

Introducción

En este informe vamos a explayarnos sobre la metodología que seguimos para explorar los datos, las mejoras de calidad que realizamos sobre los datos del archivo “eci_transactions.csv” y “eci_store_clusters.csv”, los distintos insights y los métodos de ingeniería de atributos que utilizamos para generar un dataset de buena calidad para entrenar a nuestro modelo de predicción.

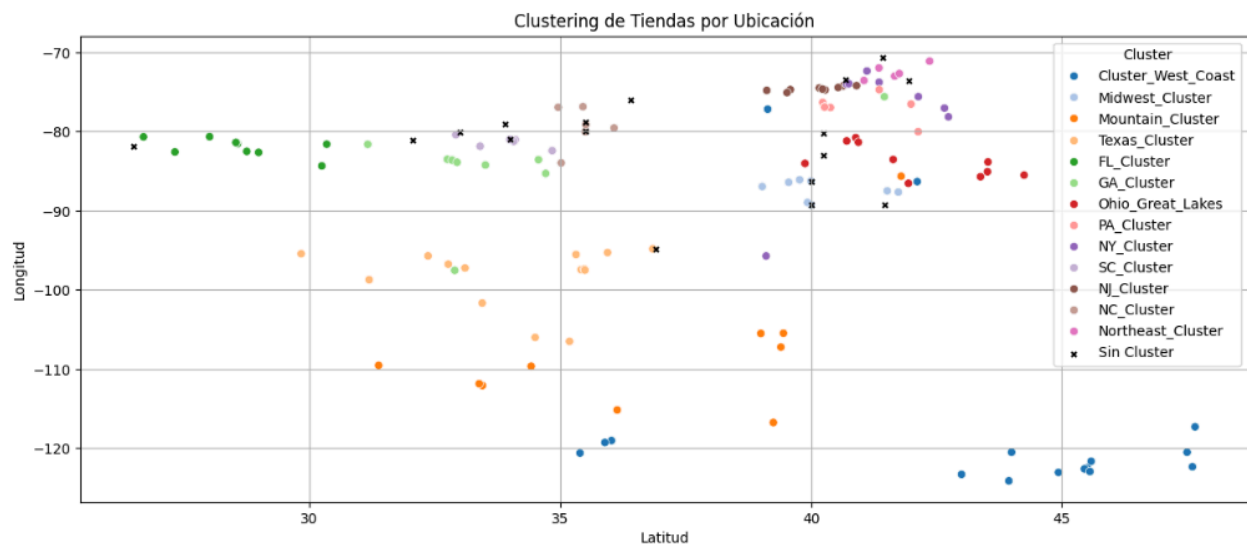
Preprocesamiento

Primero, seleccionamos la información relevante al problema de los siguientes archivos:

- “eci_transactions.csv”:
 - date
 - store_id
 - sku
 - quantity
 - price
 - total_sales
 - subgroup
- "eci_stores_clusters.csv":
 - store_id
 - store_name
 - cluster
- “eci_product_master.csv”:
 - sku
 - category
 - group
 - subgroup
- “eci_stores.csv”:
 - store_id

El dataset “eci_customers_data” tenía muy mala calidad y ninguna clave foránea que nos permitiera relacionar a los clientes con sus transacciones o con alguna tienda, entonces lo descartamos pero nos parece muy relevante que se reúna esa información para futuros análisis.

Las mejoras de calidad que realizamos fue para el dataset de “eci_store_cluster” que encontramos 17 tiendas que no tenían un cluster asignado. Lo que hicimos fue utilizar una api externa para obtener las coordenadas de todas las tiendas sin cluster asignado a partir de sus direcciones y poder determinar a qué cluster pertenecían visualmente



Para el archivo de “eci_transactions” notamos que la columna “quantity” tenía valores nulos pero que fueron fácilmente recuperables con las columnas “total_sales” y “price”.

