



# Tecnológico de Monterrey

**Proyecto: Análisis de Patrones de Ventas**

**Documento técnico**

Helena Eridani Escandón López - A01659511

Franck Rodolfo Méndez Saint Louis - A01659339

Armando Atanasio Navarrete Yépez - A01658529

Fátima Quiroz Romero - A01660757

Aplicación de métodos multivariados en ciencia de datos

Grupo 602

Juliho David Castillo Colmenares

Jonathan Montalvo Urquizo

TEC CCM

Miércoles, 3 de diciembre de 2025

## Índice

<b>I. OBJETIVO.....</b>	<b>2</b>
<b>II. DATOS.....</b>	<b>2</b>
<b>III. LIMPIEZA/FEATURE ENGINEERING.....</b>	<b>2</b>
<b>IV. EDA.....</b>	<b>3</b>
<b>V. CLUSTERING.....</b>	<b>8</b>
<b>VI. MODELOS.....</b>	<b>10</b>
<b>VII. RESULTADOS.....</b>	<b>11</b>
<b>VIII. LIMITACIONES Y RECOMENDACIONES.....</b>	<b>15</b>

## I. OBJETIVO

El presente proyecto tiene como finalidad analizar los patrones históricos y las ciclicidades de ventas en un dataset simulado que refleja las dinámicas del mercado argentino durante 2024. A pesar de su naturaleza sintética, el objetivo primordial es que los hallazgos generados brinden propuestas estratégicas accionables y replicables que apoyen la toma de decisiones críticas en los frentes de abasto, eficiencia operativa y planificación comercial. A través de técnicas de análisis multivariado y exploración estructurada de datos, se busca comprender la dinámica de la demanda y sus variaciones por región, categoría y producto. Con ello, el proyecto busca generar insumos analíticos accionables que permitan comprender información vital para la planificación estratégica: qué, dónde y cómo se vende, al igual de identificar cuál es el impacto de estas dinámicas en la operación comercial.

En términos de negocio y análisis, este estudio persigue los siguientes objetivos:

- Identificar tendencias, estacionalidades y picos de demanda, permitiendo anticipar las necesidades de inventario y mejorar la disponibilidad de productos.
- Comparar el desempeño de categorías y productos entre regiones, clasificándolos en escalas de importancia e identificando los 5 productos más y menos vendidos por zona.
- Analizar el comportamiento de compra mediante métricas clave como ventas totales, ticket promedio, métodos de pago y rotación regional de productos y categorías.
- Caracterizar el comportamiento de los clientes, diferenciando perfiles frecuentes y recurrentes por región, para entender qué compran, con qué periodicidad y a través de qué métodos de pago.
- Segmentar productos mediante técnicas de clustering para identificar patrones de comportamiento comunes en términos de ingreso generado, dependencia regional, actividad temporal y dispersión geográfica.
- Evaluar la sensibilidad de la demanda a cambios de precio, identificando categorías elásticas e inelásticas, y estimar el efecto de escenarios de pricing sobre los ingresos regionales.
- Simular escenarios de ajuste de precios para proyectar cambios esperados en los ingresos por región.

## II. DATOS

El análisis se desarrolló utilizando un conjunto de datos proveniente de un repositorio público de Kaggle, el cual simula el comportamiento del mercado argentino durante 2024. Este dataset contiene 3,029 observaciones, 25 variables y está conformado por cinco fuentes principales: **ventas**, **productos**, **clientes**, **categorías** y **métodos de pago**. Todas las tablas fueron integradas mediante claves comunes —principalmente *ID\_Producto*, *ID\_Cliente* y *Método\_Pago*— para construir una vista unificada del comportamiento comercial. A continuación, se desglosa la información del dataset:

### ❖ Fuentes de datos

- **Ventas:** incluye ID de venta, fecha, cliente, producto, cantidad, método de pago y estado de la transacción.
- **Productos:** ID de producto, nombre, categoría y precio unitario.
- **Clientes:** ID de cliente y región asociada.
- **Categorías:** clasifica cada producto en familias comerciales.
- **Métodos de pago:** tipo de pago registrado en cada transacción.

### ➤ Cobertura y limitaciones

- La región NOA presenta ausencia de datos entre febrero y mayo, lo que afecta parcialmente los análisis de estacionalidad y elasticidad.
- La base no incluye costos, inventarios ni información logística, por lo que el estudio se centra exclusivamente en el comportamiento de la demanda y el ingreso generado.
- Al ser un dataset simulado, su propósito es reflejar patrones realistas, mas no representar cifras reales del mercado argentino.

## III. LIMPIEZA/FEATURE ENGINEERING

### ❖ Construcción del dataset analítico

Para habilitar el análisis económico, se construyó la métrica central del estudio:

$$\text{Ingreso} = \text{Precio unitario} \times \text{Cantidad vendida}$$

Posteriormente se integraron todas las tablas mediante llaves de producto y cliente, generando un dataset enriquecido que permite análisis por región, categoría, producto, cliente y dimensiones temporales.

### ➤ **Enriquecimiento temporal**

Con el fin de evaluar estacionalidad diaria, semanal y mensual, así como picos y valles de demanda, a partir de la fecha de venta se derivaron:

- Año
- Mes (numérico) y nombre del mes
- Día del mes
- Día de la semana

### ➤ **Estructuras derivadas y variables adicionales**

Se prepararon las siguientes estructuras para habilitar los análisis comparativos, las segmentaciones y las evaluaciones económicas desarrolladas en las fases posteriores del estudio:

- **Dataset de productos:** ingreso total, cantidad total vendida y número de tickets.
- **Matrices región–producto:** permiten evaluar qué productos son críticos en cada región y asignarlos a escalas A–B–C.
- **Variables de comportamiento por producto:**
  - porcentaje del ingreso proveniente de Buenos Aires,
  - porcentaje de ventas en fin de semana,
  - número de regiones en las que el producto tiene actividad comercial,
  - distribución proporcional de ventas por región.
- **Dataset categoría–región:** utilizado para estimar elasticidades precio–demanda por categoría y simular escenarios de pricing a nivel regional.

## IV. EDA

Esta sección tuvo como propósito adquirir una mayor comprensión acerca de la estructura general del negocio sobre el que se está trabajando, identificar patrones preliminares de ventas y validar la coherencia interna del dataset antes de avanzar hacia técnicas de segmentación, elasticidad y simulación. Además, se hizo énfasis en describir el comportamiento histórico sin emitir conclusiones estratégicas, permitiendo establecer una primera línea base del funcionamiento comercial durante 2024. A continuación, se presentan los principales hallazgos descriptivos obtenidos:

- ❖ **KPIs globales del negocio:** Se calcularon indicadores generales para caracterizar la actividad

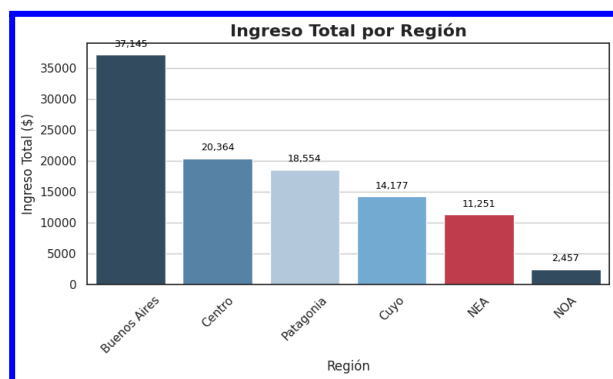
comercial agregada, con el fin de dimensionar el tamaño del dataset y el volumen transaccional del sistema comercial simulado.

- **Ingreso total anual:** \$103,947.36
- **Número total de ventas (tickets):** 3,000
- **Cientes únicos:** 326
- **Productos únicos:** 38
- **Ticket promedio:** \$34.65
- ❖ **Comparación regional:** Para evaluar disparidades geográficas en volumen y valor, se analizaron ingresos, cantidad de ventas y ticket promedio por región.

