

Resumen Ejecutivo

Implementación de Analítica a Gran Escala

Global Retail Corp

Edgar Daniel Diaz Nava



Sección I

Metodología de trabajo

(Scrum + DevOps + DataOps)



Introducción del Proyecto

Objetivo general

Generar un pipeline de datos que convierta **información cruda en insights de negocio** útiles para plantear estrategias de retención y crecimiento de ventas a agentes de negocio, cumpliendo buenas prácticas y gobernanza de datos.

Enfoque Ágil

El enfoque **Scrum** permite gestionar el desarrollo iterativo y colaborativo del pipeline, integrando continuamente feedback del negocio, mientras que **DataOps** garantiza la calidad, trazabilidad y automatización del ciclo de vida del dato.

Relevancia al negocio

Este proyecto permite a la cadena retail anticiparse a la demanda y **entender patrones de comportamiento** en combinaciones clave de tienda–producto–cliente, se **fortalecen las decisiones** comerciales con datos confiables y actualizados.

Selección de metodologías

SCRUM

Marco ágil que organiza el trabajo en ciclos cortos (sprints), **permitiendo entregar valor incremental** con retroalimentación continua del negocio.

DataOps

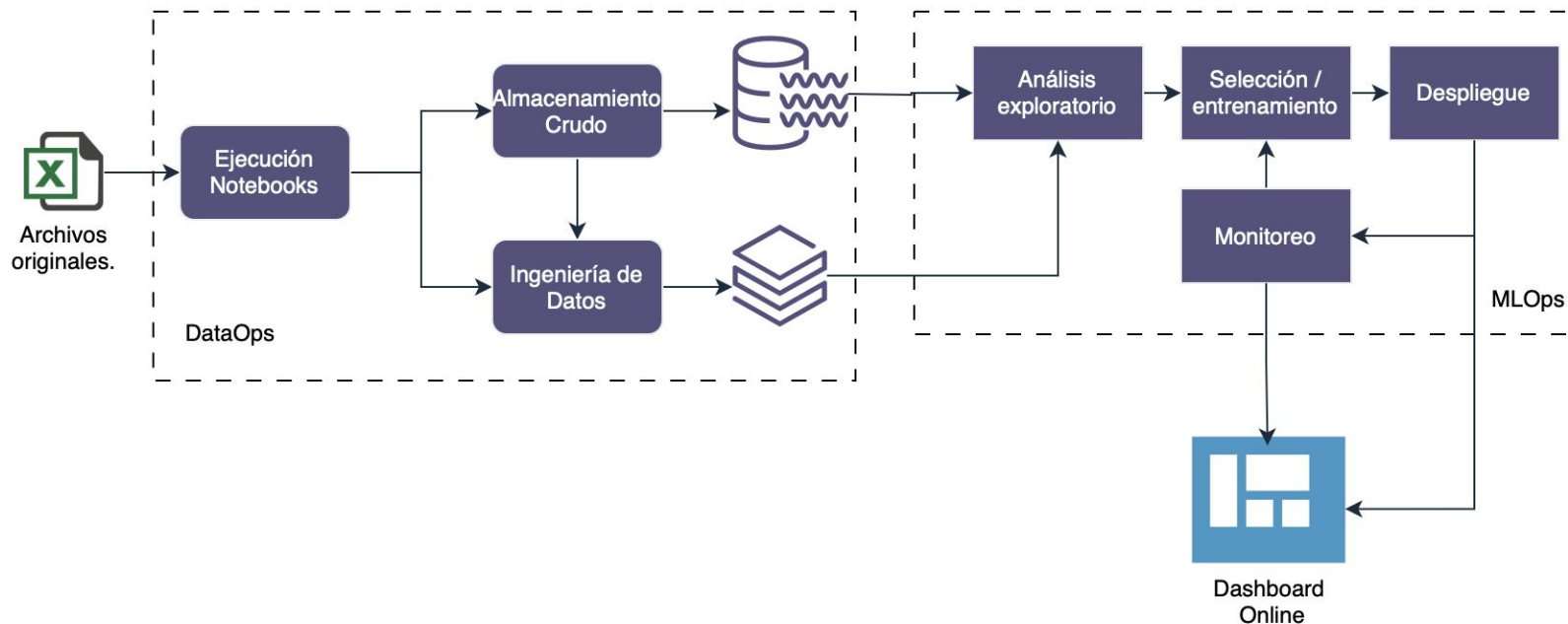
Conjunto de prácticas para **automatizar y asegurar la calidad** del ciclo de vida de los datos, desde la ingesta hasta el consumo analítico.

