil al cliente. El comportamiento del mercado industrial está sujeto a variaciones de demanda influidas por factores económicos, lo que crea la

necesidad de contar con herramientas que permitan analizar tendencias, anticipar la demanda y optimizar la gestión de los recursos.

Figura 1.1. Mapa esquemático de sedes de la empresa



Fuente: elaboración propia.

Revisión de la Literatura y Estado de la Cuestión

Revisión de la literatura

La inteligencia de negocios se ha consolidado como un componente esencial de la gestión de la información organizacional, al permitir transformar datos en conocimiento útil para la toma de decisiones estratégicas. Según Kimball y Ross (2013), la implementación de modelos dimensionales y almacenes de datos facilita el análisis multidimensional y proporciona una visión integral del negocio. Inmon (2005) plantea que un almacén de datos constituye un repositorio centralizado, orientado a la toma de decisiones, que integra datos de diversas fuentes y habilita su análisis histórico; ello resulta fundamental en organizaciones que manejan grandes volúmenes de datos, como las empresas de insumos industriales, donde la gestión eficiente de inventarios, ventas y clientes es crítica.

Los procesos ETL (Extract, Transform, Load) cumplen un papel fundamental en la preparación de los datos. Golfarelli y Rizzi (2009) señalan que la calidad de los datos es un factor crítico para el éxito de los sistemas de inteligencia de negocios, pues la información inconsistente o incompleta afecta negativamente la toma de decisiones. En la

infraestructura tecnológica, la literatura destaca la importancia de sistemas de gestión de bases de datos robustos que soporten tanto el procesamiento transaccional como el analítico; en este trabajo se seleccionó SQL Server 2022 por su capacidad para manejar grandes volúmenes de información, su soporte para procedimientos almacenados, triggers y vistas, y su integración nativa con entornos Windows.

El uso de lenguajes orientados a la analítica se ha vuelto una práctica común en las soluciones modernas. Provost y Fawcett (2013) subrayan que las técnicas de minería de datos permiten descubrir patrones ocultos en grandes volúmenes de información. En este contexto se emplean bibliotecas como Pandas para la manipulación de datos, scikit-learn para los algoritmos de segmentación, Plotly para las visualizaciones interactivas y SQLAlchemy para la gestión segura de conexiones. La visualización, además, es un componente esencial de la entrega de información: según Sharda, Delen y Turban (2020), los tableros interactivos permiten explorar los datos de forma dinámica y democratizar el acceso a la información, para lo cual se adoptó Streamlit como marco de desarrollo de la interfaz.

Estado de la cuestión

Se evidencia una tendencia a integrar la tecnología de bases de datos con lenguajes de programación analíticos y herramientas de visualización, construyendo arquitecturas de inteligencia de negocios más flexibles y escalables. Estas soluciones no se limitan al análisis descriptivo, sino que incorporan analítica predictiva, minería de datos e inteligencia artificial, con lo que mejoran la capacidad de las organizaciones para anticipar comportamientos y gestionar mejor sus procesos.

Pese a los avances tecnológicos, muchas empresas presentan todavía un bajo nivel de madurez analítica, con dependencia de reportes estáticos y falta de integración de sus datos. Esta investigación resalta la pertinencia de soluciones basadas en herramientas como SQL Server, Python y Streamlit como alternativas viables para construir plataformas analíticas eficientes, accesibles y alineadas con las necesidades del negocio.

Capítulo 2. Marco Conceptual

El marco conceptual define los términos clave utilizados a lo largo de la investigación y, a la vez, contextualiza las decisiones tecnológicas adoptadas, justificándolas desde la literatura especializada en bases de datos y analítica avanzada. Con ello se establece una

conexión coherente entre la teoría y la práctica que fortalece el sustento metodológico del proyecto. La Figura 2.1 sintetiza, a modo de nube de conceptos, las herramientas, los procesos y las capacidades que articulan la propuesta.

Figura 2.1. Nube de conceptos del marco de referencia



Fuente: elaboración propia.

Bases de datos relacionales y sistemas transaccionales (OLTP)

Las bases de datos relacionales constituyen el núcleo de los sistemas de información empresariales, al estructurar los datos en tablas relacionadas mediante claves primarias y foráneas que garantizan integridad y coherencia lógica. El modelo relacional se fundamenta en la teoría de conjuntos y el álgebra relacional, lo que proporciona una base formal para la manipulación estructurada de la información. En muchos entornos, estas bases soportan sistemas transaccionales conocidos como OLTP (Online Transaction Processing), diseñados para registrar operaciones en tiempo real como ventas, pagos, registros de clientes y actualizaciones de inventario (Silberschatz, Korth y Sudarshan, 2020).

Estos sistemas priorizan la rapidez de escritura y la consistencia de los datos, apoyándose en las propiedades ACID para asegurar la confiabilidad de cada transacción. La normalización reduce la redundancia y evita inconsistencias, aunque incrementa la complejidad cuando se requieren consultas analíticas extensas. Aunque el entorno OLTP es esencial para la operación diaria, no resulta óptimo para los análisis históricos agregados ni para las exploraciones multidimensionales de alto volumen (Silberschatz et al., 2020).

Data Warehouse y arquitectura analítica

El almacén de datos surge como respuesta a las limitaciones analíticas de los sistemas transaccionales. Se define como un repositorio centralizado orientado a temas específicos del negocio, integrado, no volátil y variante en el tiempo, diseñado para facilitar la toma de decisiones estratégicas (Inmon, 2005). Su propósito es consolidar datos históricos de diversas fuentes en una estructura optimizada para consultas analíticas complejas.

La arquitectura analítica establece una separación clara entre los sistemas OLTP y OLAP (Online Analytical Processing), lo que permite que cada entorno se especialice según su función. Mientras el OLTP gestiona las operaciones diarias, el OLAP habilita análisis agregados y comparaciones multidimensionales sin afectar el rendimiento operacional. Esta separación se considera una buena práctica fundamental en la ingeniería de datos moderna (Kimball y Ross, 2013).

Procesos ETL (Extract, Transform, Load)

El proceso ETL constituye el mecanismo mediante el cual los datos se trasladan desde los sistemas fuente hacia el almacén de datos. En la extracción se obtienen los datos de las bases operacionales; en la transformación se aplican reglas de limpieza, validación, estandarización y enriquecimiento; y en la carga, los datos transformados se integran en las estructuras dimensionales del almacén analítico (Inmon, 2005).

La calidad del proceso ETL impacta directamente la confiabilidad del análisis posterior: una transformación incorrecta puede generar inconsistencias que afecten los indicadores estratégicos. Por ello, la automatización, la documentación y el control de estos procesos son elementos críticos de cualquier arquitectura de inteligencia de negocios (Kimball y Ross, 2013).