### Ingresos por región

El ranking de ingresos anuales muestra una fuerte concentración en:

- 1) **Buenos Aires:** ~37,144
- 2) **Centro:** ~20,364
- 3) **Patagonia:** ~18,554
- 4) **Cuyo:** ~14,177
- 5) **NEA:** ~11,251
- 6) **NOA:** ~2,457



*Figura 1. Ingreso anual por región. Buenos Aires registra el mayor nivel de ingresos, mientras que NOA presenta el menor volumen del sistema.*

### Ticket promedio por región

Se presenta una dinámica distinta al volumen ya que la región de **NOA** registra el ticket más alto (~39), mientras que **NEA** el más bajo (~32.5) y las demás oscilan entre 33 y 36. Este contraste muestra que **valor y volumen no están correlacionados directamente a nivel regional**, aunque esta interpretación más profunda se desarrolla posteriormente en Resultados.

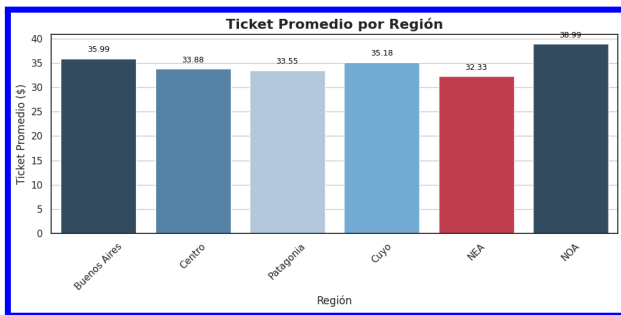


Figura 2. Ticket promedio por región. El valor del ticket promedio no sigue la misma jerarquía que el volumen total de ventas.

### Relación ingreso total – ticket promedio

El análisis conjunto confirma tres patrones descriptivos que ayudan a entender la heterogeneidad regional antes de aplicar modelos más avanzados:

- Buenos Aires: alto volumen, valor de ticket medio-alto
- NOA: valor de ticket alto, volumen muy bajo
- NEA: valor de ticket bajo y volumen bajo

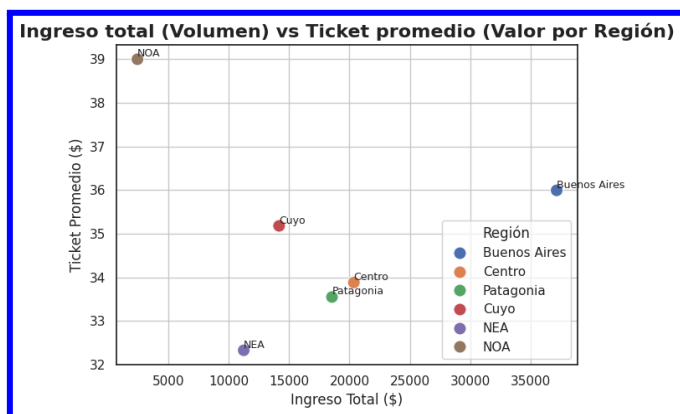


Figura 3. Relación entre ingreso total y ticket promedio por región. Las regiones presentan combinaciones distintas entre volumen y valor por ticket, sin una tendencia lineal aparente.

### ❖ Estacionalidad temporal

Con el enriquecimiento temporal realizado previamente, se pudo hacer una evaluación acerca de los patrones diarios, semanales y mensuales de ingresos:

#### ➤ Patrón mensual

El ingreso presenta variaciones mes a mes, observándose algunos con mayor actividad comercial y otros con bajas relativas, lo que permite identificar picos moderados en periodos específicos del año (febrero-marzo) y valles que permiten observar estacionalidad (mitad de año); Buenos Aires es el principal motor de esta variabilidad.

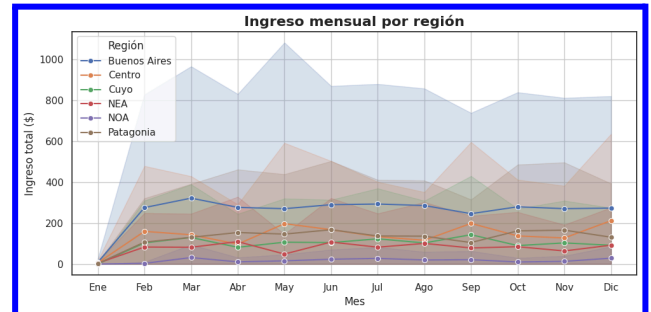


Figura 4. Ingreso mensual por región. La línea de Buenos Aires domina la variabilidad total del sistema.

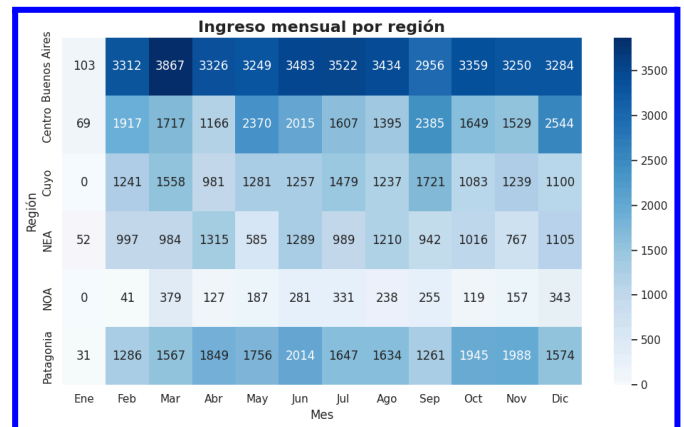


Figura 5. Heatmap de ingreso mensual por región. Muestra los contrastes estacionales entre regiones.

#### ➤ Patrón semanal

El análisis por día de la semana muestra diferencias en el comportamiento del consumidor entre días hábiles y fines de semana. Estos patrones permiten detectar ritmos de compra que se retomaron posteriormente en modelos de segmentación de productos.

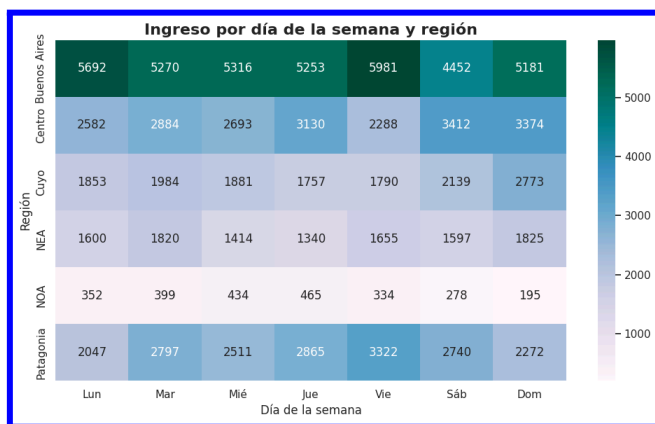


Figura 6. Ingreso por día de la semana y región. Se distinguen diferencias entre días hábiles y fines de semana, con variaciones por región.

### ➤ Patrón diario

El ingreso muestra fluctuaciones esperables a lo largo del mes, con picos y valles regulares que reflejan variaciones normales en la demanda diaria, sin presentar comportamientos atípicos o inconsistentes.



Figura 7. Variación del ingreso por día del mes. Se observan picos naturales (días 2 y 30) y valles moderados (días 8 y 31), sin anomalías en el comportamiento

### ❖ Desempeño por categoría

Después de evaluar las categorías para identificar cuáles representan mayor participación en el ingreso total para contextualizar la participación de cada familia de productos, se observó que algunas categorías dominan el portafolio con mayor volumen de ventas, mientras que otras presentan menor recurrencia, pero con tickets más altos. Adicionalmente, se pudo identificar la contribución de cada categoría en cada región y su relevancia proporcional dentro del portafolio.

