

Proyecto: Análisis de Patrones de Ventas

Fecha de entrega: Lunes 1 de diciembre

Lenguaje y stack sugerido: Python, R y herramienta de visualización (Power BI / Looker Studio / Streamlit).

1) Objetivo

Analizar patrones y ciclicidades de ventas para:

- **Detectar tendencias y estacionalidad** por zona, sector y categoría.
- **Optimizar rotación de productos y asignación de recursos** (inventario, fuerza de ventas).
- **Anticipar demanda** a corto/mediano plazo mediante modelos predictivos.
- **KPIs base (ejemplos)**: Ventas totales, #unidades, ticket promedio, margen bruto, rotación (días de inventario), share por categoría-sector, % por método de pago, crecimiento interanual/mensual.

2) Enfoque de negocio

El análisis se orienta a decisiones en tres frentes:

1. **Abasto y surtido**: qué productos priorizar por temporada y zona.
2. **Eficiencia operativa**: reducción de costos por sobre-stock/stock-out y logística.
3. **Planificación comercial**: promociones y pricing en picos y valles de demanda.

Preguntas guía:

- ¿Qué categorías y zonas muestran **picos estacionales** y cuándo?
- ¿Dónde hay **oportunidad de reasignar inventario** sin afectar servicio?
- ¿Qué métodos de pago **crecen/caen** y cómo impactan el ticket y la frecuencia?

3) Datos y alcances

- **Fuente**: dataset de Kaggle (proporcionado)
<https://www.kaggle.com/datasets/dataregina/datasets-para-proyecto-bi?select=categorias.csv>
- **Suposiciones**: datos representativos, con cobertura temporal suficiente (≥ 12 meses) para estacionalidad.
- **Fuera de alcance (opcional)**: optimización logística avanzada, elasticidades de precio por product-SKU (si no hay precio).

4) Metodología (fases y entregables intermedios)

4.1 Recolección

- Descarga, diccionario de datos y **data contract** (nombres, tipos, llaves, granularidad).

4.2 Calidad, limpieza y preparación

- Detección y tratamiento de **nulos, duplicados, outliers**.
- **Normalización/estandarización** donde aplique; codificación de categorías.
- Construcción de **variables derivadas**: semana, mes, estacionalidad, RFM, margen, etc.
- **Bitácora de decisiones** (qué, por qué y cómo se transformó).

4.3 EDA (Análisis Exploratorio)

- Distribuciones, correlaciones, **series de tiempo** (tendencia, estacionalidad).

- Evolución por **zona/sector/categoría** y por **método de pago**.
- Mapas/heatmaps temporales (calendario) para ciclicidades.

4.4 Segmentación y Clustering (peso alto)

- **K-means / DBSCAN / GMM** para agrupar productos o zonas por patrón de demanda.
- Selección de k por **inercia/codo** y **silhouette**; justificación del algoritmo.
- Interpretación de clústeres: temporada fuerte, volatilidad, mix de pago, margen.

4.5 Modelado Predictivo (peso alto)

- Enfoques posibles (elegir y justificar):
 - **Regresión** (lineal múltiple/regularizada) para demanda por categoría/zona.
 - **Árboles/boosting** para importancia de variables y no linealidades.
 - **Series de tiempo** (ARIMA/ETS/Prophet) a nivel agregado o por clúster.
 - **Redes neuronales** (opcional si hay datos suficientes).
 - **Validación**: CV temporal / holdout, métricas (RMSE, MAPE, MAE), baseline vs modelo.
 -

4.6 Visualización y storytelling

- **Dashboard** con KPIs, tendencias y estacionalidad; filtros por zona/sector/categoría.
- Gráficas clave: descomposición de serie, estacionalidad, mapa de clústeres, importancia de variables.

4.7 Recomendaciones accionables

- **Surtido y compra**: top categorías por temporada/zonas y semanas de cobertura sugeridas.
- **Operación**: reasignación de inventario, ventanas de mantenimiento/descanso.
- **Comercial**: promociones por clúster/temporada; alertas tempranas de quiebre.

5) Criterios de evaluación (rubrica y pesos)

- **Calidad de datos y trazabilidad**: 15% (bitácora, decisiones reproducibles).
- **EDA y hallazgos sustantivos**: 15% (claridad, insight, narrativa).
- **Clustering (Fase 4.4)**: 30% (selección, métrica, interpretación de negocio).
- **Modelado predictivo (Fase 4.5)**: 30% (diseño, métricas, comparación con baseline).
- **Dashboard & storytelling ejecutivo**: 10% (usabilidad, claridad, foco en decisión).

6) Entregables

1. **Documento técnico** (PDF/MD) con:
 - a. Objetivo, datos, limpieza/feature engineering, EDA, clustering, modelos, resultados, limitaciones y **recomendaciones**.
2. **Código** (Python o R) **reproducible**:
 - a. Estructura de repo (src/, data_raw/, data_clean/, notebooks/, reports/).
 - b. README con instrucciones de ejecución y requerimientos.
 - c. Semillas y configuración para replicar resultados.
3. **Dashboard** (Power BI/Looker/Streamlit) con KPIs, filtros y visualizaciones clave.
4. **Presentación ejecutiva** (10–12 diapositivas):
 - a. Problema, enfoque, 3–5 hallazgos, impacto esperado, plan de acción.

7) Criterios de aceptación (checklist)

- Diccionario de datos y bitácora de limpieza.
- Métricas de calidad y evidencia de tratamiento de nulos/outliers.
- EDA con insights accionables (no solo gráficas).
- Clustering con justificación de hiperparámetros y lectura de negocio.
- Modelos comparados vs baseline, con métricas.
- Dashboard funcional y guía de uso.
- Recomendaciones concretas (qué, dónde, cuándo, cuánto).
- Código ejecuta “end-to-end”.