

Sprint #3

Grupo 4

MIEMBROS DEL EQUIPO



Juan Ignacio Berti

- Científico de Datos
- jignacio.berti@gmail.com

Agustín Chienello

- Analista de Datos
- aguschienello27@gmail.com

M. Soledad García Ortiz

- Ingeniería de Datos
- mariasoledad.garciaortiz@gmail.com

Mario Suaza

Mentor

La entrada exitosa en un nuevo mercado de e-commerce requiere una comprensión profunda de diversos factores, como los tiempos de envío, estrategias de promoción efectivas y la selección estratégica de productos.

El **objetivo principal** de este proyecto es realizar una predicción precisa de las ventas y llevar a cabo un análisis de los métodos de pago más utilizados por los clientes. Además, se buscará comprender la distribución geográfica de los vendedores para asegurar tiempos de entrega predecibles y eficientes.

Objetivos Específicos:

- **Predicción de las ventas**
- **Análisis de los estados con mayores ventas**
- **Análisis de los métodos de pago más utilizados**
- **Análisis de la satisfacción del cliente**

Año

Todas

Categoria

Todas



Menú Principal

Total de ventas

\$1.415.881.547

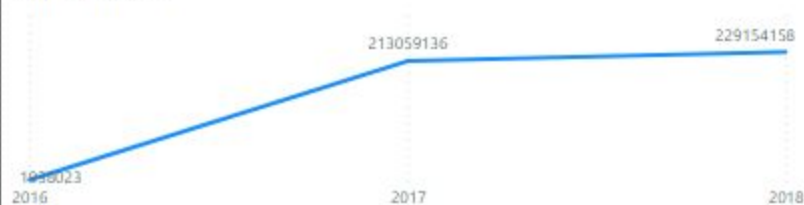
TOP 10 categorías más vendidos

\$848.113.678

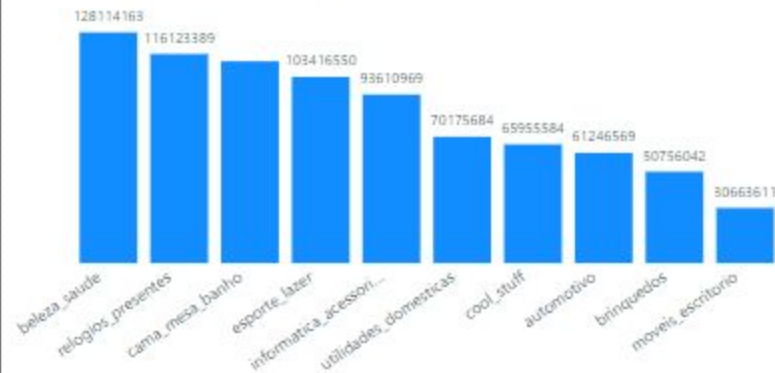
TOP 10 categorías menos vendidas

\$1.499.896

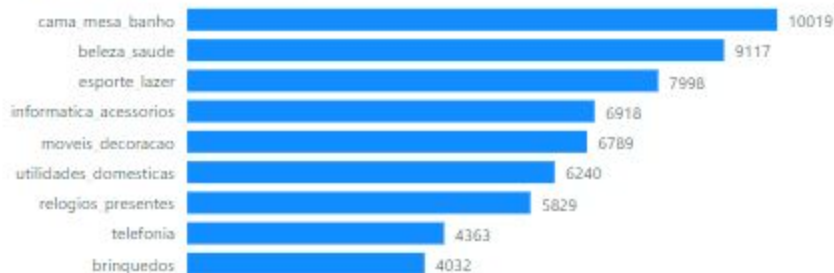
Ventas por Año



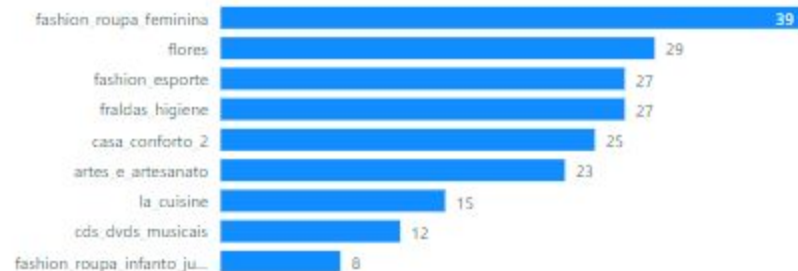
TOP 10 Total de ventas por categoría



TOP 10 Categorias más vendidas



TOP 10 Categorias menos vendidas

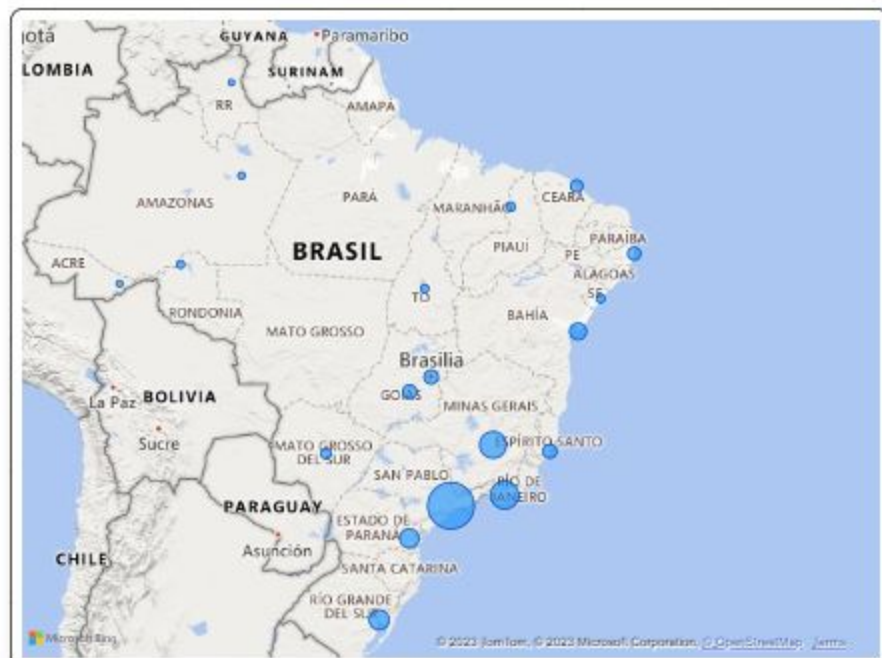




VENDEDORES POR ESTADO



ÓRDENES POR ESTADO



Año

Todas

Categoría

Todas



Menú Principal

KPI Órdenes procesadas

103886

Tasa de conversión de pagos

1,04

Promedio de puntaje de clientes

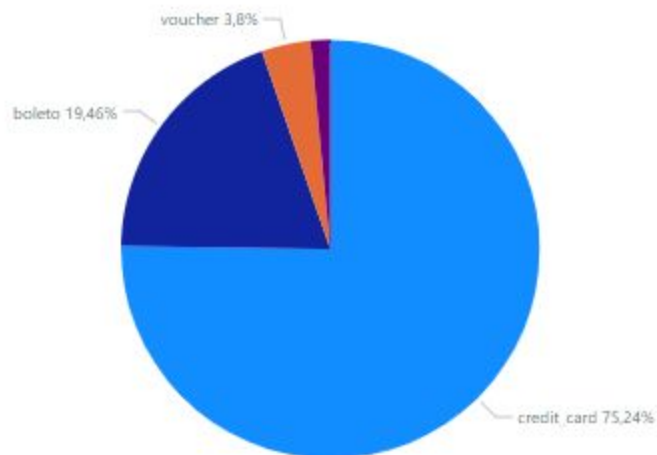
4,09

Promedio de tiempo de entrega (Días)