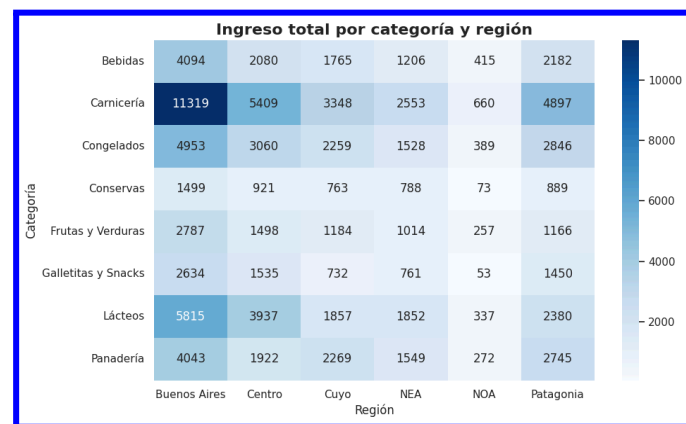


Figura 8. Ingreso total por categoría y región. Carnicería tiene los mayores ingresos en todas las regiones.

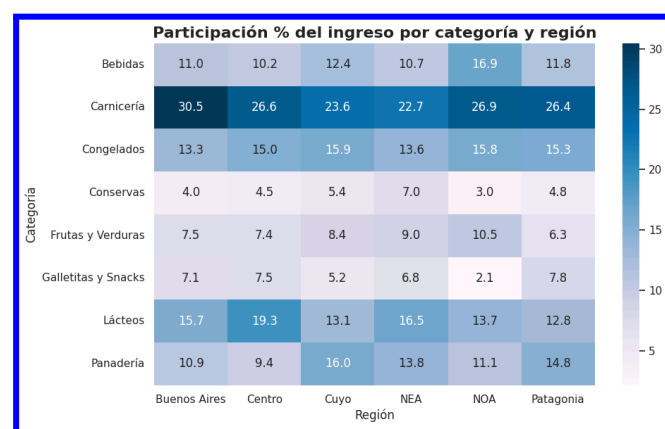


Figura 9. Participación porcentual del ingreso por categoría y región.

De forma exploratoria, observando los gráficos obtenidos, se realizó el hallazgo de algunas categorías concentran mayor volumen en todas las regiones, mientras que otras muestran presencia constante pero de menor contribución relativa. Con esto se concluye que la composición varía moderadamente entre zonas.

❖ **Análisis de productos por región:** Se evaluó la importancia de cada producto dentro de su región y su contribución al ingreso total. Esto permite observar que existen composiciones distintas por región, tanto en productos de alto como de bajo ingreso.

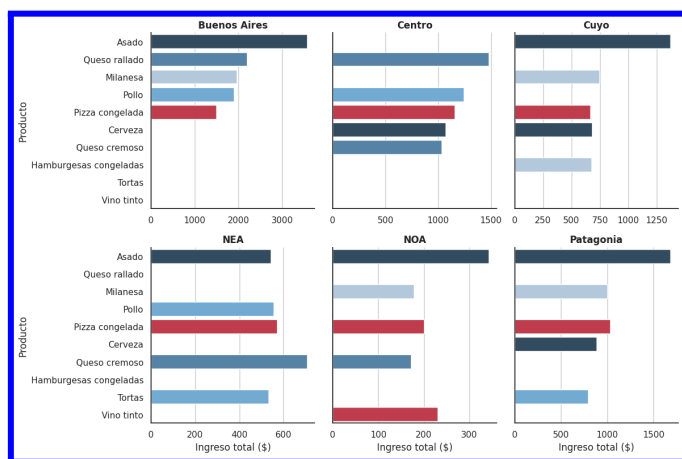


Figura 10. Top 5 productos por región

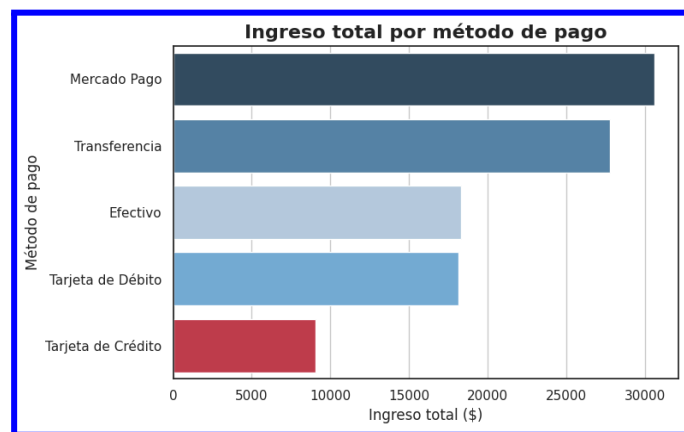


Figura 12. Ingreso total por método de pago. Mercado Pago y Transferencia dominan tanto en tickets como en monto.

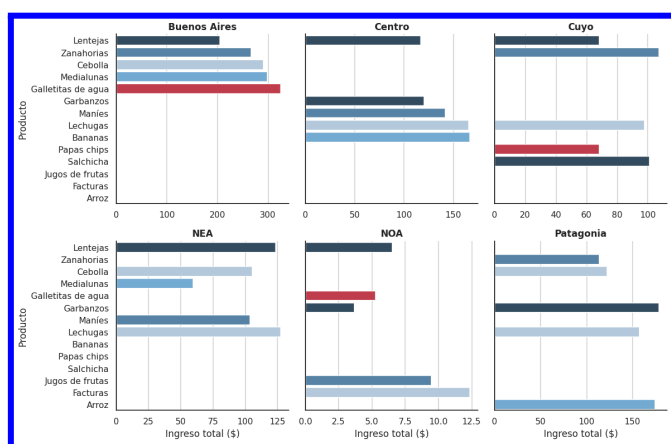


Figura 11. Bottom 5 productos por región

## ❖ Métodos de pago

Se revisó la distribución y frecuencia de los métodos de pago utilizados en las transacciones, observando que existen métodos predominantes que concentran la mayoría de las ventas y que la distribución varía por región y por categoría, observándose preferencias locales en el comportamiento de pago. Esto brinda indicios de que existen relaciones entre métodos de pago con categorías y productos, las cuales se desarrollarán más adelante.

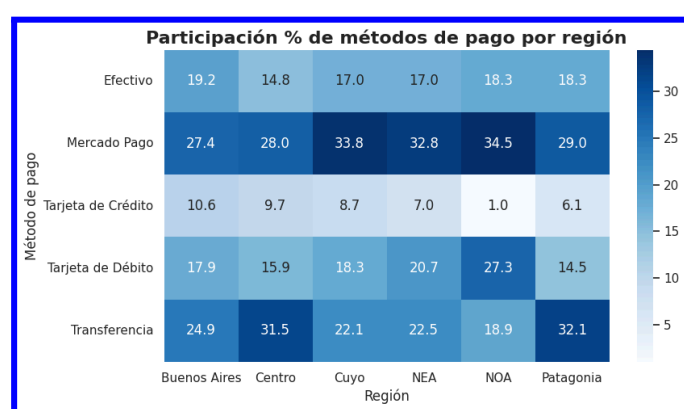


Figura 13. Participación % de métodos de pago por región.

Con esto, se observa que las preferencias de pago cambian según la región: algunas privilegian billeteras digitales, mientras que otras mantienen un uso relativamente mayor de efectivo o débito.

## ➤ Por categoría y región

Posteriormente, se examinó cómo varían los métodos de pago según la combinación entre región y categoría de producto. El objetivo fue identificar patrones operativos —no estratégicos— que muestren si ciertas familias de productos tienden a concentrarse en determinados medios de pago dentro de cada zona geográfica. Los siguientes gráficos revelan interacciones específicas entre métodos de pago y categorías claves.

