

Proyecto SF LKR (Análisis de Patrones de Ventas)

Luis A. Morales, Paulina D. Arroyo, Rodrigo J. Ortiz

Resumen

Este documento detalla un análisis técnico de patrones de ventas, segmentación de clientes y productos, así como modelado predictivo para apoyar decisiones estratégicas en inventario, marketing y operaciones. Se aplican técnicas de limpieza de datos, análisis exploratorio, clustering y modelos de regresión avanzada.

1 Objetivo

Detectar tendencias y patrones de demanda, clasificar productos y clientes según su comportamiento de compra y predecir los ingresos futuros mediante el uso de modelos estadísticos y técnicas avanzadas de machine learning. Todo ello con el propósito de fortalecer la toma de decisiones estratégicas en áreas clave del negocio, como el abasto, la planeación de surtido, la segmentación de clientes y la definición de estrategias de pricing. Estas capacidades analíticas permiten anticipar cambios en el mercado, optimizar niveles de inventario, identificar oportunidades de crecimiento, mejorar la rentabilidad y aumentar la eficiencia operativa en toda la cadena de valor.

2 Limpieza y Preparación de Datos

Para garantizar la calidad del análisis, se realizó un proceso de limpieza y preparación de datos. Se estandarizó el campo **Precio_Unitario** y se integraron las bases de ventas, clientes y productos mediante uniones internas. Se eliminaron columnas irrelevantes y se generaron variables clave como **Cantidad_dinero**. Además, se incorporaron datos de **Stock** y **Categoría**, lo que permitió obtener métricas agregadas como ventas por categoría. Se aplicó detección de outliers en

precios, conservando aquellos justificables. Finalmente, se estimó el stock restante y la cantidad total vendida por producto.

3 Análisis Exploratorio de Datos (EDA)

El análisis exploratorio fue fundamental para comprender la estructura de los datos, detectar patrones iniciales de comportamiento y validar hipótesis. Nos permitió identificar variables relevantes y entender la relación entre variables. Además, reveló diferencias significativas entre regiones, categorías y momentos temporales, lo cual sirvió como base para seleccionar variables en los modelos de clustering y predicción, así como para formular recomendaciones accionables en términos de marketing, inventario y estrategia comercial.

Se analizó la distribución de ingresos por venta (**Cantidad_dinero**), observándose una asimetría positiva: la mayoría de las transacciones se concentran en ingresos bajos, con algunos productos de alto valor que generan ingresos elevados.

La matriz de correlación reveló una fuerte relación entre **Cantidad_dinero**, **Precio_Unitario** ($r = 0.71$) y **Cantidad** ($r = 0.62$), lo que valida su uso como variable objetivo en modelos posteriores.

A nivel temporal, se detectaron variaciones en la cantidad de productos vendidos por día y patrones estacionales. Un análisis por región mostró que Buenos Aires lidera en ingresos, mientras que las categorías con mayores ventas fueron “Carnicería” y “Lácteos”.

El heatmap por día de la semana y mes destacó un mayor volumen de ventas los martes, viernes y domingos, así como en los meses de julio, septiembre y diciembre.

Finalmente, se comparó el stock inicial con el restante, identificando productos con baja rotación (sobre-stock) y productos estrella con alta salida, útiles para segmentación y optimización de inventario.

4 Clustering de Productos (KMeans, k=4)

Variables utilizadas: cantidad vendida, stock actual, porcentaje de venta.

- **Cantidad vendida:** refleja la demanda del producto.
- **Stock actual:** permite evaluar si existe sobre-stock o falta de inventario.
- **Porcentaje de venta:** mide eficiencia del producto al convertir inventario en ventas.

Estas variables permiten segmentar los productos según su comportamiento operativo y comercial, lo cual es esencial para tomar decisiones de inventario, promociones y rotación.

Resultados del clustering:

- **Cluster 0 - Sobre-stock:** Alta existencia pero baja salida. Riesgo financiero por acumulación.
- **Cluster 1 - Rotación Saludable:** Balance entre ventas y stock. Ideal para inventario base.
- **Cluster 2 - Estables:** Sin variaciones extremas. Requieren seguimiento, no intervención inmediata.

- **Cluster 3 - Estrella:** Alta rotación y demanda. Prioridad en campañas y disponibilidad continua.

5 Clustering de Clientes (KMeans, k=6)

Variables utilizadas: frecuencia de compra, gasto total, ticket promedio, variedad de categorías.

- **Frecuencia de compra:** mide lealtad y recurrencia.
- **Gasto total:** identifica a los clientes más rentables.
- **Ticket promedio:** diferencia entre alto valor por transacción o volumen por frecuencia.
- **Variedad de categorías:** indica amplitud del interés o especialización.

Estas variables permiten agrupar a los clientes según comportamiento económico y hábitos de consumo, lo cual facilita la personalización de estrategias comerciales.