Modelado dimensional

El modelado dimensional es el enfoque predominante para diseñar estructuras dentro de un almacén de datos. Organiza la información en tablas de hechos, que contienen métricas cuantificables, y tablas de dimensiones, que describen los contextos de análisis como el tiempo, el cliente, el producto o la región. La estructura más común es el esquema estrella, caracterizado por su simplicidad y su eficiencia en las consultas (Kimball y Ross, 2013).

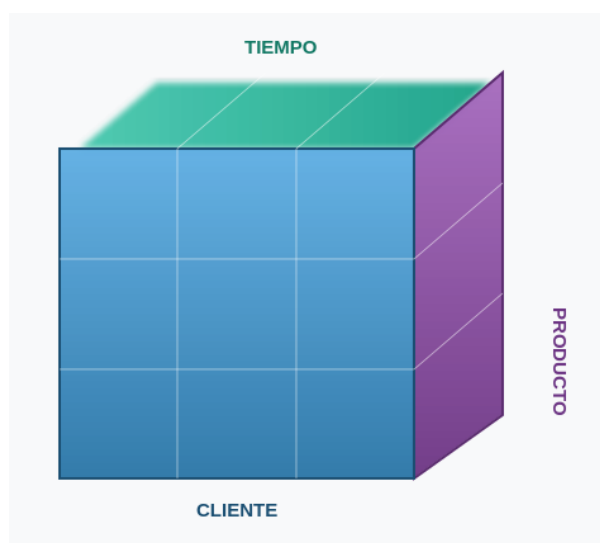
A diferencia del modelo relacional altamente normalizado, el modelo dimensional prioriza el rendimiento y la facilidad de uso. Las jerarquías definidas dentro de las dimensiones permiten realizar análisis en distintos niveles de granularidad, lo que facilita la comprensión del negocio desde múltiples perspectivas y optimiza las consultas agregadas (Golfarelli y Rizzi, 2009).

Procesamiento analítico en línea (OLAP)

El procesamiento analítico en línea permite realizar análisis multidimensionales sobre los datos históricos almacenados en el almacén. Las operaciones fundamentales incluyen slice, dice, drill-down, roll-up y pivot, que facilitan la exploración dinámica de la información desde distintos niveles de agregación (Chaudhuri y Dayal, 1997).

El OLAP permite analizar el desempeño por región, evaluar tendencias temporales y comparar resultados entre categorías de productos o segmentos de clientes. Esta capacidad de análisis interactivo mejora la rapidez en la generación de conocimiento estratégico y fortalece la toma de decisiones basada en evidencia cuantitativa (Chaudhuri y Dayal, 1997). La Figura 2.2 ilustra la estructura de un cubo multidimensional con las dimensiones de tiempo, cliente y producto.

Figura 2.2. *Cubo OLAP multidimensional*



Fuente: elaboración propia.

Inteligencia de negocios (BI)

La inteligencia de negocios integra tecnologías, metodologías y procesos orientados a transformar los datos en información útil para la toma de decisiones estratégicas. Abarca

tableros, indicadores clave de desempeño, reportes analíticos y visualizaciones interactivas que facilitan la interpretación de resultados en distintos niveles organizacionales (Sharda et al., 2020).

Más allá de la tecnología, la inteligencia de negocios promueve una cultura organizacional basada en datos, donde las decisiones se sustentan en análisis cuantitativos y evidencia empírica. Una correcta definición de métricas alineadas con los objetivos estratégicos permite evaluar el desempeño y ajustar las estrategias de manera oportuna (Sharda et al., 2020).

Segmentación de clientes y minería de datos

La minería de datos aplica técnicas estadísticas y algoritmos de aprendizaje automático para descubrir patrones significativos en grandes volúmenes de información. Entre sus aplicaciones más relevantes en el ámbito empresarial está la segmentación de clientes, que agrupa individuos con comportamientos similares mediante técnicas como el agrupamiento (Han, Kamber y Pei, 2012).

La segmentación facilita la personalización de las estrategias comerciales, optimiza las campañas de marketing y mejora la retención de clientes. Al identificar patrones de compra, frecuencia y valor monetario, las organizaciones pueden diseñar acciones diferenciadas orientadas a maximizar el valor del cliente a largo plazo (Han et al., 2012).

Análisis multidimensional y visualización analítica

El análisis multidimensional permite examinar la información desde múltiples dimensiones de forma simultánea, combinando variables como el tiempo, el producto, el cliente y la región. Esta perspectiva integral fortalece la interpretación estratégica del desempeño organizacional y facilita la identificación de oportunidades de mejora (Golfarelli y Rizzi, 2009).

La visualización analítica complementa este proceso al traducir datos complejos en representaciones gráficas comprensibles. El uso de aplicaciones web interactivas democratiza el acceso a la información y promueve una cultura basada en datos; la accesibilidad y la claridad en la presentación de los resultados resultan determinantes para la adopción efectiva de las soluciones de inteligencia de negocios (Sharda et al., 2020).

Capítulo 3. Marco Metodológico

Tipo de investigación

El estudio corresponde a una investigación aplicada, cuyo objetivo principal es desarrollar una solución tecnológica orientada a resolver un problema específico del contexto empresarial: la falta de herramientas analíticas que permitan transformar los datos operacionales en información estratégica. La investigación aplicada se caracteriza por utilizar conocimientos teóricos existentes para diseñar soluciones prácticas que mejoren los procesos organizacionales y optimicen la gestión de la información (Hernández Sampieri, Fernández Collado y Baptista Lucio, 2018).

En este proyecto se integran conceptos de bases de datos, analítica empresarial, inteligencia de negocios y minería de datos con el fin de construir una arquitectura que permita analizar patrones de comportamiento de clientes, desempeño de ventas y tendencias del negocio. La investigación no se limita a la generación de conocimiento teórico, sino que busca implementar una solución funcional que demuestre la aplicabilidad de los principios estudiados en un entorno de negocio.

Alcance investigativo

El alcance de la investigación es descriptivo y analítico. Es descriptivo porque identifica y analiza características relevantes del comportamiento de los datos relacionados con ventas, clientes e inventario, lo que permite describir patrones, tendencias y relaciones mediante técnicas de análisis multidimensional y visualización (Hernández Sampieri et al., 2018). Es analítico porque, además de describir, explora relaciones entre variables mediante minería de datos, segmentación y modelos predictivos, con el objetivo de generar conocimiento más profundo sobre el comportamiento del negocio y facilitar la identificación de oportunidades de mejora.

Enfoque

El enfoque de la investigación es cuantitativo, dado que se basa en el análisis de datos estructurados provenientes de sistemas de información empresariales. Este enfoque se caracteriza por el uso de mediciones numéricas, análisis estadístico y modelado de datos para identificar patrones y generar conclusiones objetivas (Hernández Sampieri et al., 2018). En el desarrollo del proyecto se emplearon técnicas de exploración estadística, análisis multidimensional OLAP, segmentación mediante agrupamiento y modelos predictivos basados en regresión y series temporales.