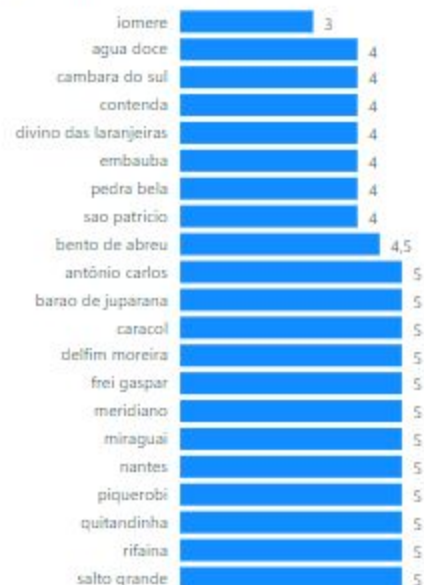
12,50

Medios de pago

● credit_card ● boleto ● voucher ● debit_card ● not_defined

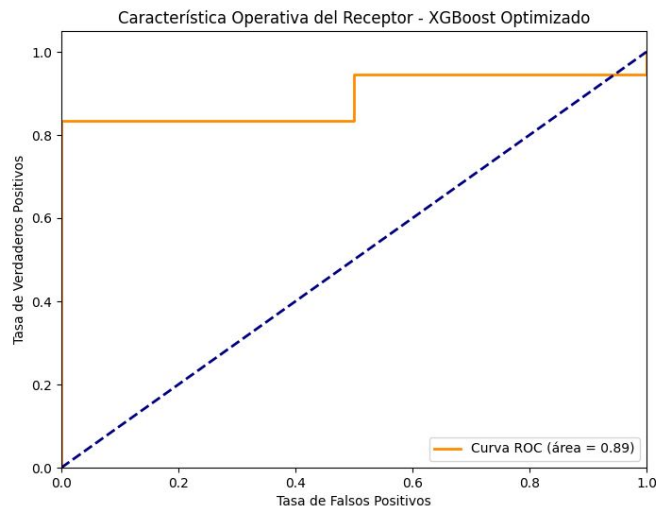


Entrega promedio en dias por ciudad



Modelo de Entrenamiento de PLN con XGBoost

- Proceso de PLN con tokenización y lematización
- Vectorización del texto de las opiniones de los usuarios con el paquete TfidfVectorizer
- Clasificación con XGBoost
- Alta precisión en la predicción de opiniones (>85%)



Consideraciones adicionales:

- En el dataset hay más de 300 reseñas relacionadas directamente con el tiempo de entrega.
- Es la variable con más correlación a la hora de clasificar una opinión como positiva o negativa.
- Es menester destinar recursos a la optimización de la logística del sistema de entrega y los canales de retroalimentación con las opiniones de los usuarios.
- El hecho de que haya tres veces más opiniones positivas que negativas evidencia un patrón pero la diferencia puede profundizarse si se emplean las recomendaciones esgrimidas en esta presentación.

Producto ML en Streamlit


Random Forest
R2=0.906

Menú

G4

Ventas

BUSINESS



INTELLIGENCE

G4

Objetivo principal

El objetivo principal de este proyecto es realizar una predicción precisa de las ventas y llevar a cabo un análisis de los métodos de pago más utilizados por los clientes. Además, se buscará comprender la distribución geográfica de los vendedores para asegurar tiempos de entrega predecibles y eficientes.

Objetivos específicos

- Análisis de los estados con mayores ventas
- Análisis de los métodos de pago más utilizados
- Análisis de la satisfacción del cliente
- Predicción de las ventas