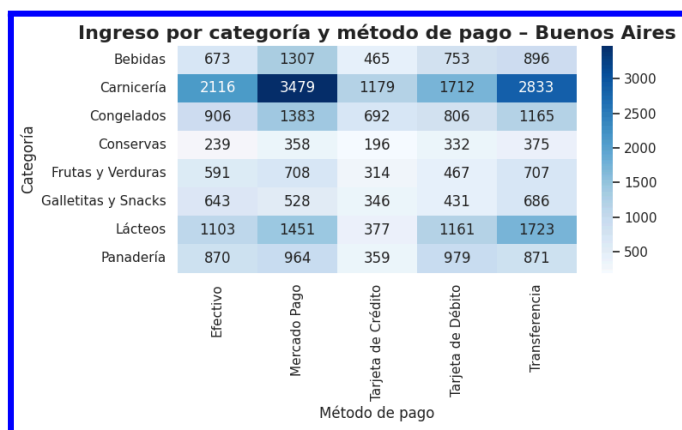


Figura 14. Distribución de métodos de pago por categoría en Buenos Aires.

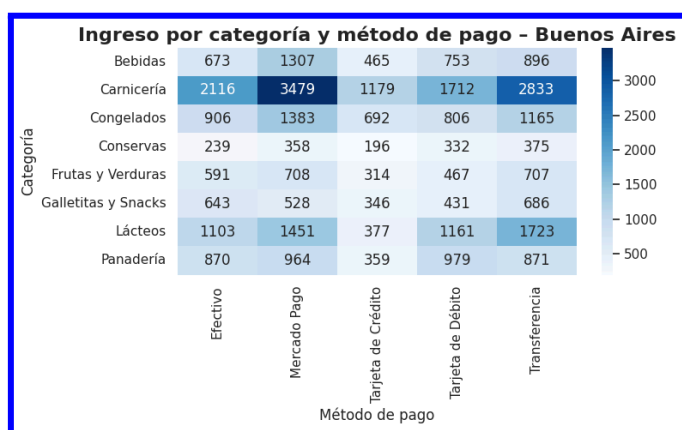


Figura 15. Distribución de métodos de pago por categoría en NOA.

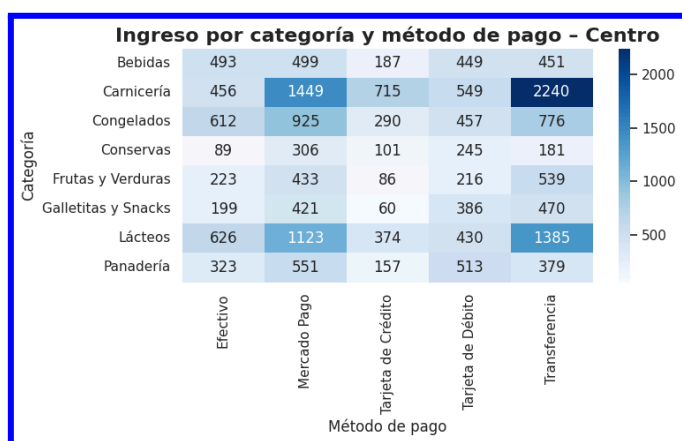


Figura 16. Distribución de métodos de pago por categoría en Centro.

Habiendo observado la distribución de los métodos de pago por categoría, ahora se analizó específicamente para los productos de mayor contribución en cada región. Este enfoque permite observar si los artículos más relevantes del portafolio presentan preferencias de cobro particulares, información útil

para comprender hábitos transaccionales antes de aplicar modelos más complejos.

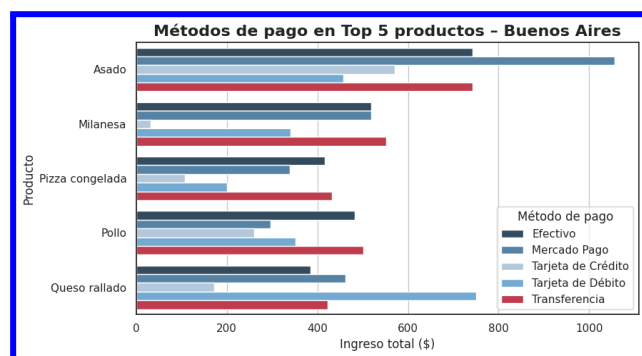


Figura 17. Métodos de pago en Top 5 productos – Buenos Aires.

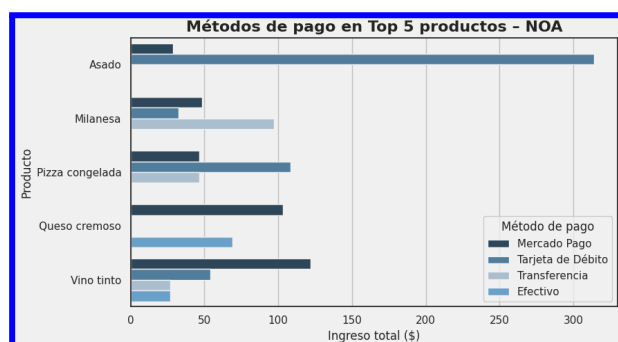


Figura 18. Métodos de pago en Top 5 productos – NOA.

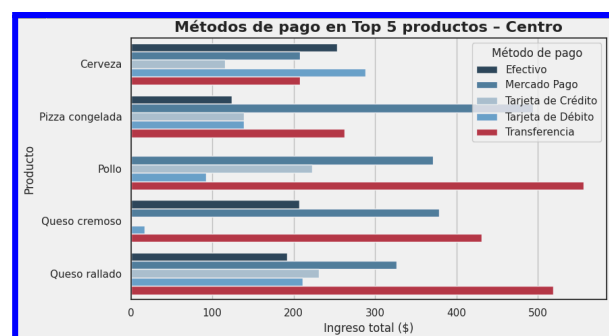


Figura 19. Métodos de pago en Top 5 productos – Centro.

### ❖ Perfil preliminar de clientes

Finalmente, se describió el comportamiento agregado de los clientes en cada región, evaluando tanto su volumen como su frecuencia de compras durante el periodo analizado. Este análisis permitió establecer una primera caracterización descriptiva de la base de clientes, sin segmentar aún formalmente, pero preparando el camino para análisis posteriores de clusters y patrones de recurrencia.



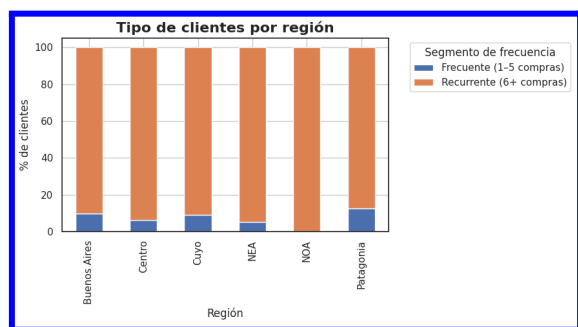


Figura 20. Distribución de segmentos de clientes por región.

## V. CLUSTERING

En esta sección se segmentó el portafolio de productos a partir de patrones de comportamiento comercial observados en el dataset, es decir, que en lugar de operar únicamente con categorías estáticas, este análisis buscó identificar grupos naturales de productos que compartieran dinámicas similares. Dichas similitudes podrían encontrarse en términos de ingreso generado, dependencia geográfica y hábitos temporales de compra; esto permitió comprender qué tipos de productos tienen un rol estratégico distinto dentro del negocio, habilitando líneas de acción orientadas a surtido, pricing y optimización regional.

Para ello se construyó un dataset analítico específico de productos que resume sus características fundamentales:

- **Tamaño económico:** ingreso total anual y ticket promedio por producto.
- **Dependencia geográfica:** porcentaje del ingreso proveniente de Buenos Aires y número de regiones donde el producto realmente tiene ventas.
- **Temporalidad:** porcentaje de ventas generadas durante fines de semana.

Estas variables capturan la dispersión, estabilidad y criticidad comercial de cada producto dentro del sistema; a partir de ellas, se procedió al escalado y posterior aplicación de K-Means.

