

Máster en Big Data & Analytics (Business Intelligence)

TRABAJO FINAL DE MÁSTER

Modelo predictivo para la estimación de ventas de Tiendas Aditivo.



Tutor:

Aleix Ruiz de Villa Robert

Alumnos:

Andrés Castro Ganoza
Edmundo Charabati Gindi
David Figuerola Serrano
Óscar Maestre Cuscó

Barcelona, septiembre 2023

Índice Trabajo Final de Máster (TFM)

| | | |
|----|---|----|
| 1. | Presentación de Tiendas Aditivo | 3 |
| 2. | Contexto y objetivo del TFM | 6 |
| 3. | Transformación de datos | 7 |
| | Fase 1: obtención y preparación de los datos: | 7 |
| | Fase 2: selección de datos: | 8 |
| | Fase 3: codificación de datos: | 8 |
| 4. | Modelos predictivos y resultados | 9 |
| 5. | Uso del modelo predictivo a futuro | 16 |
| 6. | Conclusiones y recomendaciones | 17 |

1. Presentación de Tiendas Aditivo

Aditivo es una empresa textil mexicana con más de medio siglo de trayectoria, que se ha consolidado como una de las líderes en la industria mexicana. Fundada como un pequeño taller de confección ha evolucionado hasta convertirse en un importante referente en el mercado textil nacional gracias a su enfoque en la innovación y la calidad.

El portafolio de productos de Aditivo es amplio y variado, abarcando más de diez familias de productos. Este rango incluye desde accesorios como gorras hasta prendas de vestir como camisetas en diversas tallas, así como diferentes estilos como urbano o deportivo, adaptándose a las necesidades y preferencias de una amplia gama de clientes. Esta diversificación de productos es clave para satisfacer las demandas del mercado y asegurar una cobertura integral de las preferencias de consumo.

Actualmente Aditivo opera 80 puntos de venta estratégicamente distribuidos para maximizar su alcance y penetración en el mercado. Estos se encuentran ubicados en ejes comerciales de las principales ciudades de México, centros comerciales, estaciones de transporte público y establecimientos con los que se tienen alianzas estratégicas como el caso de *Price Shoes*¹.

Además de la red de puntos de venta físicos, Aditivo ha diversificado sus canales de venta para incluir ventas por catálogo a través de asociados y una plataforma de venta online.

Esta estrategia multicanal no solo amplía su mercado objetivo, sino que también facilita a los clientes el acceso a sus productos de manera más conveniente y flexible.

La venta por catálogo permite a los asociados distribuir los productos de Aditivo de manera personalizada, obteniendo además precios exclusivos, mientras que la venta online proporciona una opción cómoda para los consumidores modernos que prefieren realizar compras desde la comodidad de sus hogares.



En términos de gestión de inventario y oferta de productos, Aditivo maneja más de 600 SKU's (*Stock Keeping Units*), lo que refleja la diversidad y profundidad de su catálogo². La gestión eficiente de un número tan considerable de SKU's es crucial para asegurar la disponibilidad de productos y mantener altos niveles de satisfacción del cliente. Este enfoque en la gestión del

¹ <https://www.priceshoes.com/> es una empresa que ofrece moda en zapatos, ropa y joyas de venta por catálogo con presencia en toda la República Mexicana y parte de América Latina.

² Para Aditivo cada Catálogo equivale a una colección de productos que por lo general se mantiene en piso de venta entre 3 y 4 meses.

inventario y la oferta de productos demuestra el compromiso de Aditivo con la calidad y la atención al detalle en su operación diaria.

A lo largo de los años, la empresa amplió su gama de productos para incluir una variedad de textiles, desde prendas de vestir hasta accesorios como gorras, carteras o bisutería.

En 2009, Aditivo decidió diversificar sus operaciones e incursionar en el sector minorista bajo su propia marca. Este nuevo enfoque llevó a la apertura de cuatro tiendas propias, pero el intento inicial no tuvo el éxito esperado. La combinación de una estrategia que intentaba satisfacer simultáneamente a clientes mayoristas y minoristas, junto con ubicaciones desfavorables y la falta de experiencia en el sector *retail*, resultó en el cierre de estas tiendas.



Tras este revés, Aditivo optó por una nueva estrategia: la venta por catálogo. En 2010, se inauguró la primera tienda de catálogo en un centro de *Price Shoes*. Esta iniciativa resultó ser exitosa y marcó el inicio de una etapa de crecimiento sostenido para la empresa. En 2012, Aditivo tomó la decisión estratégica de dejar de fabricar mercancía para terceros y enfocarse exclusivamente en su marca. Este cambio permitió a la empresa redimensionar su fábrica y dedicarse de lleno al negocio de retail por catálogo.

