

Análisis y Modelado Analítico de Transacciones de Supermercado

Objetivo General

Diseñar y desarrollar una **solución tecnológica integral** que permita analizar y visualizar el comportamiento de las transacciones de un supermercado, con el fin de generar valor a partir de los datos disponibles mediante analítica descriptiva y diagnóstica.

Contexto

Se cuenta con un **dataset de transacciones de supermercado** que contiene información como:

- ID de transacción
- Fecha y hora de la compra
- ID de cliente
- ID de producto
- Categoría del producto
- Cantidad comprada
- Tienda o punto de venta

Nota: No se dispone de precios ni montos de pago. Los estudiantes deberán generar indicadores y métricas relativas (por ejemplo, volumen de compras, frecuencia, diversidad de productos, etc.) que permitan realizar los análisis solicitados.

Entrega esperada

Los estudiantes deben presentar una **solución tecnológica funcional** que cumpla con los siguientes componentes:

Resumen Ejecutivo

Debe incluir visualizaciones y métricas que respondan a las siguientes preguntas:

Indicador	Descripción	Visualización sugerida
Total de ventas	Total de unidades vendidas (suma de cantidades)	Indicador numérico o gráfico de barra
Número de transacciones	Conteo total de transacciones registradas	Indicador numérico
Top 10 productos	Productos más comprados según volumen o frecuencia	Gráfico de barras horizontales
Top 10 clientes	Clientes con mayor número de compras	Gráfico de barras horizontales
Días pico de compra	Días con mayor número de transacciones	Serie de tiempo o heatmap diario
Categorías más rentables	Inferir rentabilidad según volumen o frecuencia relativa	Gráfico de pastel o barras

Visualizaciones Analíticas

Deben crearse las siguientes visualizaciones, con el objetivo de explorar la estructura y el comportamiento de los datos:

Tipo	Descripción	Objetivo
Serie de tiempo	Ventas por día o semana	Identificar tendencias y estacionalidad
Boxplot	Distribución de totales por cliente o categoría	Detectar outliers o comportamientos atípicos
Heatmap	Correlación entre variables numéricas (por ejemplo, frecuencia, cantidad promedio, diversidad de productos)	Explorar relaciones entre variables

Análisis Avanzado

Implementar los resultados deben estar acompañados de su interpretación:

A. Segmentación de Clientes

Aplicar **K-Means o clustering jerárquico** para segmentar clientes según su comportamiento de compra:

- Variables sugeridas: frecuencia, número de productos distintos, volumen total, diversidad de categorías.
- Entregar: visualización del clustering, descripción de cada grupo y recomendaciones de negocio.

B. Recomendador de Productos

Desarrollar un **modelo básico de recomendación** usando técnicas de filtrado colaborativo o reglas de asociación:

- **Dado un cliente:** sugerir productos complementarios o similares a los que ha comprado.
- **Dado un producto:** recomendar otros productos que suelen comprarse junto a él.

C. Generación de nuevos resultados

Implementar el sistema de tal manera que al incorporar nuevos datos como fuente de información se generen los resultados de análisis requeridos.

Entregables

1. **Código fuente/documento ejecutable:** Notebook o aplicación (Streamlit/Dash/Shiny/Power BI).
2. **Informe técnico:** PDF o Markdown con:
 - Descripción de los datos.
 - Metodología de análisis.
 - Principales hallazgos visuales.
 - Resultados del modelo de segmentación o recomendación.
 - Conclusiones y posibles aplicaciones empresariales.
3. **Presentación (5-10 min):** resumen de los resultados clave y demostración funcional.

Criterios de evaluación

Criterion	Weight
Clarity and quality of visualizations	20%
Depth of descriptive analysis	20%
Correct implementation of advanced analysis (clustering or recommender)	25%
Incorporation of new data	25%
Presentation and documentation	10%