En primer lugar se construyó un dataset enfocado solamente en productos, con el que se buscaba caracterizar el rol de cada uno dentro del negocio y se utilizó como insumo principal del modelo de clustering. En dicho set se integraron:

- 1) **Desempeño económico por producto**
  - a) Ingreso total
  - b) Frecuencia en tickets
  - c) Cantidad total vendida
  - d) Ticket promedio por producto
- 2) **Distribución geográfica del ingreso**

- a) Porcentaje del ingreso proveniente de Buenos Aires
  - b) Número de regiones en las que el producto registra ventas
- 3) **Comportamiento temporal:** Porcentaje de ingreso generado en fines de semana

### ❖ Selección de variables y escalado

Las variables seleccionadas para el modelo fueron:

- **Ingreso\_Total**
- **Ticket\_Promedio\_Prod**
- **Pct\_BA**
- **N\_Regiones\_Activas**
- **Pct\_Finde**

Todas fueron estandarizadas mediante **StandardScaler**, con el objetivo de homogeneizar escalas y asegurar que ningún atributo dominara el proceso de segmentación.

### ❖ Determinación del número óptimo de clusters, asignación y caracterización

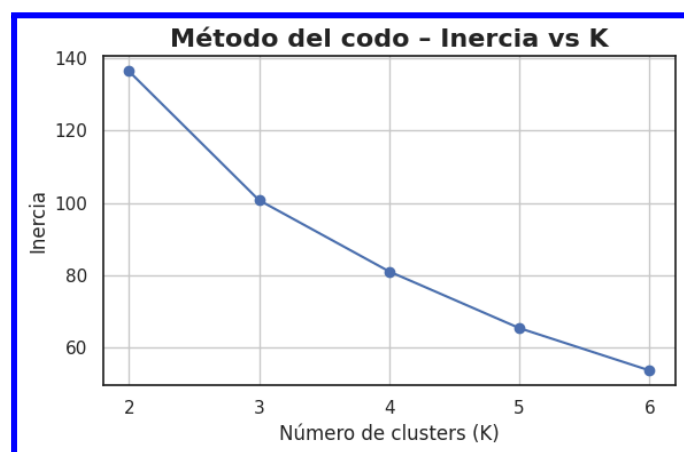


Figura 21. Método del codo: la inercia muestra una disminución abrupta hasta  $K=3$ , donde comienza a estabilizarse.

Se aplicaron dos métodos de validación interna:

- **Método del codo (Inercia vs. K)**
  - La reducción abrupta de inercia se estabiliza en  $K \approx 3$  o 4.
  - A partir de  $K=3$  las mejoras marginales disminuyen.

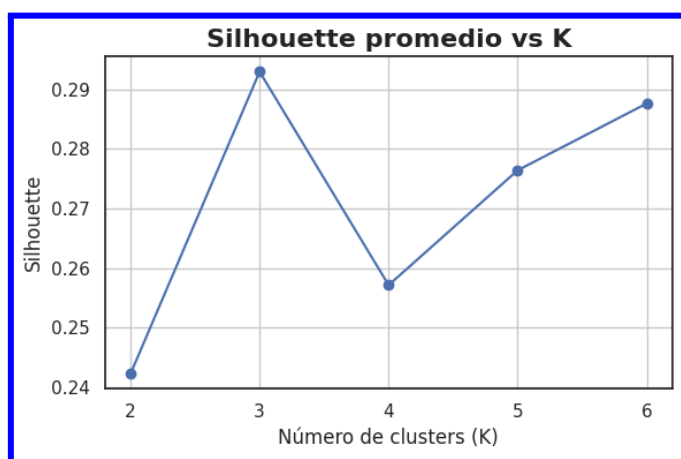


Figura 22. Coeficiente de Silhouette promedio: el valor máximo se obtiene en K=3, indicando la separación más estable entre clusters.

- **Coeficiente de Silhouette**
  - El valor máximo observado fue **Silhouette  $\approx 0.293$  en K=3**.
- K=3 presentó mejor separación relativa entre grupos que valores superiores.

Se utilizó **K=3**, ya que ambos métodos lo mostraban como la estructura más apropiada para el conjunto de productos. Posteriormente, se entrenó el modelo K-Means y se asignó cada producto a un grupo. Posteriormente se calcularon métricas promedio por clúster para interpretar los patrones dominantes, lo que permitió identificar diferencias claras en tamaño económico y distribución territorial e interpretar segmentos con roles distintos en el negocio.

Cluster	Productos	Ingreso Promedio	Ticket Promedio	% BA Promedio	Regiones Activas	% Finde
0	15	4,336	54.4	0.36	6	0.29
1	19	1,694	21.7	0.33	6	0.29
2	4	1,680	21.3	0.37	5	0.31

#### ❖ Composición por categoría dentro de cada clúster

La distribución de categorías dentro de cada grupo permitió entender si los clusters capturan comportamientos comunes más allá de la clasificación comercial tradicional, por ejemplo:

- **Cluster 0:** compuestos principalmente por **Carnicería (33%)**, **Congelados (20%)** y **Lácteos (20%)**, de alto ingreso y mayor contribución por ticket.
- **Cluster 1:** más diversificado, con presencia fuerte de **Frutas y Verduras (26%)**, **Conservas (21%)** y

**Snacks (16%)**, productos de rotación media y dispersión amplia.

- **Cluster 2:** pequeño, con productos más específicos y menor dispersión geográfica.

Esto evidencia que los clusters no replican categorías formales, sino patrones reales de comportamiento económico y territorial.

#### ❖ Visualización del clustering mediante PCA

Para facilitar la interpretación visual, se aplicó **PCA** para reducir la dimensionalidad a dos componentes principales.

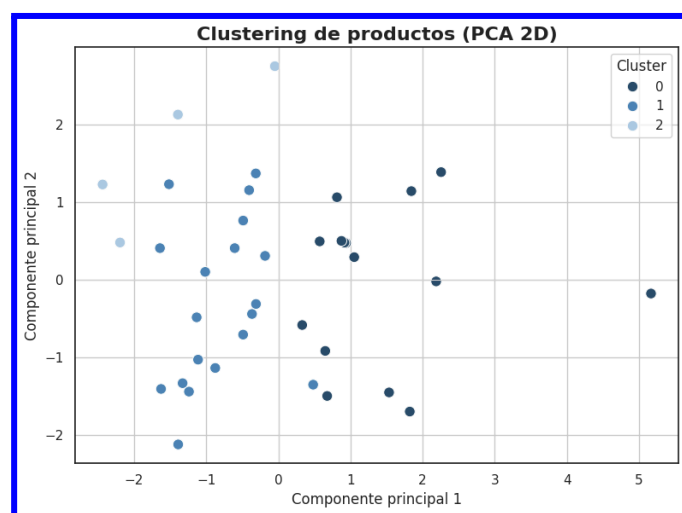


Figura 23. Visualización del clustering de productos mediante PCA (2 componentes). A pesar del ruido visual, se aprecian patrones de dispersión y concentración que diferencian el rol comercial de cada grupo.

Aunque los clusters presentan cierta superposición, la proyección 2D permite observar:

- Un grupo compacto y de mayor tamaño económico (Cluster 0) desplazado hacia valores positivos de la primera componente.
- Productos más dispersos en clusters 1 y 2, indicando variabilidad interna mayor en productos de ticket pequeño o comportamiento más estacional.

Para contextualizar el significado de cada grupo, se seleccionaron los tres productos con mayor ingreso dentro de cada clúster:

- **Cluster 0 (alto valor):** Asado, Milanesa, Pizza congelada
- **Cluster 1 (rotación media):** Facturas, Papas fritas, Pan francés
- **Cluster 2 (rotación baja / nicho):** Salchicha, Galletitas de chocolate, Medialunas

Estos ejemplos muestran que los productos con mayor impacto económico se concentran en el clúster 0, mientras que los productos con menor presencia o dispersión geográfica se ubican en el clúster 2.

## VI. MODELOS

A continuación, se describen los modelos analíticos implementados para estimar la sensibilidad precio–demanda y evaluar escenarios alternativos de pricing a nivel regional. Estos se usaron para cuantificar cómo distintas variaciones de precios impactan el ingreso esperado, considerando las diferencias estructurales entre regiones y categorías.