Para la ingeniería de atributos, utilizamos 2 técnicas utilizadas para la predicción sobre series de tiempo: rolling windows sobre la demanda y los atributos de fourier. La primera da un sentido de la tendencia que sigue la demanda de nuestro producto. Nosotros tomamos para los 7 y 14 días anteriores la “rolling sum” y “rolling mean” de la demanda. Los atributos de fourier nos permite reemplazar los datos de día y mes por

senos y cosenos con distintos periodos. Esto nos permite detectar la tendencia estacional de la demanda. Ya que da la sensación de un continuo y eliminamos el “falso orden” de los meses del 1 al 12, ya que el mes diciembre y enero son muy parecidos pero muy distintos numéricamente.

Modelo de predicción

Utilizamos el algoritmo XGBoost para predecir la demanda, probamos 2 enfoques de predicción un modelo de predicción recursiva y otro simple y entrenarlo sobre distintos conjuntos de datos, todo el dataset y entrenado sobre cada categoría de datos. Estos modelos los validamos con datos históricos y contra el score público del kaggle y los alimentamos con distintos tipos de atributos. Intentamos darles señales lo más relevantes y distintas posibles. Los atributos que utilizamos fueron clusters, rolling windows y los atributos de fourier. Esto nos permitió darle al modelo un sentido geográfico de las tiendas, de tendencia y temporal.

Hicimos un post procesado para un subgrupo que no tuvo ninguna venta en los 3 años de datos de los que tenemos registro. Es el subgrupo “Basketball” así que todas sus predicciones las rellenamos con 0, para ser coherentes con lo que se fue observando.

Elasticidad y Precio de equilibrio

a esta parte del proyecto, se trabajó con el archivo “eci_transactions.csv”, que contiene información detallada de cada operación, incluyendo la fecha (DATE), el producto (SKU), la cantidad vendida (QUANTITY) y el precio aplicado (PRICE). En una primera etapa fue necesario organizar y limpiar los datos, de modo que pudieran ser utilizados en las estimaciones posteriores.

En particular, se identificó que un mismo producto podía registrar múltiples observaciones en un mismo día, ya sea porque el SKU se vendió en diferentes tiendas o porque se registraron varias operaciones en esa fecha. Para evitar que esta multiplicidad distorsionara los cálculos, se decidió agrupar los registros diarios por SKU, sumando las cantidades vendidas y calculando un precio promedio. Con este procedimiento se obtuvo una serie temporal diaria de ventas y precios para cada producto, lo cual permitió trabajar con información consolidada y representativa.

Una vez preparadas las series, se procedió al cálculo de la elasticidad a través de dos enfoques complementarios.

El primero de ellos fue el **método del arco**, que utiliza únicamente los valores máximos y mínimos de precio y cantidad observados en el período analizado. Con esta información se calcula una elasticidad promedio que refleja cómo se comporta la demanda entre los extremos del rango de precios y ventas. Este método es particularmente útil porque es simple de implementar y ofrece una visión rápida de la sensibilidad del producto sin necesidad de modelos estadísticos complejos. Sin embargo, su principal limitación es que se basa solo en dos puntos, dejando de lado gran parte de la información disponible.

Para superar esta restricción, también se implementó un **modelo de regresión log-log**. En este caso, se transformaron los datos aplicando logaritmos tanto a las cantidades vendidas como a los precios, y se estimó un modelo de regresión lineal entre ambas variables. El coeficiente asociado a la variable $\log(P)$ representa la elasticidad. Este segundo método es más robusto porque aprovecha toda la información de la serie temporal, en lugar de centrarse únicamente en los valores extremos. Además, permite reducir el impacto de la variabilidad puntual de algunos días, al capturar una tendencia general en el comportamiento de la demanda.

El resultado de esta etapa fue un DataFrame consolidado que para cada SKU incluye la elasticidad estimada por ambos métodos. Esto permite comparar resultados y detectar posibles inconsistencias, a la vez que brinda un rango de referencia sobre la sensibilidad de cada producto.

Por otro lado, si bien la estimación de elasticidad es un paso clave, para poder derivar precios óptimos resulta necesario complementarla con información sobre el contexto actual de cada producto. Por esta razón, se incorporó un análisis de los precios vigentes en la última fecha disponible de la base de datos.

