

OPTIMIZACIÓN DE LAS OPERACIONES EN INSTACART MEDIANTE ANÁLISIS DE CANASTA DE MERCADO

Josué Gerardo Lozano Suárez 3081284

Paloma Garza Muñoz 03084376

Norberto Suaste Hernández 03088785

INTRODUCCIÓN

Instacart maneja millones de pedidos y productos. Comprender cómo compran sus clientes permite tomar mejores decisiones.

Este proyecto aplicó técnicas de análisis de canasta de mercado para identificar asociaciones frecuentes entre productos.

Con esta información, se propone una estrategia de mejora operacional y comercial para la empresa.



CONJUNTO DE DATOS

Los datos utilizados provienen del repositorio de análisis de mercado de Instacart e incluyen:

- **products.csv**: catálogo de productos.
- **orders.csv**: historial de pedidos de los clientes.
- **order_products__prior.csv**: productos incluidos en pedidos anteriores.

Se trabajó con los primeros 10,000 pedidos, lo cual permitió un análisis representativo sin comprometer el rendimiento computacional.



PROCESO DE ANÁLISIS

- **Integración de datos:** Se unieron tablas de productos y pedidos mediante merge para obtener una base consolidada.
- **Agrupación de transacciones:** Cada pedido se transformó en una lista de productos.
- **Codificación binaria (one-hot encoding):** Para aplicar el algoritmo Apriori, se creó una matriz donde cada fila es un pedido y cada columna un producto.
- **Aplicación de Apriori:** Se identificaron combinaciones de productos con alta frecuencia de compra conjunta.
- **Extracción de reglas de asociación:** Se obtuvieron reglas con alta confianza y valor de lift, que indican relaciones fuertes y útiles.

HALLAZGOS CLAVE



Existen productos que se compran juntos con alta frecuencia, como leche y cereales, o pan y mantequilla.

Algunas combinaciones tienen confianza superior al 60% y lift mayor a 2, lo que indica una relación fuerte entre los productos.

Estos patrones son valiosos para personalizar la experiencia de compra y tomar decisiones operativas.

PROPUESTAS DE MEJORA PARA INSTACART

Recomendaciones personalizadas

- Sugerir productos adicionales al cliente mientras realiza su compra, basados en patrones descubiertos. Esto puede aumentar el ticket promedio.

Promociones inteligentes

- Ofrecer descuentos en combos o productos relacionados, aumentando la probabilidad de compra conjunta.

Optimización de inventario

- Si dos productos se venden juntos, se puede prever su demanda conjunta y evitar quiebres de stock o sobreinventario.

Eficiencia logística

- Diseñar rutas de recolección (picking) que agrupen productos relacionados, reduciendo el tiempo y costo de preparación de pedidos.

IMPACTO ESPERADO

+Ingresos: Aumento del ticket promedio gracias a ventas cruzadas basadas en recomendaciones.

+Satisfacción del cliente: Compras más rápidas, completas y personalizadas.

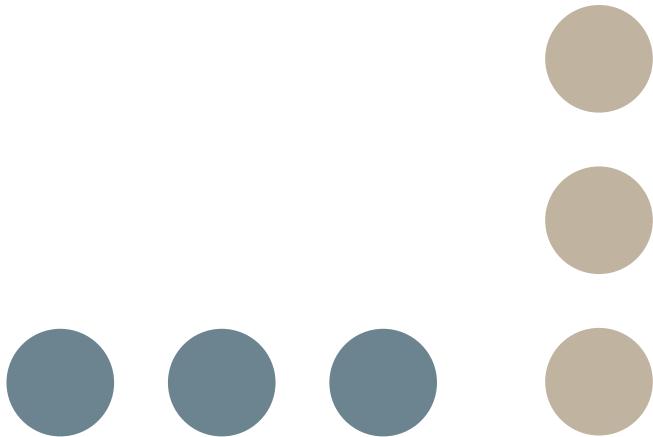
- Tiempos operativos: Reducción en el tiempo de recolección de productos en almacenes.

- Costos de inventario: Menor probabilidad de exceso o escasez de productos.

CONCLUSIONES DEL ANÁLISIS



- El análisis de canasta de mercado es una herramienta poderosa para convertir datos históricos en decisiones de negocio.
- A partir de simples transacciones se pueden obtener reglas valiosas que guían promociones, recomendaciones y logística.
- Es recomendable que Instacart integre estos análisis en sus sistemas y los actualice regularmente con nuevos datos.



PREGUNTAS FINALES

¿Cuáles fueron tus retos para hacer la preparación de los datos y cómo los resolviste?

Uno de los principales retos fue el tamaño del conjunto de datos. Dado que contiene millones de registros, fue necesario filtrar los primeros 10,000 pedidos para asegurar una ejecución eficiente. Además, integrar correctamente los datasets (productos, pedidos y productos por pedido) requería claridad en las claves primarias/foráneas. Para resolver esto, usamos merge con pandas asegurando que los campos coincidieran correctamente.



PREGUNTAS FINALES

¿Qué problemas se presentaron y cómo los resolviste al construir los modelos de datos para hacer los análisis solicitados?

Un desafío técnico fue la transformación de los datos a formato de "one-hot encoding" para usar el algoritmo Apriori. Este paso es crucial para representar las transacciones como una matriz binaria. Hubo que hacer un groupby adecuado y aplicar correctamente la codificación con pd.crosstab. También se filtraron reglas con bajos valores de soporte o confianza para obtener asociaciones realmente útiles.



PREGUNTAS FINALES

¿Qué aprendiste, cuáles son tus conclusiones del proyecto y qué tienes que mejorar?

Aprendímos a aplicar técnicas de minería de datos en un caso real y masivo. Vimos cómo los datos pueden aportar valor directamente al negocio. Mis conclusiones son que el análisis de canasta de mercado es una herramienta poderosa para personalizar la experiencia del cliente y optimizar la operación.



Gracias!