Desde entonces, Aditivo ha experimentado un crecimiento significativo, expandiendo su presencia en más de la mitad de los estados de la República Mexicana, con un énfasis particular en las regiones del centro y sureste del país. Actualmente, la empresa es gestionada por la segunda y tercera generación de la familia fundadora.

Pese a su éxito, Aditivo enfrenta actualmente un desafío importante: la disminución anual en las ventas por catálogo, sin una comprensión clara de las causas de esta tendencia. A pesar de esta problemática, la empresa ha retomado la venta a cadenas comerciales como parte de su estrategia para adaptarse a las cambiantes condiciones del mercado y asegurar su sostenibilidad a largo plazo.



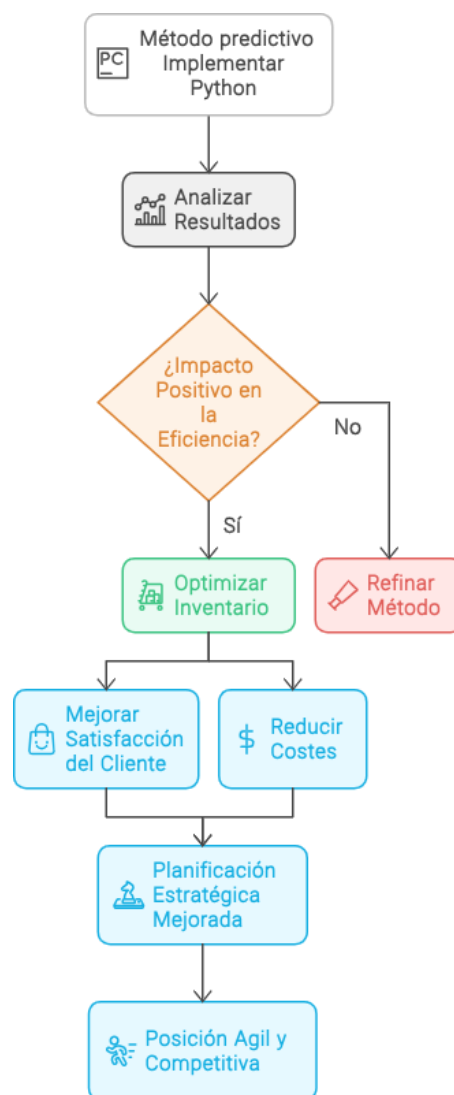
En la imagen se puede apreciar la expansión actual (septiembre de 2024) de los estados de México donde Aditivo tiene presencia en tiendas físicas, aunque por venta por catálogo y web venden a todo México.

2. Contexto y objetivo del TFM

El presente Trabajo Final de Máster (TFM) se centra en la identificación del método predictivo más eficaz para proyectar las ventas de la empresa textil mexicana Aditivo, utilizando el lenguaje de programación *Python*. Este estudio es esencial para la optimización de la planificación estratégica y operativa de la empresa, minimizando los costes de almacenaje y permitiendo una gestión más eficiente de sus recursos y una respuesta proactiva a las fluctuaciones del mercado.

La capacidad de anticipar la demanda con mayor exactitud permitirá a la empresa ajustar sus procesos de producción y adquisición de materiales, resultando en una mejora significativa en la eficiencia operativa y una disminución de costes. Además, la optimización del inventario garantizará una disponibilidad adecuada de productos en los puntos de venta, mejorando la satisfacción del cliente y reduciendo el capital inmovilizado en productos que no tienen una alta rotación. En conjunto, estas mejoras contribuirán a una gestión más ágil y rentable, fortaleciendo la posición competitiva de la empresa en el mercado.