### ❖ Estimación de elasticidad precio–demanda

Para evaluar la relación entre precio e ingreso por categoría y región, se estimaron elasticidades empleando modelos de regresión logarítmica del tipo:

$$\log(\text{Ingreso}) = \beta_0 + \beta_1 * \log(\text{Precio}) + \varepsilon$$

donde el coeficiente ( $\beta_1$ ) representa la elasticidad precio–demanda. Este enfoque permite interpretarla directamente como el cambio porcentual esperado en la demanda frente a un cambio porcentual en el precio.

Los modelos se ajustaron a nivel **categoría–región** utilizando el dataset previamente construido, el cual resume las ventas históricas con precios simulados consistentes con el comportamiento del mercado argentino. Las categorías se clasificaron posteriormente como:

- **Inelásticas** ( $|\beta_1| < 1$ ), en estas se puede subir el precio y aumentará el ingreso, sin riesgo inmediato ni generalizado de perder demanda.
- **Elásticas** ( $|\beta_1| \geq 1$ ), son sensibles a los cambios de precio: si se sube, existe el riesgo inmediato de perder demanda.

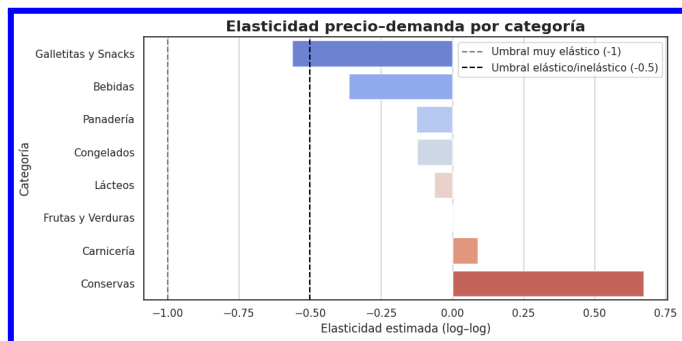


Figura 24. Elasticidad precio–demanda por categoría. Comparación de elasticidades estimadas mediante regresión log–log; se incluyen líneas de referencia para identificar umbrales de categorías muy elásticas y elásticas/inelásticas.

### ❖ Simulación de escenarios de pricing

A partir de las elasticidades estimadas, se modelaron tres escenarios para evaluar su impacto sobre el ingreso total por región. Cada escenario modifica los precios por categoría y proyecta el ingreso simulado utilizando:

$$\text{Ingreso simulado} = \text{Precio Nuevo} * \text{Cantidad Estimada}$$

donde la cantidad estimada se ajusta según la elasticidad correspondiente.

Los escenarios evaluados fueron:

- **Escenario 1 — Aumento general de precio (+10%)**  
Se aplicó un incremento del 10 % a todas las categorías del portafolio, sin distinguir elasticidad. Este escenario sirve como línea base de comparación para las dos siguientes simulaciones.
- **Escenario 2 — Ajuste focalizado (+10% solo a categorías inelásticas)**  
El incremento del 10 % se aplicó únicamente a las categorías clasificadas como inelásticas en cada región. Este modelo permite evaluar estrategias de pricing selectivo orientadas a maximizar ingreso sin deteriorar la demanda.
- **Escenario 3 — Descuento estratégico (–10% en categorías muy elásticas)**  
Se aplicó una reducción del 10 % únicamente a categorías consideradas altamente elásticas. Aquí se buscó evaluar si un descuento podría aumentar el ingreso total mediante un mayor volumen de ventas.

Los resultados simulados fueron consolidados en una estructura analítica denominada **impacto\_regional**, donde se incluyen: región, escenario, ingreso actual, ingreso simulado, variación absoluta y variación porcentual del ingreso. Esta tabla permite comparar el efecto económico de cada escenario en todas las regiones y fue exportada para su análisis en la etapa final de la ejecución del código utilizado para el análisis del dataset.

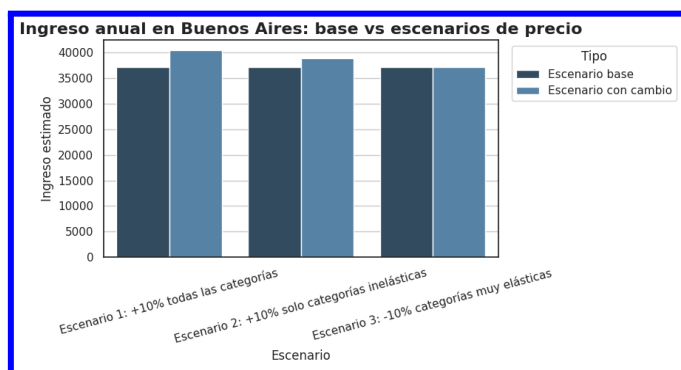


Figura 25. Comparación del ingreso simulado en Buenos Aires bajo tres escenarios de variación de precios; se observa que los aumentos de precio generan mejoras sostenidas en el ingreso anual.

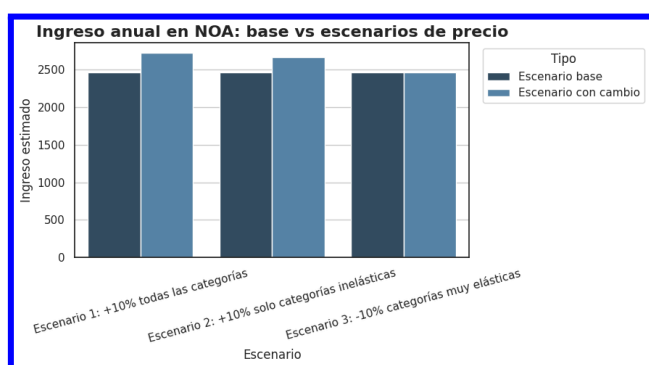


Figura 26. Ingreso anual en NOA: base vs escenarios de precio. La región muestra incrementos proporcionales mayores ante aumentos de precios; los descuentos no traen ningún beneficio.

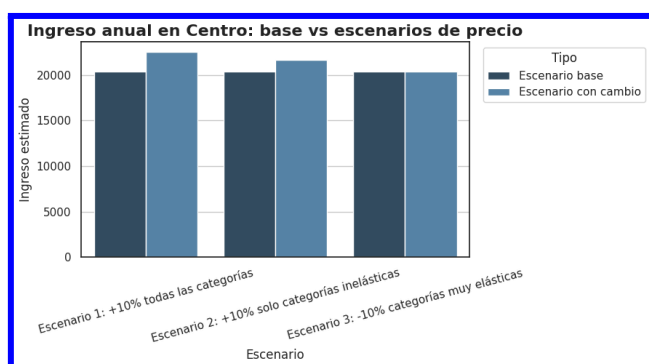


Figura 27. Ingreso anual en Centro: base vs escenarios de precio. Los escenarios de aumento de precio elevan el ingreso en forma moderada, mientras que los descuentos no producen mejoras relevantes.

Para complementar el análisis individual por región, se incluye a continuación una visualización comparativa que sintetiza el efecto porcentual de cada escenario de precios en todas las regiones del país. Esta gráfica resume de forma integrada el impacto económico esperado bajo cada estrategia.

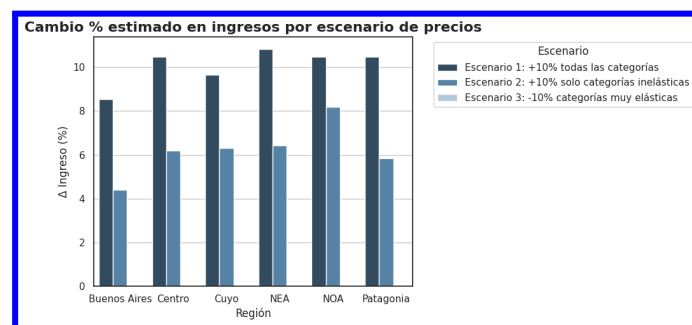


Figura 28. Cambio porcentual estimado del ingreso por región bajo distintos escenarios de pricing. El escenario 1 muestra el mayor crecimiento promedio; el escenario 2 el mejor equilibrio riesgo-beneficio; el escenario 3 no genera mejoras.