Diseño

El diseño corresponde a un estudio no experimental de tipo transversal. Es no experimental porque el investigador analiza los fenómenos tal como ocurren en su contexto, sin manipular las variables (Hernández Sampieri et al., 2018). Es transversal porque los análisis se realizan sobre un conjunto de datos recopilados durante un período específico, lo que permite evaluar el comportamiento del negocio en ese intervalo. A partir de esa información se construyen los modelos analíticos, los tableros y las herramientas de exploración que permiten comprender la dinámica del sistema empresarial y generar recomendaciones basadas en evidencia. La Figura 3.1 resume la estructura metodológica del estudio.

Figura 3.1. Estructura jerárquica del enfoque metodológico



Fuente: elaboración propia.

Población y muestreo

La población del estudio está conformada por el conjunto total de registros generados por los sistemas de la distribuidora de insumos industriales, que incluye información de clientes corporativos, productos, órdenes de venta, cotizaciones e inventario mensual por sucursal. Estos datos representan la totalidad de las transacciones y de los eventos comerciales registrados durante el período de análisis.

Debido a que el volumen de datos es manejable dentro del entorno analítico, se trabajó con la totalidad de los registros disponibles y no con una muestra representativa. Este enfoque censal permite obtener una visión completa del comportamiento del sistema, incrementa la precisión del análisis y reduce el riesgo de sesgos. Dado que el proyecto es una propuesta de diseño validada mediante prototipo, y con el fin de proteger información comercial sensible, los datos se generaron de forma sintética reproduciendo la estructura, el catálogo y los volúmenes característicos de una distribuidora industrial B2B.

Capítulo 4. Análisis de la Situación

Este capítulo evalúa el estado actual de la gestión de datos dentro de la distribuidora de insumos industriales e identifica las principales limitaciones, oportunidades y necesidades relacionadas con el uso de la información para la toma de decisiones. La empresa dispone de sistemas transaccionales orientados a la operación diaria (facturación, control de inventarios y gestión de ventas) que cumplen eficientemente con el registro de las transacciones, pero presentan limitaciones importantes cuando se trata de explorar la información con fines analíticos.

Análisis de la gestión de datos

Uno de los principales hallazgos es la fragmentación de los datos: la información se encuentra distribuida en múltiples fuentes sin un esquema de integración definido, lo que dificulta su consolidación y genera inconsistencias al realizar análisis globales del negocio. No existe una estructura centralizada que permita mantener históricos organizados ni trazabilidad de la información. Se identifican, además, procesos manuales para la generación de reportes que implican extraer datos de distintos sistemas, consolidarlos en hojas de cálculo y producir informes estáticos, lo que incrementa el riesgo de error y limita la oportunidad del análisis. En cuanto a la calidad, se detectan problemas de duplicidad, inconsistencia y falta de estandarización derivados de la ausencia de procesos formales de limpieza, transformación y validación.

Análisis tecnológico

La empresa cuenta con infraestructura suficiente para soportar sus operaciones transaccionales, pero carece de una arquitectura orientada al análisis de datos. No existe un almacén de datos ni un modelo analítico que separe los procesos operativos (OLTP) de los analíticos (OLAP). La ausencia de procesos ETL formalmente definidos limita la

capacidad de integrar datos de diferentes fuentes de manera consistente y automatizada, y no se dispone de herramientas de visualización avanzada ni de tableros interactivos que faciliten la exploración dinámica. Aunque la organización posee bases tecnológicas relevantes, estas no se aprovechan para generar valor analítico.

Análisis de procesos

Se observa una fuerte dependencia de actividades manuales y una falta de estandarización en la generación de reportes e indicadores. No existen flujos definidos para la gestión del ciclo de vida del dato, desde su captura hasta su análisis, ni prácticas formales de gobernanza (responsables, políticas de calidad, catálogos de datos o auditoría). La toma de decisiones se realiza de forma principalmente reactiva, con base en información histórica limitada y sin herramientas que permitan análisis predictivos o exploratorios.

Análisis organizacional y cultural

Se evidencia una baja madurez analítica, donde el uso de los datos se enfoca en la operación y no en la generación de conocimiento estratégico; no existe una cultura consolidada de toma de decisiones basada en datos. Los usuarios de negocio dependen de reportes predefinidos y carecen de herramientas para explorar la información de forma autónoma, lo que genera cuellos de botella en el acceso. No obstante, se identifica una disposición al cambio y apertura a la adopción de nuevas tecnologías, lo que representa una oportunidad clave para la implementación de soluciones de inteligencia de negocios.

Análisis del entorno y necesidades del negocio

La empresa opera en un entorno altamente competitivo, donde la optimización de inventarios, la eficiencia de la cadena de suministro y la comprensión del comportamiento del cliente son determinantes para el éxito. La falta de capacidades analíticas limita la posibilidad de identificar patrones de compra y segmentar clientes, analizar el rendimiento de productos y categorías, anticipar la demanda y optimizar inventarios, y evaluar el desempeño comercial por región o período. Esto genera una desventaja frente a organizaciones que sí utilizan analítica avanzada como soporte estratégico.

Identificación de oportunidades

Pese a las limitaciones, la organización cuenta con una base importante para evolucionar hacia un modelo analítico más maduro: dispone de datos operacionales

relevantes, sistemas estructurados, acceso a tecnologías modernas de análisis y la posibilidad de adoptar metodologías ágiles para una implementación incremental. La integración de procesos ETL, modelos dimensionales, análisis OLAP y plataformas de visualización permitiría transformar los datos operativos en información estratégica de alto valor.

Síntesis del diagnóstico

La situación actual refleja una brecha significativa entre la generación de datos y su aprovechamiento estratégico. Mientras que los sistemas actuales permiten operar el negocio de manera eficiente, no facilitan la explotación analítica de la información. Esta brecha se manifiesta en cuatro dimensiones: datos fragmentados, inconsistentes y no integrados; ausencia de arquitectura analítica en la dimensión tecnológica; procesos manuales y no estandarizados; y una cultura de baja madurez en el uso estratégico de los datos. En torno a estas evidencias se justifica el diseño de un esquema de integración y análisis de datos que permita evolucionar hacia una organización orientada a datos.