En primer lugar, se identificó la fecha más reciente en que cada SKU tuvo registro de ventas. Luego, para cada producto en esa fecha se calcularon dos estadísticas: el precio promedio y el desvío estándar del precio. El precio promedio constituye una referencia central que sintetiza el nivel actual al que se comercializa el producto. El desvío estándar, en cambio, permite medir la variabilidad de precios dentro de esa misma fecha, lo cual es relevante en casos donde un mismo SKU se comercializa en distintas tiendas o bajo distintas condiciones que generan dispersiones. De este modo, se obtuvo una caracterización del entorno de precios reciente que sirve para vincular la elasticidad con la situación actual de cada producto.

Es importante destacar que en este análisis también se detectó un caso particular: un producto que no registró ventas en la última fecha considerada. Como consecuencia, no fue posible calcular un precio promedio ni un desvío estándar para dicho SKU, lo que se reflejó en valores nulos (NaN) en la base de resultados. Si bien se trata de una excepción, constituye un recordatorio de la importancia de revisar la calidad de los datos y de contemplar que pueden existir productos con dinámicas de demanda atípicas que requieren un tratamiento diferenciado.

Con los resultados de elasticidad y precios recientes se desarrolló un esquema de cálculo para obtener un precio óptimo sugerido. Para ello se definió un factor de ajuste, que se obtiene como la elasticidad dividida por la suma de la elasticidad más uno. En este caso, se utilizó la elasticidad estimada a través del modelo de regresión log-log, dado que este enfoque es más robusto al aprovechar toda la información disponible y reducir el impacto de valores extremos. Esta elección asegura que el cálculo del precio óptimo se base en una estimación más estable y representativa del comportamiento real de la demanda.

El precio óptimo se calculó como el producto del precio promedio de la última fecha y este factor de ajuste. La interpretación es sencilla: se trata de ajustar el precio actual en función de la sensibilidad histórica de la demanda, buscando un punto intermedio entre maximizar ingresos y no afectar negativamente las cantidades vendidas.

Finalmente, el proceso completo dio como resultado un conjunto de métricas para cada SKU. Entre ellas se incluyen la elasticidad estimada por el método del arco, la elasticidad obtenida a través de regresión, el precio promedio y el desvío estándar en la última fecha, el factor de ajuste calculado y, finalmente, el precio óptimo sugerido.

Este conjunto de información se consolidó en un DataFrame y se exportó a un archivo CSV para facilitar su análisis y posterior utilización en la estrategia comercial.

Los resultados permiten observar diferencias notorias entre productos. Algunos presentan elasticidades relativamente bajas, lo cual indica que podrían tolerar incrementos de precio sin una caída significativa en las ventas. Otros, en cambio, muestran elasticidades elevadas, lo que sugiere que cualquier incremento de precio podría traducirse en una reducción proporcionalmente mayor de la cantidad demandada. Estas diferencias ofrecen un insumo valioso para segmentar las políticas de precios según la naturaleza de cada SKU.

A continuación se observa una tabla con los datos de los primeros 10 productos con mayor desviación estándar.

SKU	Precio Promediado	Desviación Estándar del Precio	Elasticidad x Regresión	Factor de Ajuste	Precio Óptimo
ELEMOSM012	348.80	24.33	-1.82	2.21	772.09
ELEAUHE011	278.70	24.13	-1.80	2.26	629.12
ELEMOAC003	340.90	22.74	-1.45	3.22	1098.71
ELEMOWE010	357.69	21.31	-1.34	3.93	1405.93
ELEMOWE012	323.93	20.61	-2.32	1.76	568.53
ELEMOSM009	261.14	20.56	-1.73	2.37	620.12
ELEAUSP008	353.80	20.19	-1.10	13.84	4896.93
ELEAUSP006	271.39	20.06	-1.55	2.83	767.78
ELECODE008	305.71	19.72	-1.62	2.62	802.33
ELEAUHE014	266.55	18.86	-1.30	4.46	1188.44

Los productos listados presentan una alta variabilidad de precios, aunque es importante remarcar que no basta con observar la desviación estándar en valores absolutos, sino también su relación con el precio promedio. No es lo mismo una desviación de 20 sobre un precio promedio de 300 (variabilidad moderada), que la misma desviación sobre un precio de 50 (variabilidad muy alta).