## VII. RESULTADOS

Después de aplicar los modelos indicados -clustering de productos, estimación de elasticidades y la simulación de escenarios-, se obtuvieron patrones, los cuales se interpretarán a continuación, al igual que sus implicaciones de negocio para surtido, planeación comercial, segmentación y estrategias de precios.

### ❖ Resultados del Clustering de Productos

El análisis de clustering reveló tres segmentos claramente diferenciados de productos que capturan patrones reales de comportamiento económico, territorial y temporal, más allá de las categorías tradicionales, lo que permite entender la estructura del portafolio con una mirada operativa y comercial; los tres clusters identificados son:

#### ➤ Cluster 0 – Productos de alto valor e impacto económico, este grupo reúne productos con:

- Ingreso promedio más alto del portafolio ( $\approx 4,336$ )
- Ticket promedio elevado
- Amplia presencia territorial ( $\approx 6$  regiones activas)
- Contribución geográfica relevante desde Buenos Aires

Estos productos muestran un rol estratégico dentro del negocio: son motores del ingreso y mantienen estabilidad tanto en la demanda como en la dispersión geográfica. Los principales ejemplos incluyen Asado, Milanesa y Pizza congelada, que coinciden con categorías de alto consumo y fuerte presencia regional. Cluster 1 – Productos de rotación media y diversificación amplia

➤ **Cluster 1:**

- Nivel de ingreso medio
- Ticket promedio moderado
- Amplia dispersión geográfica
- Patrón de compra notablemente estable entre días hábiles y fines de semana

Cabe destacar que, a diferencia del cluster 0, el valor de estos productos no proviene del ticket, sino de volúmenes frecuentes y dispersos. Entre los productos más representativos se encuentran Facturas, Papas fritas y Pan francés; sus categorías dominantes son Frutas y Verduras, Conservas y Snacks, las cuales confirman los hallazgos, ya que cuentan con amplia presencia en surtidos de supermercados.

➤ **Cluster 2 – Productos de nicho o baja participación** (más pequeño):

- Ingreso reducido
- Presencia restringida a 5 o menos regiones
- Frecuencia de compra menor o comportamiento más esporádico

Algunos productos de este cluster son Salchicha, Galletitas de chocolate y Medialunas, productos con demanda menos estable o altamente localizada. En caso de que este set de datos mostrara el comportamiento real del mercado, este segmento evidenciaría oportunidades para racionalizar inventarios y optimizar espacio de anaquel en regiones donde su rotación es mínima.

### Patrón general del portafolio

Los clusters no replican la clasificación comercial tradicional. Esto confirma que la agrupación basada en comportamiento económico y territorial es más informativa para decisiones estratégicas que las categorías estándar. La separación observada en la proyección PCA indica que hay un grupo dominante (cluster 0) con variabilidad interna baja, además de la existencia de demás grupos - Clusters 1 y 2- con mayor dispersión, reflejando heterogeneidad y sensibilidad a factores temporales y territoriales.

### ❖ Resultados del Análisis de Roles A/B/C por Categoría y por Producto

El análisis A/B/C se utilizó como una técnica de clasificación basada en la **contribución porcentual al ingreso dentro de cada región**, con el fin de identificar niveles de criticidad comercial tanto para **categorías** como para **productos individuales**. Para ello, se calcularon aportes acumulados y se asignaron tres niveles:

- **A (alto impacto):** primeros 80% del ingreso acumulado
- **B (impacto medio):** de 80% a 95%
- **C (bajo impacto):** más del 95%

Este enfoque permite distinguir las categorías y productos críticos del portafolio, las oportunidades intermedias y los elementos de baja aportación. Los resultados obtenidos complementan los hallazgos del clustering y de los modelos de elasticidad, proporcionando una perspectiva adicional para decisiones de surtido y priorización regional.

### Categorías

	Buenos Aires	Centro	Cuyo	NEA	NOA	Patagonia
Bebidas	A	A	B	A	A	B
Carnicería	A	A	A	A	A	A
Congelados	A	A	A	A	A	A
Conservas	C	C	B	B	C	C
Frutas y Verduras	B	C	B	B	B	C
Galletitas y Snacks	C	B	C	C	C	B
Lácteos	A	A	A	A	A	A
Panadería	B	B	A	A	B	A

Figura 29. Roles A/B/C por categoría y región. Las categorías se clasifican según su contribución acumulada al ingreso regional: A = alto impacto, B = impacto medio, C = bajo impacto.

El gráfico muestra que:

- **Carnicería, Lácteos, Panadería y Bebidas** se clasifican sistemáticamente como A en casi todas las regiones.
- **Congelados y Frutas y Verduras** alternan entre A y B, lo que evidencia sensibilidad geográfica.
- **Conservas y Galletitas/Snacks** permanecen en C, indicando baja aportación relativa.

## Productos

**Productos que cambian de rol entre regiones  
(A = alto impacto, B = medio, C = bajo)**

Producto	Buenos Aires	Centro	Cuyo	NEA	NOA	Patagonia
Manteca	C	B	B	B	B	A
Salchicha	A	A	C	A		A
Facturas	A	B	A	A	C	A
Bananas	B	C	A	B	B	B
Agua mineral	A	B	A	A	A	B
Jugos de frutas	B	B	A	A	C	A
Galletitas de agua	C	A	C	B	C	A
Papas chips	A	A	C	B	C	B
Arroz	B	B	B	B	C	C
Atún enlatado	B	A	A	A	B	B

Figura 30. Productos que cambian de rol A/B/C entre regiones. Se observan variaciones significativas en la importancia relativa de varios productos, indicando sensibilidad geográfica y comportamientos de demanda localizados.

El gráfico indica que:

- Algunos productos —como **manteca, salchicha, galletitas de agua y papas chips**— alteran su rol entre A, B y C dependiendo de la región.
- Los cortes de carne y algunos congelados permanecen como **A**, confirmando su rol estratégico estable.
- Productos frescos y de baja rotación se agrupan en **C**, sugiriendo oportunidades de optimización del surtido.

Este análisis confirma que el portafolio está compuesto por:

- Una **columna vertebral de productos y categorías A**, que sostienen el ingreso en todas las regiones.
- Un conjunto de **productos B**, cuyo impacto depende fuertemente del territorio.
- Un grupo **C** adecuado para revisión táctica, racionalización o estrategias de impulso selectivo.

### ❖ Resultados del Modelo de Elasticidad Precio–Demanda

La estimación econométrica permitió clasificar cada categoría según su sensibilidad al precio. Los valores obtenidos reflejan tres comportamientos principales:

- **Categorías con demanda inelástica ( $\beta_1$  entre 0 y 0.5):**

Subir su precio aumenta los ingresos

  - **Bebidas:** -0.36
  - **Congelados:** -0.12
  - **Frutas y Verduras:** 0
  - **Lácteos:** -0.06
  - **Panadería:** -0.12

Estos son productos esenciales o de reposición frecuente, por lo que su demanda se mantiene estable aun cuando el precio aumenta y un ajuste positivo en sus precios no compromete de forma significativa el volumen. Es decir, son las categorías con menor riesgo en estrategias de incremento de precios.

- **Categorías elásticas o muy sensibles al precio ( $\beta_1 \geq 1$ )**
  - Galletitas y Snacks: elasticidad negativa más alta del portafolio (categoría más sensible al cambio de precio)
  - Bebidas

En estos productos, los incrementos de precio generan una caída proporcionalmente mayor en la demanda.

Adicionalmente, como parte de los hallazgos, se considera útil mencionar también que se encontraron categorías con elasticidad positiva: Carnicería y Conservas. Estos resultados se interpretan como anomalías del modelo, atribuibles a la naturaleza simulada del dataset, ya que no siguen el comportamiento económico típico y suelen interpretarse como:

- Productos donde el precio fluctúa simultáneamente con calidad
- Efectos no capturados (promociones, sustitutos o complementariedades)
- Cambios estacionales que distorsionan la relación precio–cantidad

Se interpreta que la existencia de esta categoría se debe a que los datos no son extraídos directamente un comportamiento auténtico, si no de una simulación que imita el comportamiento del mercado real.