Capítulo 5. Propuesta de Solución

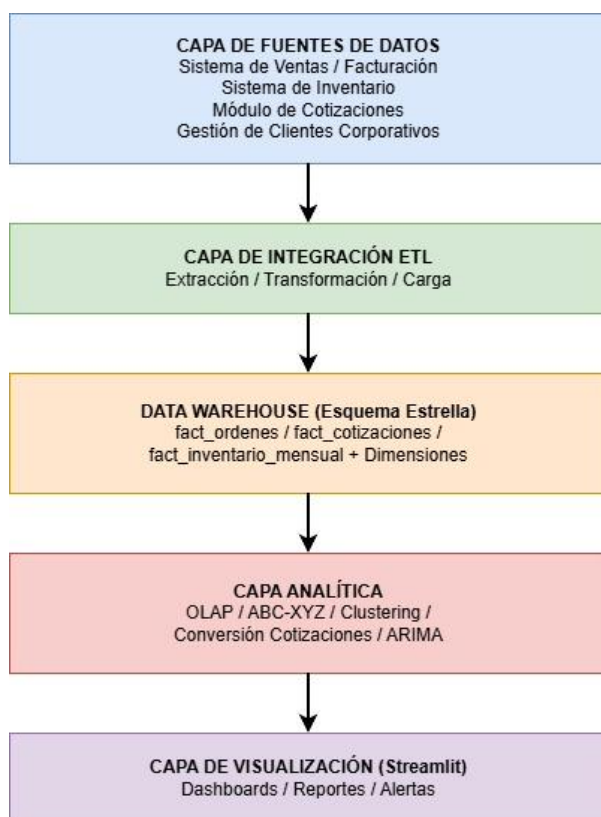
Con base en el diagnóstico, se identificó que la empresa presenta limitaciones significativas en la integración, la centralización y el aprovechamiento analítico de sus datos operativos. En respuesta, se propone el diseño de un esquema integral de integración y análisis de datos basado en una arquitectura de inteligencia de negocios, complementado con un enfoque de implementación sustentado en metodologías ágiles. La solución contempla procesos ETL, un modelo dimensional orientado al análisis, herramientas de visualización interactiva y mecanismos de análisis multidimensional que facilitan la exploración dinámica de la información.

Arquitectura general propuesta

La arquitectura se fundamenta en un modelo de múltiples capas orientado a separar los entornos operativo y analítico, lo que garantiza escalabilidad, mantenibilidad y eficiencia en el procesamiento. Se compone de una capa de fuentes de datos, integrada por los sistemas transaccionales de la organización; una capa de integración, responsable de los procesos ETL; una capa de almacenamiento analítico, constituida por el almacén de datos dimensional; una capa analítica, encargada de las consultas OLAP, los indicadores, la

segmentación y el análisis predictivo; y una capa de visualización, compuesta por tableros interactivos y por el asistente conversacional (Figura 5.1).

Figura 5.1. Arquitectura general de la solución propuesta



Fuente: elaboración propia.

Diseño del modelo de datos analítico

Se propone un modelo dimensional en esquema estrella, organizado como una constelación de hechos que comparten un conjunto común de dimensiones. El modelo incorpora tres tablas de hechos (órdenes de venta, cotizaciones e inventario mensual) y seis dimensiones compartidas, lo que permite analizar de forma coherente las ventas, el pipeline comercial y el inventario. La Tabla 5.1 resume las tablas de hechos y la Tabla 5.2, las dimensiones.

Tabla 5.1. Tablas de hechos del modelo dimensional

Tabla de hechos	Grano	Métricas principales
fact_ordenes	Línea de orden	Cantidad, precio, descuento, monto, costo y margen
fact_cotizaciones	Línea de cotización	Estado, conversión a orden y tiempo de respuesta
fact_inventario_mensual	Producto × sucursal × mes	Existencias, rotación y clase ABC/XYZ

Fuente: elaboración propia.

Tabla 5.2. Dimensiones compartidas del modelo

Dimensión	Atributos representativos
dim_tiempo	Fecha, día, mes, trimestre, año
dim_cliente	Empresa, sector industrial, tipo y tamaño de cuenta
dim_producto	Categoría, marca y tipo de insumo
dim_sucursal	Sede, región y capacidad
dim_proveedor	Proveedor, país de origen y confiabilidad
dim_geografia	Provincia, cantón y distrito

Fuente: elaboración propia.

Proceso de integración ETL propuesto

Se plantea un flujo ETL automatizado en tres fases. En la extracción se obtienen periódicamente los datos desde los sistemas fuente mediante conexiones directas a las bases operativas. En la transformación se aplican reglas de negocio: eliminación de duplicados, corrección de formatos, normalización de categorías, validación de integridad referencial y limpieza de valores nulos. En la carga, los datos se insertan en el almacén mediante cargas incrementales. Este proceso garantiza la calidad, la consistencia y la disponibilidad continua de la información.

Herramientas tecnológicas

Con base en la revisión bibliográfica y en las necesidades identificadas, se seleccionaron las herramientas que resume la Tabla 5.3.

Tabla 5.3. Herramientas tecnológicas seleccionadas

Componente	Herramienta	Justificación
Base de datos	SQL Server 2022	Robustez empresarial; soporte OLTP/OLAP
ETL	Python (pyodbc, Pandas)	Flexibilidad y automatización
Analítica	Python (scikit-learn, statsmodels)	Minería de datos y modelos predictivos
Visualización	Streamlit + Plotly	Tableros interactivos
Asistente	API de Anthropic Claude	Consultas en lenguaje natural

Fuente: elaboración propia.

Aplicación de metodologías ágiles

La implementación se guio por principios de metodologías ágiles, particularmente Scrum, con el fin de asegurar adaptabilidad y mejora continua. El desarrollo se dividió en sprints incrementales que entregaron valor de forma progresiva: levantamiento de requerimientos e identificación de fuentes; diseño de la arquitectura de datos; construcción de los procesos ETL; desarrollo del modelo dimensional; construcción de los tableros y del

asistente; y validación con retroalimentación. Este enfoque permitió entregas incrementales y ajustes continuos según las necesidades del proyecto.

Beneficios esperados

- Centralizar la información empresarial en una única fuente confiable.
- Reducir los tiempos de generación de reportes.
- Mejorar la calidad de los datos.
- Facilitar el análisis multidimensional del negocio.
- Detectar patrones de compra y segmentar clientes.
- Optimizar la gestión de inventarios.
- Fortalecer la toma de decisiones basada en datos.

Hoja de ruta de implementación

Para asegurar una ejecución ordenada se estableció una hoja de ruta dividida en fases incrementales bajo metodología ágil, según la Tabla 5.4.

Tabla 5.4. *Cronograma de implementación por fases*

Fase	Actividad	Duración
1	Levantamiento de requerimientos	2 semanas
2	Identificación y análisis de fuentes	2 semanas
3	Diseño de la arquitectura técnica	2 semanas
4	Desarrollo de procesos ETL	4 semanas
5	Construcción del Data Warehouse	3 semanas
6	Diseño de tableros e indicadores	3 semanas
7	Pruebas y validación	2 semanas
8	Capacitación y despliegue	1 semana

Fuente: elaboración propia.

Matriz de riesgos

Toda implementación tecnológica presenta riesgos. La Tabla 5.5 identifica los principales y sus estrategias de mitigación.

Tabla 5.5. *Matriz de riesgos del proyecto*

Riesgo	Nivel	Estrategia de mitigación
Baja calidad de datos fuente	Crítico	Implementar validaciones en el ETL
Resistencia al cambio	Alto	Capacitación y gestión del cambio
Retrasos en la integración	Medio	Uso de Scrum y sprints

Riesgo	Nivel	Estrategia de mitigación
Sobrecarga de infraestructura	Medio	Evaluación previa de capacidad
Falta de adopción por usuarios	Alto	Tableros intuitivos y capacitación
Cambios en requerimientos	Alto	Gestión ágil del backlog

Fuente: elaboración propia.