DevOps

Metodología que integra desarrollo y operaciones para **facilitar la entrega continua**, confiable y segura de productos de software y modelos.

Scrum organiza el trabajo en sprints iterativos con entregas constantes de valor, alineando a los equipos técnicos y de negocio. DataOps asegura que los datos fluyan con calidad y trazabilidad en cada etapa del pipeline, mientras que DevOps automatiza el despliegue y mantenimiento del modelo y dashboard en producción, garantizando velocidad, seguridad y estabilidad operativa.

Ciclo de vida del proyecto



Sección II

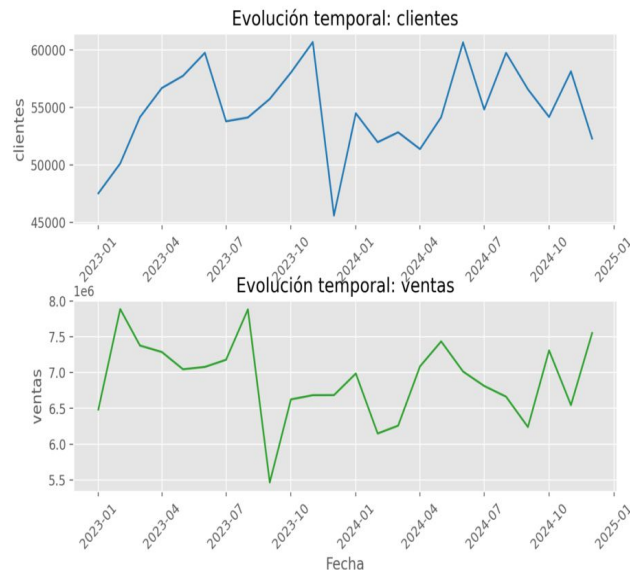
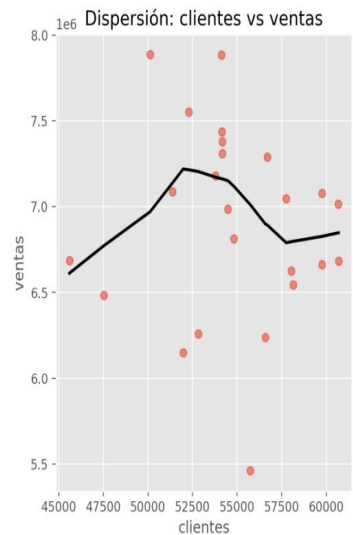
Resultados del EDA

Visualizaciones y KPIs



KPIs de negocio

- El desempeño de las tiendas y los productos muestran una tendencia cíclica, pero independiente unos de otros.
- En cuanto a la relación entre clientes y ventas, se ve que no hay una relación estadísticamente fuerte entre una y otra, lo que sugiere indagar otras variables,
- Más del 50% de los clientes representan al menos 100 ventas.



Sección III

Producto Final

Estrategia de Implementación



Mockup: Primer implementación

Se propone la construcción de un Dashboard que contenga los siguientes secciones:

- Análisis Exploratorio
- Predicciones
- Gobierno de datos

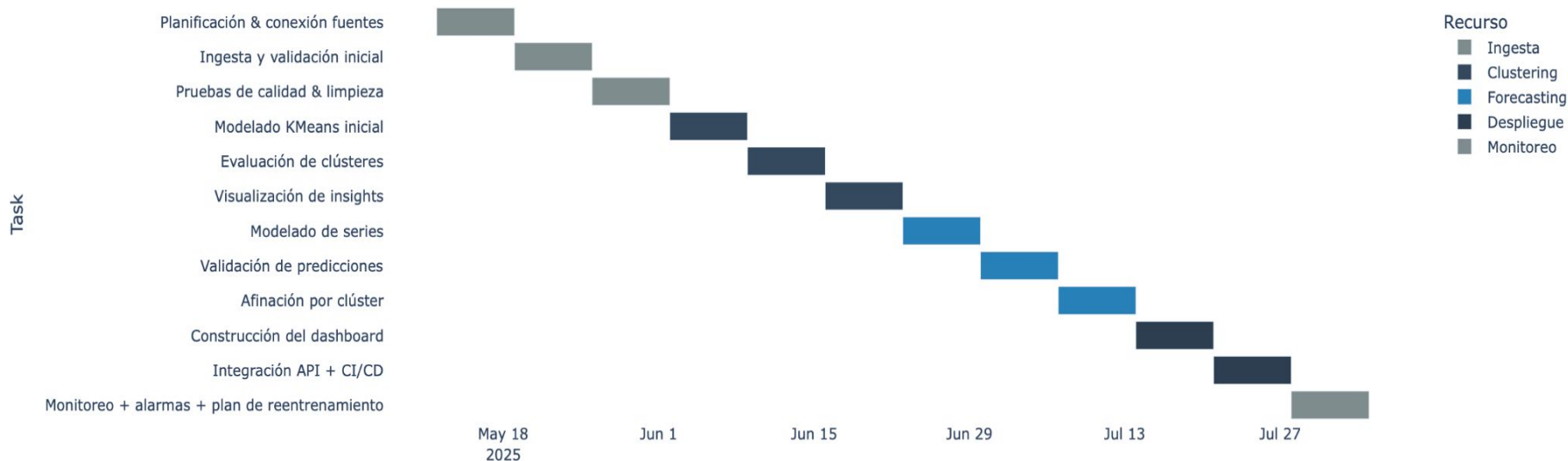
Así se cumple con una función de análisis, planeación y gobernanza de datos.



Una implementación de la propuesta se puede ver [aquí](#).

Plan de Implementación

Proyecto de Desarrollo Analítico a Gran Escala



Gobierno de Datos y Seguridad

Se implementa un **control de versiones** sobre los datos históricos utilizados y los modelos generados (clustering y forecasting), permitiendo **reproducibilidad total** del pipeline y facilitando auditorías.

Cada dataset, variable clave y modelo entrenado se documenta en un **Catálogo de Datos**, incluyendo origen, transformación, frecuencia de actualización y responsables. Garantizando **transparencia y alineación**.

Los modelos serán **reentrenado automáticamente** de forma mensual o al detectar cambios significativos en los patrones de los datos (**data drift**).

Se establecerán triggers para lanzar **validaciones cruzadas y comparar** el rendimiento del nuevo modelo.

Se aplica una **lista de verificación** para asegurar que el pipeline **respete regulaciones de privacidad** (como la LFPDPPP)

Sección IV

Cierre

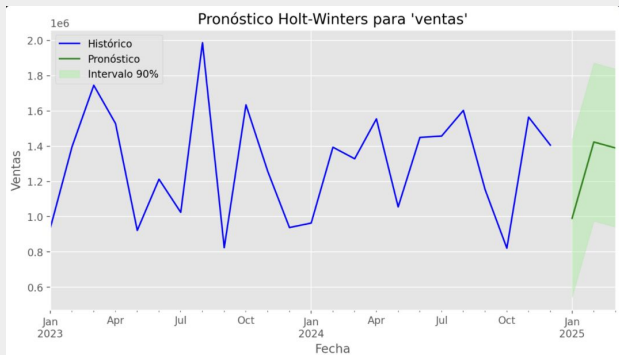
Propuesta Estratégica



Propuesta aplicada

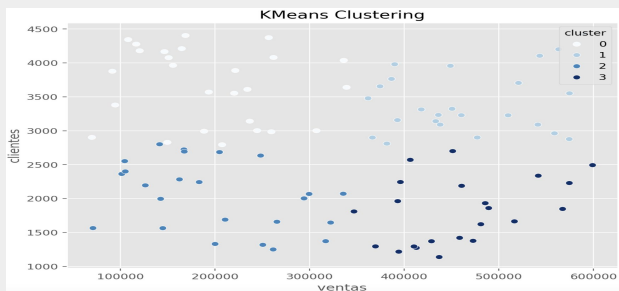
Pronóstico

Pronóstico de ventas y clientes que permitan anticipar escenarios y plantear estrategias.



Clusters

Clasificar regiones-productos según su desempeño en cada mes.



Escalabilidad

Los pronósticos para las ventas y clientes se pueden extender a diferentes niveles de granularidad, regiones y productos.

La clusterización se puede extender y su base en Spark permite el escalamiento a más datos.