### ❖ Resultados de la Simulación de Escenarios de Pricing

La simulación evaluó tres estrategias de precio, estimando el impacto sobre los ingresos anuales de cada región, con el fin de determinar qué políticas contribuirían a maximizar el ingreso total y el nivel de riesgo que conllevan.

- **Escenario 1: +10% en todas las categorías**
  - **Resultado general:** Incremento consistente del ingreso en todas las regiones, con un promedio de **+10.08%**.
  - **Regiones con mayor impacto:**
    - NEA (+10.84%)
    - Centro (+10.48%)
    - Patagonia (+10.49%)



El resultado arrojado es predominantemente inelástico, lo que permite incrementos de precio generalizados sin comprometer el ingreso. Este resultado es coherente con la composición del portafolio, donde gran parte del peso económico está en Lácteos, Panadería, Frutas y Verduras y Bebidas, todas inelásticas.

- **Escenario 2: +10% solo en categorías inelásticas**
  - **Resultado general:** Aumento moderado, con un promedio de **+6.23%**.
  - **Regiones más favorecidas:**
    - NOA (+8.18%)
    - NEA (+6.43%)
    - Cuyo (+6.30%)

Consideramos esta como la **estrategia de bajo riesgo**, pues actúa únicamente sobre productos cuya demanda es estable y no existe amenaza inminente de una disminución en la demanda. Asimismo, la percibimos como la opción más viable para implementarse en regiones pequeñas o sensibles, como NOA, donde cambios amplios podrían generar mayor volatilidad.

- **Escenario 3: -10% en categorías muy elásticas**
  - **Resultado general:** Sin variación significativa (**0.0% en todas las regiones**), debido a que el dataset no contiene categorías clasificadas como “muy elásticas” bajo los criterios utilizados. Esto confirma que el portafolio no incluye productos con elasticidades extremas o que estos no presentan variación suficiente para estimarse.

Lo arrojado por este escenario demuestra que los descuentos no generan incremento apreciable en el ingreso total. Esto indica que el portafolio depende más de categorías inelásticas que de productos impulsivos o altamente sensibles al precio.

## ❖ Comparación y Conclusiones por Región

### Buenos Aires

- Es la región con el mayor volumen de ingreso base ( $\approx 37,000$ ), lo que amplifica el impacto absoluto de cualquier cambio.
- El escenario 1 genera el mayor incremento absoluto del portafolio.
- El escenario 2 ofrece mejoras moderadas pero seguras, reflejando una demanda relativamente más sensible

Esta es una región amplia y estable donde ambas estrategias de aumento funcionan, aunque con una elasticidad algo mayor

que en el resto del país, es decir, con un comportamiento levemente más sensible a los aumentos de precio.

### NOA

- Responde positivamente a los aumentos de precio, aun teniendo el menor ingreso base ( $\sim 2,450$ ).
- Presenta incrementos porcentuales altos ante aumentos del 10%, lo que evidencia una demanda poco sensible en escenarios de subida.
- El escenario 3 no tiene ningún beneficio, ya que los descuentos no aumentan los ingresos.

Esta región se caracteriza por una demanda poco sensible a los aumentos, pero también por una baja respuesta a descuentos, por lo que las reducciones de precio no generan pérdidas pero tampoco aportan beneficios adicionales.

### NEA, Centro, Cuyo y Patagonia

- Todas reaccionan positivamente a incrementos del 10%, mostrando comportamientos consistentes entre sí.
- Obtienen aumentos porcentuales estables y previsibles, reflejando baja sensibilidad al precio.
- Exhiben elasticidades alineadas con la inelasticidad general del portafolio de categorías.

En estas regiones, cualquier política de aumento gradual es viable y genera incrementos sostenidos de ingreso.

## ❖ Hallazgos globales

- 1) El portafolio está dominado por categorías inelásticas, lo que permite aumentar precios sin riesgo significativo.
- 2) Los productos de mayor impacto económico se concentran en el cluster 0, que corresponde también a categorías de ingreso estable.
- 3) Los roles A/B/C complementan los clusters confirmando que los productos estratégicos (cluster 0) corresponden también a categorías y productos clasificados como A en la mayoría de las regiones.
- 4) Los descuentos no generan beneficios, lo cual desecha la idea de que promover categorías sensibles impulsa el ingreso total.
- 5) La separación por clusters revela que las dinámicas económicas reales no coinciden con las categorías comerciales tradicionales.
- 6) Las estrategias de aumento de precio óptimas dependen del peso relativo de las categorías inelásticas, especialmente en regiones pequeñas donde su efecto se magnifica.

## VIII. LIMITACIONES Y RECOMENDACIONES

A pesar de la solidez y rigurosidad metodológica empleada en el análisis, consideramos importante reconocer los límites estructurales del dataset y del enfoque estadístico aplicado. Cabe destacar que estos condicionantes no invalidan los hallazgos, pero sí establecen el marco dentro del cual deben ser interpretados.

### 1) Naturaleza sintética del dataset

Como se mencionó en diversas ocasiones dentro del reporte, el conjunto de datos proviene de una simulación diseñada para emular el comportamiento del mercado argentino durante 2024. Aunque refleja patrones plausibles, no captura completamente el comportamiento real del mercado, la complejidad del comportamiento del consumidor ni los picos esperados durante fechas de alto consumo. Por lo tanto, los resultados deben interpretarse como una aproximación analítica coherente bajo este supuesto, pero no como una representación fiel y directa del comportamiento del mercado. No obstante,

### 2) Ausencia de costos, inventarios y márgenes comerciales

El análisis se concentra únicamente en ingresos, lo cual limita la evaluación integral de estrategias de pricing, dado que variaciones en el precio podrían tener efectos diferenciales sobre margen y rotación.

Al no contar con costos unitarios ni restricciones logísticas, las recomendaciones se orientan exclusivamente a maximizar ingreso y no necesariamente utilidad. Adicionalmente, los datos proporcionados acerca del almacenamiento de productos no pudieron ser utilizados para realizar recomendaciones reales, debido a que no reflejaban el estado actual o real de los productos ni inventarios.

### 3) Elasticidades estimadas a nivel categoría–región

El modelo asume respuestas homogéneas dentro de cada categoría en cada región. En un entorno real, la elasticidad puede variar entre productos específicos, presentaciones y marcas, especialmente en categorías con fuerte sustitución interna. Esto implica que las recomendaciones de precio operan al nivel “familia de productos” y no necesariamente al nivel artículo.

Habiendo mencionado las limitaciones que se reconocen en este análisis, se **recomiendan** las siguientes líneas de acción para fortalecer tomas de decisión futuras:

### 1) Profundizar análisis con datos reales, costos e inventarios

En pro de migrar de un enfoque basado en ingreso a uno basado en utilidad, habilitar decisiones de pricing, surtido y abastecimiento con impacto financiero directo, consideramos que una siguiente fase de este proyecto debería incorporar datos adicionales, que ayuden a apegarse al comportamiento de la industria, tales como:

- Costos unitarios
- Rotación y días de inventario
- Restricciones logísticas por región

### 2) Ajustar elasticidades al nivel producto o subcategoría

Como se mostró en este análisis, la elasticidad por categoría es útil para estimaciones globales, pero puede ocultar diferencias relevantes. Para poder implementar estrategias de precio más precisas y con menor riesgo, se recomienda avanzar hacia modelos más granulares que permitan:

- Elasticidad por producto (SKU)
- Elasticidad por marca
- Elasticidad diferenciada entre canales o regiones con alta dispersión

### 3) Validar escenarios de pricing con experimentos controlados (A/B tests)

Antes de una realizar una implementación real, se recomienda ejecutar pruebas piloto por región para poder ajustar parámetros antes de una aplicación generalizada y confirmar la previsibilidad del modelo; algunas pruebas propuestas son:

- Incrementos controlados en categorías inelásticas
- Reducción selectiva en categorías elásticas
- Evaluación semanal del impacto en volumen y rentabilidad