Consideraciones finales de la propuesta

La integración de estos elementos permite establecer no solo un diseño conceptual de la solución, sino también una visión práctica de ejecución, validación y sostenibilidad. La incorporación de la arquitectura, la planificación de la implementación y el análisis de riesgos fortalece la viabilidad técnica del proyecto y refleja una planificación integral alineada con las buenas prácticas de gestión de proyectos tecnológicos. La propuesta descrita en este capítulo se concretó en un prototipo funcional cuyos resultados se presentan en el capítulo siguiente.

Capítulo 6. Implementación y Resultados

Este capítulo presenta los resultados obtenidos tras implementar el esquema propuesto en un prototipo funcional y poblarlo con datos sintéticos representativos de una distribuidora de insumos industriales. Los análisis se ejecutaron sobre el almacén de datos ya cargado, y sus cifras provienen directamente de consultas al almacén.

Descripción del sistema implementado y del conjunto de datos

El prototipo se construyó sobre SQL Server 2022, con la base transaccional (SumIndustrial_OLTP) y el almacén analítico (SumIndustrial_DW) en bases separadas. El procesamiento se realizó en Python (Pandas, scikit-learn, statsmodels y SQLAlchemy), la interfaz en Streamlit con Plotly, y el asistente conversacional sobre la API de Anthropic Claude. La construcción incluyó una migración deliberada: la plataforma partió de un dominio análogo de comercio electrónico y se reescribió por completo hacia el dominio industrial, lo que exigió regenerar el catálogo, corregir varios errores del proceso ETL y volver a poblar el almacén; entre esas correcciones estuvo la generación de la dimensión de tiempo a partir de las fechas reales de las órdenes.

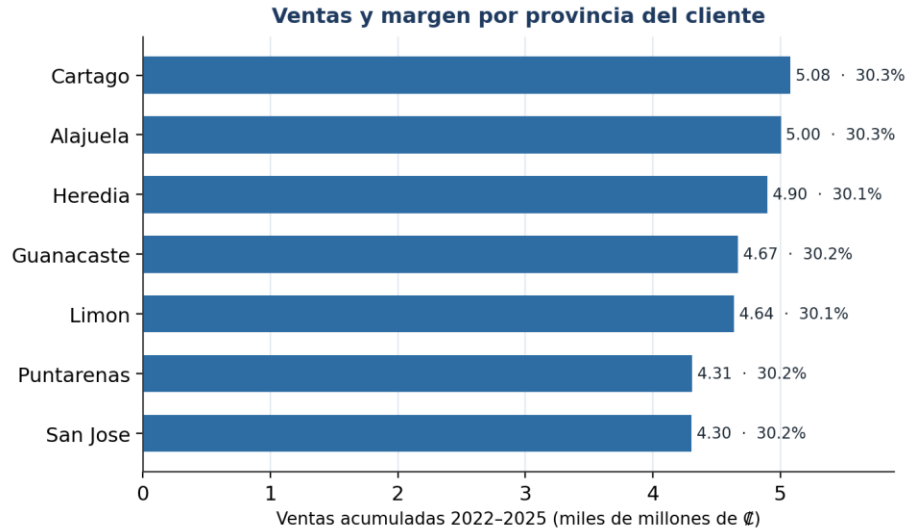
El almacén quedó conformado por tres tablas de hechos y seis dimensiones. Las tablas de hechos reunieron 32.383 líneas de orden, 72.048 líneas de cotización y 450.000 registros de inventario mensual, sobre un catálogo de 2.500 referencias, 3.000 clientes

corporativos, 30 proveedores y cinco sucursales, para el período 2022–2025. En conjunto, las órdenes no canceladas sumaron ₡32.901.986.330 en ventas, con un margen bruto del 30,2 %, 7.700 órdenes procesadas, un ticket promedio de ₡1.070.715 y 476.107 unidades vendidas.

Análisis OLAP multidimensional

La exploración OLAP confirmó un negocio geográficamente equilibrado. La Figura 6.1 presenta las ventas acumuladas y el margen por provincia del cliente: Cartago encabeza con ₡5.080 millones y San José cierra con ₡4.302 millones, dentro de un rango estrecho y con un margen prácticamente uniforme cercano al 30 % en todas las regiones, lo que sugiere una política de precios centralizada.

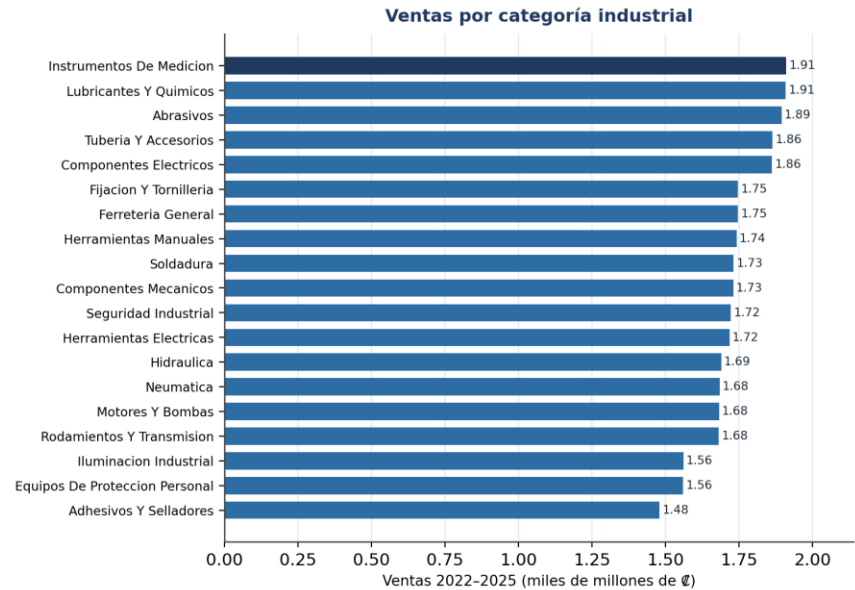
Figura 6.1. *Ventas y margen por provincia del cliente, 2022–2025*



Fuente: elaboración propia. El segundo valor de cada barra indica el margen porcentual.

Por categoría, las ventas se distribuyen de forma relativamente homogénea entre las veinte líneas industriales (Figura 6.2). Instrumentos de medición lidera con ₡1.910 millones, seguido de lubricantes y químicos y de abrasivos; los márgenes oscilan entre el 29,6 % y el 31,0 %.

Figura 6.2. Ventas por categoría industrial, 2022–2025



Fuente: elaboración propia.