Resultados del clustering:

- **Cluster 0 - Premium:** Alta frecuencia, ticket alto, variedad amplia. Ideales para programas de fidelidad.
- **Cluster 1 - Nicho:** Compradores de pocas categorías, pero frecuentes. Necesitan segmentación por línea de producto.
- **Cluster 2 - Estables:** Nivel medio en todas las variables. Segmento base.
- **Cluster 3 - Ocasionales Diversos:** Compras esporádicas, pero con exploración de muchos productos. Oportunidad de conversión.
- **Cluster 4 - Cazadores de Ofertas:** Ticket bajo, recurrencia limitada, sensibles al precio. Reaccionan bien a descuentos y promociones.
- **Cluster 5 - Dormidos/Nuevos:** Poco activos, potencial desaprovechado. Clave en estrategias de bienvenida y reactivación.

6 Modelos Predictivos

1. Regresión Lineal Múltiple (OLS por Clúster)

- Permite entender relaciones internas entre precio, cantidad, stock y venta por segmento.
- Resultados expresan el peso de cada variable sobre los ingresos.
- Fácil de interpretar para equipos no técnicos (entender qué variable mueve más el ingreso dentro de cada grupo).
- Se aplicó por cada clúster de productos, lo que permite capturar relaciones específicas por segmento.
- Explica cómo Precio Unitario, Cantidad y Stock influyen en los ingresos.

2. Modelos de Boosting

Técnicas utilizadas: Gradient Boosting, AdaBoost, XGBoost.

- Capturan relaciones no lineales y combinaciones complejas.
- Mejoran significativamente el RMSE y el R^2 comparado con OLS.
- Ideales para predecir ingresos futuros por segmento incluso con variabilidad alta.
- Útiles para un forecast operativo, planeación de compras y análisis “what-if”.

La implementación de modelos predictivos aporta una mayor precisión en la planeación de la demanda, permitiendo ajustar inventarios con base en estimaciones más confiables y reducir compras innecesarias mientras se mejora la disponibilidad de productos. Además, actúan como un soporte para decisiones de pricing y estrategias promocionales, al identificar cómo distintos factores impactan directamente en los ingresos. Por último, establecen una base técnica robusta para desarrollar modelos futuros más avanzados como

series de tiempo o técnicas de forecasting multivariante que podrán seguir incrementando la capacidad predictiva del negocio.

7 Resultados

- Clasificación precisa de productos para decisiones de stock.
- Segmentación de clientes que permite marketing personalizado.
- Modelos predictivos robustos con alta capacidad de generalización.

8 Limitaciones y Oportunidades

Limitaciones

- Datos limitados a comportamiento histórico.
- Ausencia de factores externos como promociones, clima, competencia.
- No se incluyeron variables temporales complejas (tendencias globales).

Áreas de Oportunidad

- **Encuestas a clientes:** detectar razones de compra, satisfacción, promociones preferidas.
- **Integrar variables externas:** clima, días festivos, eventos, redes sociales.
- **Canales de feedback:** reviews, atención al cliente, menciones en redes.
- **Datos de recompra:** tiempo entre compras, productos relacionados.

9 Recomendaciones

- Enfocar inventario en productos del Cluster 3.

- Personalizar promociones por tipo de cliente.
- Aplicar estrategias de retención y reactivación según cluster.
- Hacer encuestas post-compra para enriquecer el análisis.
- Reentrenar modelos con datos actualizados cada trimestre.

paldada por datos, reduciendo la incertidumbre en áreas clave como compras, logística, pricing y promociones. Además, abre el camino para la evolución hacia sistemas avanzados de predicción, incluyendo modelos temporales multivariantes y la integración de fuentes externas como estacionalidad, clima, eventos o tendencias de mercado que permitirán una planificación aún más precisa y adaptable a cambios del entorno.

Conclusión

El análisis integral de patrones de ventas, apoyado en técnicas de EDA, clustering y modelos predictivos, permitió identificar comportamientos clave de productos y clientes, optimizar la gestión de inventario y mejorar la toma de decisiones comerciales. Los hallazgos obtenidos sirven como base para implementar estrategias de surtido, promociones y segmentación más efectivas. Se recomienda continuar con el monitoreo periódico del comportamiento del mercado, así como enriquecer el análisis con fuentes externas y cualitativas, como encuestas o datos de contexto, que permitan capturar aspectos no reflejados en las ventas históricas.

El uso integrado de análisis exploratorio, segmentación mediante clustering y modelos predictivos proporciona una visión completa del comportamiento comercial, del portafolio de productos y de los distintos tipos de clientes. Esta combinación metodológica permite comprender no solo qué está ocurriendo en términos de ventas, inventario y comportamiento de compra, sino también por qué sucede y cómo evolucionará en el corto y mediano plazo.

Con esta perspectiva completa, se puede optimizar la gestión de inventarios al identificar productos con sobre-stock, alta rotación o demanda estable, diseñar estrategias comerciales personalizadas basadas en perfiles reales de clientes y anticipar patrones de demanda utilizando modelos con mayor precisión que enfoques tradicionales.

El proyecto establece una base sólida para una toma de decisiones más rápida, informada y res-