En todos los casos, las elasticidades estimadas son negativas, lo que confirma la relación inversa entre precio y demanda. Productos como ELEMOSM012 (-1.82) o ELEAUSP006 (-1.55) muestran una demanda sensible al precio, mientras que ELEAUSP008 (-1.10) presenta menor sensibilidad. Este último destaca por un factor de ajuste inusualmente alto (13.84), que lleva a un precio óptimo extremo, probablemente reflejo de datos atípicos, promociones puntuales o limitaciones de la fórmula aplicada.

En general, se observa que las elasticidades más negativas producen factores de ajuste menores y precios óptimos más moderados (por ejemplo, ELEMOWE012, con -2.32 y un precio óptimo de 568.53). Esto evidencia que la elasticidad calculada mediante regresión influye de manera decisiva en la recomendación de precios. Sin embargo, es importante aclarar que la tabla corresponde solo a los 10 productos con mayor desviación estándar, por lo que no representa la totalidad de los datos.

El CSV con todos los datos completos se encuentra en el archivo .zip como "Datos_Elasticidad_Precio_Optimo.csv"

Consideraciones adicionales y oportunidades de mejora

Si bien el análisis realizado permite estimar elasticidades y sugerir precios óptimos, existen varias fuentes de información adicionales que podrían mejorar la calidad y robustez de los resultados. Por ejemplo, incorporar factores externos que influyen en la demanda, tales como cambios en la economía (inflación, variaciones de poder adquisitivo), eventos políticos o regulatorios que afecten la disponibilidad de productos, o fenómenos climáticos que modifiquen el comportamiento de compra, permitiría detectar patrones que no se observan únicamente con los datos de transacciones.

Asimismo, considerar diferencias geográficas entre tiendas o regiones podría aportar claridad sobre cómo la demanda responde al precio en distintos mercados, lo que permitiría ajustar estrategias de pricing de manera localizada. Otro factor relevante son las fechas importantes del año, como festividades, campañas promocionales o cambios estacionales, que pueden generar variaciones significativas en las ventas y, por lo tanto, afectar la elasticidad de la demanda.

Finalmente, contar con información adicional sobre inventarios, competencia o promociones activas podría enriquecer aún más el análisis, permitiendo diferenciar cuándo una caída de ventas se debe a la sensibilidad al precio o a limitaciones externas. En conjunto, estos datos complementarios permitirían construir modelos más sofisticados y precisos, capaces de capturar patrones más finos en la relación entre precio y demanda, y de generar recomendaciones de pricing más certeras y estratégicas.

Conclusiones y agradecimientos

El desarrollo de este proyecto permitió transformar un conjunto de datos dispersos e incompletos en un dataset de alta calidad, apto para entrenar modelos predictivos y extraer información valiosa sobre el comportamiento de la demanda. A través de técnicas de ingeniería de atributos, validación de modelos y estimación de elasticidades, se logró no sólo anticipar patrones de consumo, sino también proponer precios óptimos ajustados a la sensibilidad de cada producto. De este modo, consideramos que pudimos resolver el desafío planteado y conseguir una posición satisfactoria en la competencia a pesar de todas las dificultades que se nos presentaron.

Los resultados evidencian la relevancia de considerar tanto la variabilidad del precio como su relación con la demanda, ya que no todos los productos reaccionan de la misma forma ante cambios en el precio. Esto abre la posibilidad de diseñar estrategias de pricing segmentadas, más alineadas con la dinámica real del mercado.

Si bien se superaron limitaciones técnicas y de calidad de datos, el trabajo también dejó en claro la importancia de incorporar fuentes externas (económicas, estacionales, competitivas y geográficas) para mejorar la robustez de los modelos y mejorar la precisión de las recomendaciones. De este modo, el análisis realizado constituye una base sólida sobre la cual seguir construyendo soluciones más completas y estratégicas.

Finalmente, queremos darle las gracias al equipo de Alix Partner por organizar y darnos la posibilidad de participar en esta competencia, que representó tanto un interesante desafío técnico como un aprendizaje significativo para todos los integrantes del equipo.