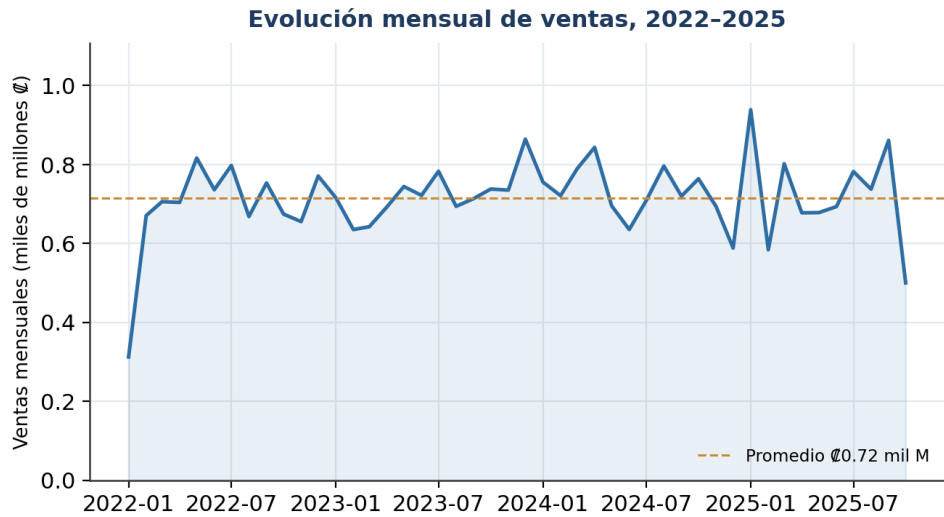
En el plano temporal, la Tabla 6.1 resume la evolución anual: las ventas se mantuvieron alrededor de ₡8.500 millones por año entre 2022 y 2024 y descendieron en 2025 (año que cubre hasta noviembre), con un margen estable en torno al 30,2 %. La Figura 6.3 muestra la serie mensual completa.

Tabla 6.1. Evolución anual de ventas, órdenes y margen

Año	Órdenes	Ventas (₡)	Margen %
2022	1.932	8.262.793.465	30,1
2023	2.053	8.677.830.349	30,2
2024	2.016	8.708.204.010	30,2
2025	1.699	7.253.158.506	30,2
Total	7.700	32.901.986.330	30,2

Fuente: elaboración propia. Cifras de órdenes no canceladas.

Figura 6.3. Evolución mensual de ventas, 2022–2025



Fuente: elaboración propia. La línea discontinua marca el promedio mensual.

Embudo de conversión de cotizaciones

En un negocio B2B la venta nace de una cotización, por lo que su conversión es un indicador central. De las 18.000 cotizaciones emitidas, 8.100 se convirtieron en orden, lo que arroja una tasa de conversión del 45,0 %, con un tiempo medio de respuesta de 16,5 días (Tabla 6.2). Más de la mitad de las cotizaciones (y del monto asociado) no llega a facturarse, lo que dimensiona el margen de mejora comercial que ofrece la gestión del pipeline.

Tabla 6.2. Embudo de conversión de cotizaciones

Estado	Cotizaciones	% del total	Monto cotizado (₡)
Convertida	8.100	45,0	40.446.315.715
Rechazada	3.240	18,0	16.089.483.999
Vencida	2.700	15,0	13.395.486.717
Pendiente	2.160	12,0	10.781.366.453
Aprobada	1.800	10,0	8.962.521.350

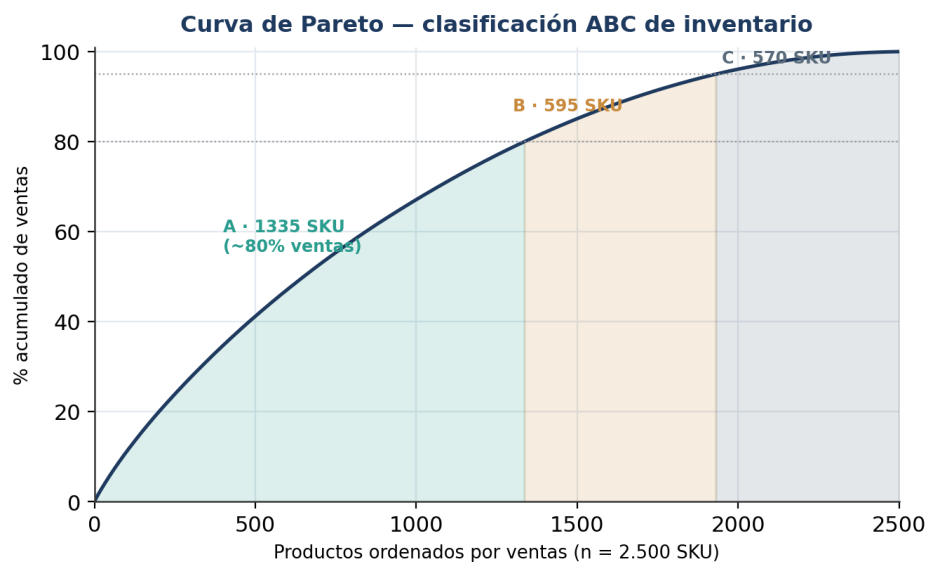
Fuente: elaboración propia.

Clasificación ABC/XYZ del inventario

La gestión de inventario se apoyó en dos clasificaciones complementarias. La clasificación ABC, basada en la contribución acumulada a las ventas, asignó 1.335 referencias a la clase A (que concentra el 80 % de las ventas), 595 a la clase B y 570 a la clase C (Figura 6.4). La clasificación XYZ, basada en la variabilidad de la demanda

mensual, distinguió tres grupos que orientan las políticas de reposición y el stock de seguridad (Tabla 6.3).

Figura 6.4. Curva de Pareto de la clasificación ABC de inventario



Fuente: elaboración propia sobre 2.500 referencias.

Tabla 6.3. Clasificación XYZ por variabilidad de la demanda

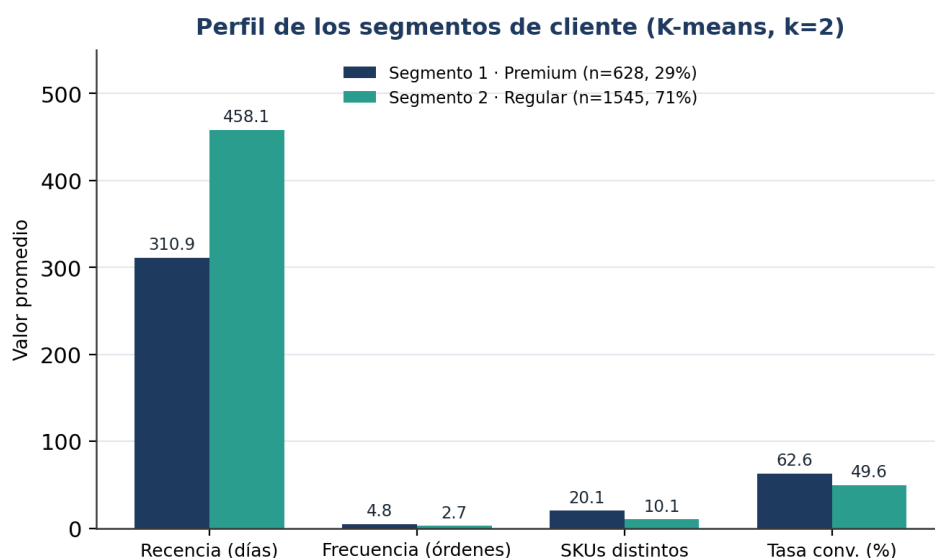
Clase	Referencias	Criterio (coef. de variación)	Estrategia sugerida
X	500	$CV < 0,5$ (demanda estable)	Reposición automática
Y	818	$0,5 \leq CV < 1,0$ (variable)	Revisión periódica
Z	1.182	$CV \geq 1,0$ (irregular)	Stock de seguridad alto

Fuente: elaboración propia.