Menú

G4

Ventas

Predicción de ventas

Introduce los datos para predecir las ventas

Las variables Medio de pago son binarias

1 si se utilizó ese medio de pago y 0 si no se utilizó

Las variables Categoría y Ciudad son categóricas

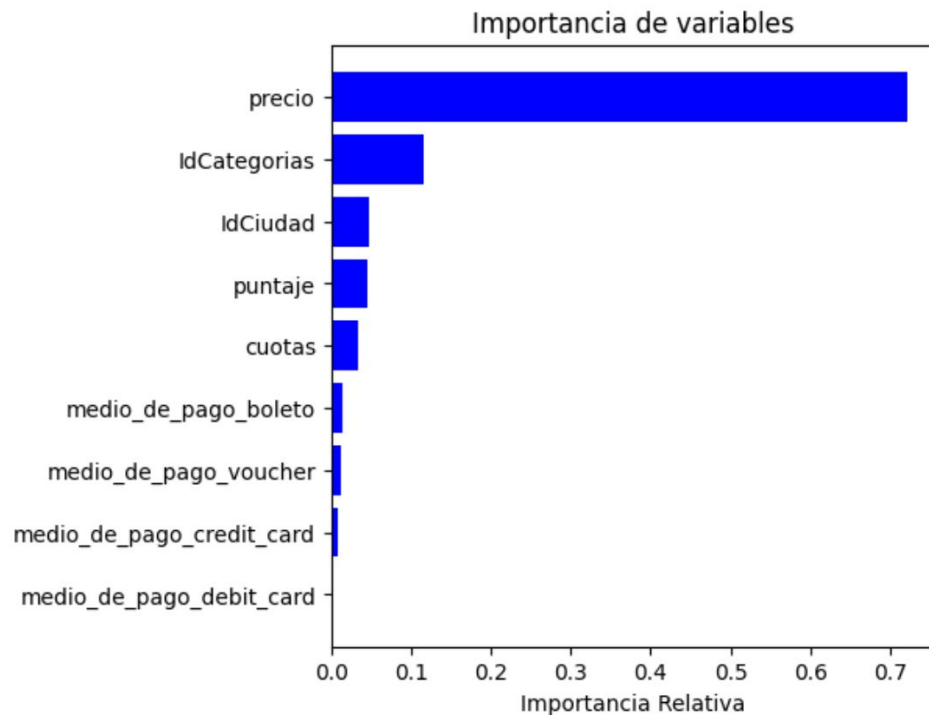
se debe ingresar el Id correspondiente

Datos de entrada

IdCategorias	medio_de_pago_boleto
0,00 - +	0,00 - +
precio	medio_de_pago_credit_card
0,00 - +	0,0 - +
cuotas	medio_de_pago_debit_card
0,00 - +	0,0 - +
puntaje	medio_de_pago_voucher
0,00 - +	0,0 - +
IdCiudad	
0,00 - +	

predecir

Variables de importancia



Recomendaciones:

Migración de datos a la nube (Cloud Computing) y automatización de procesos

Acceder mediante un sistema remoto al software, al procesamiento de datos y al almacenamiento de archivos.

- No hay costos iniciales de infraestructura
- Se puede pagar para obtener recursos adicionales solo si se necesitan y en el momento que se necesiten
- Inmediatez
- Concurrencia





Google Cloud Platform

Compute



Identity & Security



Networking



Big Data



Storage and Databases



Machine Learning



CARACTERÍSTICAS:

- Escalabilidad
- Seguridad
- Fiabilidad
- Herramientas de desarrollo
- Integración con otras herramientas de Google
- Precios competitivos

ALGUNAS VENTAJAS:

- Economía y control de costos
- Flexibilidad
- Multinube
- Backup seguro
- Control de calidad
- Recuperación ante desastres
- Servicio sostenible

Esta automatización no solo mejora la velocidad y la precisión del proceso, sino que también permite el monitoreo continuo y el mantenimiento, asegurando la seguridad y cumplimiento normativo.

Conclusiones/Plan de acción

Automatización de procesos de ETL

1

Contratar el servicio de GCP para almacenar, gestionar datos y automatizar el proceso de ETL

Optimización logística del sistema de entregas

2

El análisis de las opiniones de los usuarios arrojó una tendencia a valorar los tiempos de entrega por sobre el resto de las variables.

Mejor distribución geográfica de vendedores

3

Encontramos una tendencia muy marcada en la cual donde menos vendedores hay por región, se registran menor cantidad de ventas.

MUCHAS GRACIAS!

GRUPO 4