Segmentación de clientes mediante K-means

La segmentación se aplicó sobre los 2.173 clientes con dos o más órdenes. Los tres criterios de selección (codo, silueta y Davies-Bouldin) coincidieron en que dos grupos constituyen la solución más estable (silueta = 0,34; Davies-Bouldin = 1,13). El Segmento 1, de mayor valor, agrupa al 28,9 % de los clientes, con mayor frecuencia de compra y una tasa de conversión del 62,6 %; el Segmento 2, mayoritario (71,1 %), reúne clientes menos frecuentes y con mayor recencia (Figura 6.5 y Tabla 6.4). Para la empresa, esto implica que cerca de dos de cada tres clientes activos son cuentas de baja frecuencia con potencial de reactivación, mientras que un núcleo minoritario concentra el valor.

Figura 6.5. Perfil comparado de los segmentos de cliente (K-means, $k = 2$)



Fuente: elaboración propia.

Tabla 6.4. Caracterización de los segmentos de cliente

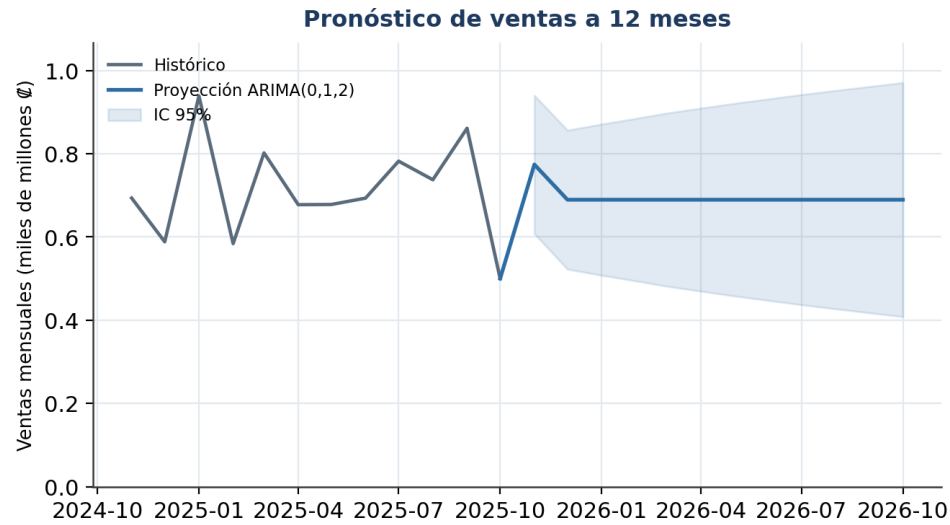
Segmento	Clientes	%	Recencia (días)	Frecuencia	Monto total (C\$)	Conversión %
1 · Premium	628	28,9	311	4,78	22.596.683	62,6
2 · Regular	1.545	71,1	458	2,66	10.471.005	49,6

Fuente: elaboración propia.

Pronóstico de demanda con ARIMA

El pronóstico se construyó sobre la serie mensual de ventas. La prueba de Dickey-Fuller aumentada rechazó la presencia de raíz unitaria ($p < 0,001$) y la búsqueda por criterio de Akaike seleccionó un modelo ARIMA(0,1,2). Sobre el conjunto de prueba de seis meses, el modelo alcanzó un error porcentual absoluto medio del 13,8 % (RMSE = C\$112,7 millones; MAE = C\$87,4 millones). La proyección a doce meses se estabiliza alrededor de C\$689 millones mensuales, con intervalos de confianza que se amplían de forma progresiva (Figura 6.6).

Figura 6.6. *Pronóstico de ventas a doce meses con intervalos de confianza al 95 %*



Fuente: elaboración propia. Modelo ARIMA(0,1,2).

Modelo de regresión

Se entrenó un modelo de regresión por bosque aleatorio para estimar el número de transacciones por grupo de tiempo, categoría, sucursal y vendedor. Durante el desarrollo se detectó una fuga de información: en una primera corrida, el margen total (una transformación directa de las ventas) dominaba el modelo con una importancia del 96 % e inflaba artificialmente el ajuste. Al confirmar el problema, se excluyeron el margen y el monto totales del conjunto de predictores. Tras la corrección, el modelo explicó el 31,7 % de la varianza en el conjunto de prueba (R^2 de entrenamiento = 0,72; validación cruzada = 0,31 $\pm$ 0,01); la caída del R^2 respecto de la primera corrida es, precisamente, la evidencia de que el modelo inicial estaba contaminado. Las variables de mayor importancia se muestran en la Tabla 6.5.

Tabla 6.5. *Importancia relativa de variables en el modelo de regresión (sin fuga)*

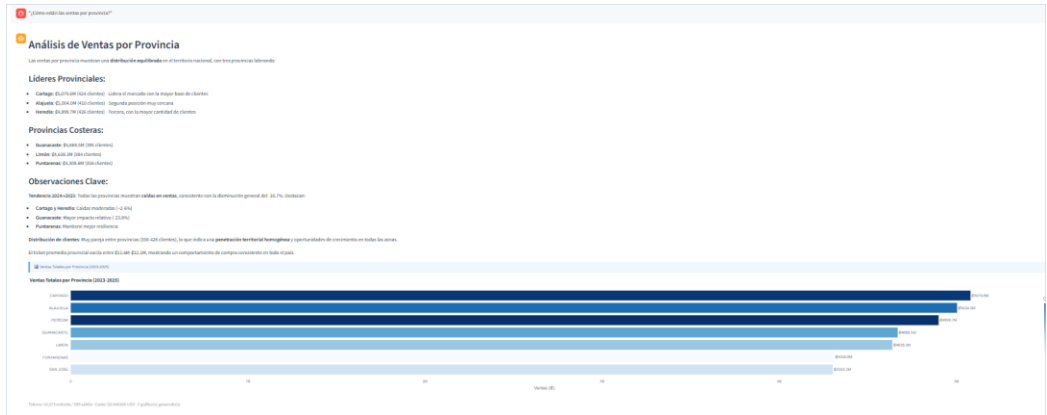
Variable	Importancia	Variable	Importancia
Unidades totales	0,453	Vendedor	0,049
Descuento promedio	0,106	Mes	0,041
Precio promedio	0,101	Día de la semana	0,038
Cantidad promedio	0,090	Sucursal	0,031
Categoría	0,059	Año	0,021

Fuente: elaboración propia. Variable objetivo: número de transacciones.

Asistente conversacional y verificación

El asistente conversacional constituye la capa de acceso más externa del sistema: traduce una pregunta en lenguaje natural a una consulta sobre el modelo dimensional, la ejecuta a través del cubo OLAP sobre el almacén poblado por el ETL y devuelve una respuesta redactada junto con una visualización automática. Para comprobar que las cifras provienen del almacén y no del modelo de lenguaje, se formularon preguntas cuyas respuestas podían contrastarse con los resultados ya tabulados. La Figura 6.7 muestra una de esas interacciones, con la respuesta, el gráfico y el detalle de consumo de la API.

Figura 6.7. Interacción con el asistente: consulta en lenguaje natural, respuesta y visualización



Fuente: captura del prototipo. El pie de la respuesta detalla el consumo de la API (tokens y costo en USD).

La Tabla 6.6 resume la verificación. En las tres consultas cruzables con el resto del capítulo las cifras coincidieron; una cuarta, sobre órdenes canceladas (un dato no tabulado), devolvió un total de 400 cancelaciones (tasa del 4,9 %) obtenido en vivo del almacén, y el propio asistente advirtió que el desglose anual era una estimación proporcional y no un valor consultado. Cada consulta con su visualización representó un costo por consumo de tokens cercano a los USD 0,03–0,04.

Tabla 6.6. Verificación de las respuestas del asistente contra los datos del estudio

Consulta en lenguaje natural	Respuesta del asistente	Contraste	Resultado
¿Cómo están las ventas por provincia?	Cartago lidera con ₡5.079,6 M	Figura 6.1	Coincide
Comparación 2023 vs 2024 vs 2025	₡8.677,8 / 8.708,2 / 7.253,2 M	Tabla 6.1	Coincide
Análisis de márgenes por categoría	Fijación y Tornillería, 31,0 %	Figura 6.2	Coincide
¿Cuántas órdenes se cancelaron por año?	400 canceladas; tasa 4,9 %	Dato no tabulado (en vivo)	Verificada

Fuente: elaboración propia.

Conclusiones

El proyecto diseñó, implementó y validó un esquema de integración y análisis de datos para una distribuidora de insumos industriales, con lo que se cumplió el objetivo general. El trabajo evolucionó desde una propuesta conceptual hacia un prototipo funcional que produce resultados analíticos reales sobre un almacén de datos poblado, y demostró que responde a la brecha diagnosticada entre la generación de datos y su aprovechamiento estratégico.

En relación con los objetivos específicos, el diagnóstico permitió caracterizar las fuentes de datos del negocio y confirmar una madurez analítica baja en cuatro dimensiones (datos, tecnología, procesos y cultura). Sobre esa base, la arquitectura por capas, que separa el entorno transaccional del analítico, centralizó la información y habilitó su explotación sin penalizar la operación diaria. El modelo dimensional en esquema estrella, organizado como una constelación de hechos con seis dimensiones compartidas, facilitó la exploración multidimensional del negocio desde perspectivas que el usuario reconoce con naturalidad.

El esquema de análisis con inteligencia artificial se concretó en cinco análisis y un asistente conversacional. La exploración OLAP describió el desempeño por región, categoría y período; la clasificación ABC/XYZ ordenó el catálogo para la gestión de inventario; la segmentación mediante K-means distinguió dos arquetipos de cliente; el embudo de cotizaciones cuantificó una conversión del 45 %; y el pronóstico ARIMA proyectó la demanda con un error razonable. El asistente, además, tradujo preguntas en lenguaje natural a consultas ejecutadas sobre el almacén, y sus respuestas coincidieron con los datos tabulados, lo que confirmó que recupera cifras reales y no verosímiles. La incorporación de metodologías ágiles guió el desarrollo incremental del prototipo, y el conjunto de resultados evidenció la viabilidad técnica y organizacional de la propuesta.

La contribución del trabajo es doble. Por un lado, aporta evidencia de diseño en un dominio poco documentado (la distribución industrial B2B multisede) frente al predominio del comercio minorista en la literatura, incorporando la lectura del pipeline de cotizaciones y la clasificación ABC/XYZ del inventario. Por otro, muestra que el valor de una solución de inteligencia de negocios depende tanto de la accesibilidad de su interfaz como de la sofisticación de sus técnicas.

El estudio reconoce sus límites. Se trata de una propuesta validada sobre datos sintéticos y un único caso, por lo que sus cifras ilustran el comportamiento del esquema y no el de un mercado real. La clasificación ABC resultó poco concentrada por la dispersión uniforme de la demanda sintética; el modelo de regresión explicó una fracción moderada de la varianza tras corregir la fuga de información; y el pronóstico ARIMA asume que la dinámica histórica se mantiene. Estas limitaciones trazan las líneas futuras: instanciar el esquema sobre datos reales de la empresa en una fase de prototipo productivo, enriquecer el pronóstico con variables externas de demanda sectorial, incorporar inferencia causal para medir el efecto de las decisiones comerciales, consolidar una función de gobernanza de datos y dimensionar el costo por consumo de tokens del asistente al escalar la solución a muchos usuarios.

Referencias

- Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM SIGMOD Record*, 26(1), 65–74.
<https://doi.org/10.1145/248603.248616>
- Golfarelli, M., & Rizzi, S. (2009). *Data warehouse design: Modern principles and methodologies*. McGraw-Hill.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3.^a ed.). Morgan Kaufmann.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2018). *Metodología de la investigación* (6.^a ed.). McGraw-Hill.
- Inmon, W. H. (2005). *Building the data warehouse* (4.^a ed.). Wiley.
- Kimball, R., & Ross, M. (2013). *The data warehouse toolkit: The definitive guide to dimensional modeling* (3.^a ed.). Wiley.
- McKinney, W. (2010). Data structures for statistical computing in Python. En *Proceedings of the 9th Python in Science Conference* (pp. 51–56).
<https://doi.org/10.25080/Majora-92bf1922-00a>
- Microsoft. (2022). *SQL Server 2022 documentation*. Microsoft Learn.
<https://learn.microsoft.com/sql/sql-server/>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Plotly Technologies Inc. (2023). *Plotly Python graphing library documentation*.
<https://plotly.com/python/>
- Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media.
- Sharda, R., Delen, D., & Turban, E. (2020). *Business intelligence, analytics, and data science: A managerial perspective* (4.^a ed.). Pearson.
- Silberschatz, A., Korth, H. F., & Sudarshan, S. (2020). *Database system concepts* (7.^a ed.). McGraw-Hill.

Streamlit Inc. (2023). *Streamlit documentation*. <https://docs.streamlit.io/>