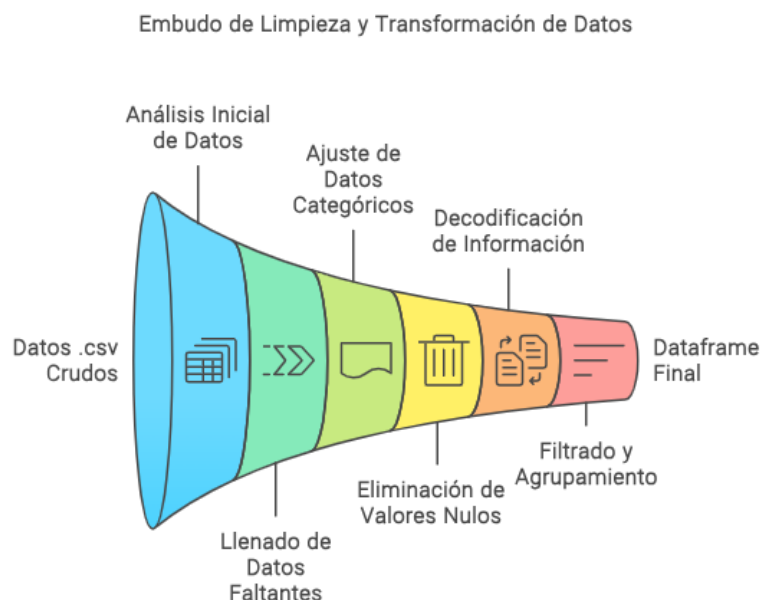
Este trabajo se plantea como una Prueba de Concepto (PoC) cuyo objetivo es demostrar la viabilidad y el potencial de implementar un modelo predictivo avanzado para la gestión de la demanda en Aditivo. Se busca proporcionar a la dirección de la empresa una base sólida y basada en datos que les permita evaluar los beneficios de una mayor inversión en tecnologías de *machine learning* para optimizar su inventario y mejorar la eficiencia operativa. La decisión final sobre la implementación a gran escala de este modelo dependerá de los resultados presentados en esta PoC y de su alineación con los objetivos estratégicos de la empresa.



El trabajo se basa en datos de ventas recolectados durante los últimos dos años y medio (30 meses de enero 2022 hasta junio 2024). Este periodo ofrece una ventana suficiente para captar tanto las tendencias estacionales como las fluctuaciones anuales, proporcionando una base sólida para el desarrollo de modelos predictivos precisos. Los datos históricos de ventas incluyen diversas variables relevantes que influyen en el comportamiento del consumidor y las ventas, tales como promociones, eventos especiales y tendencias de moda.

3. Transformación de datos

Una vez recopilados los datos enviados por Aditivo, se aplicaron una serie de pasos con el fin de contar con un *dataset final* listo para ser usado en los modelos predictivos. Estos pasos fueron agrupados en tres fases de transformación: preparación y limpieza, selección y codificación.



Fase 1: obtención y preparación de los datos:

Después de recibir el archivo .csv por parte de Aditivo con la información de las ventas de todos los almacenes y todos los artículos de los últimos 30 meses (hasta junio 2024), se cuantificó alrededor de 5 millones de registros y aproximadamente 40 columnas de datos. Esta información requería un procesamiento exhaustivo antes de poder utilizarla en un modelo predictivo.

Se corrigieron las inconsistencias y se llenaron los datos faltantes, como códigos postales o información de direcciones. También se ajustaron varias columnas categóricas, incluyendo el tipo de tienda, categoría, clima, entre otras. Además, se eliminaron todos los valores nulos, ya que el modelo predictivo no los acepta.

Toda la información recibida estaba codificada, pero también se nos proporcionaron catálogos con las descripciones correspondientes, por lo que se sustituyeron los códigos por sus descripciones largas; por ejemplo, en lugar de que el departamento apareciera como "9C", se cambió a "Caballeros". Este proceso se aplicó a unos 15 campos codificados de manera similar.

Una vez transformados los datos, se agruparon todas las ventas por semana. Posteriormente, se realizó un filtrado de datos para refinar aún más el conjunto. Se tomaron en cuenta solo los departamentos de "Chavos" y "Chavas", ya que son el foco principal de Aditivo. También se eliminaron las tiendas tipo Outlet y los SKUs no resurtibles, enfocando el análisis solo en almacenes y productos activos.

Tras este proceso, el conjunto de datos quedó reducido aproximadamente a 136.000 registros y 24 columnas. Este resultó como el nuevo *dataset* que se usó en la siguiente fase.

Fase 2: selección de datos:

En primer lugar, se seleccionaron los 75 mejores almacenes basándose en sus ventas. Se eliminaron los almacenes de distribución que no son almacenes de venta para que no generen distorsiones en los resultados.

De cada uno de estos almacenes, se identificaron sus 10 mejores SKUs³, asegurando así que se evaluarán los artículos más representativos en términos de ventas en cada almacén.

Se hallaron las semanas en las que no había ventas y se rellenaron esos vacíos con ceros a partir de la semana en que el SKU comenzó a venderse. Es importante destacar que no se rellenaron las semanas anteriores al lanzamiento del SKU con ceros, ya que el producto aún no existía en ese momento. Por ejemplo, si un SKU comenzó a venderse en la semana 30 y no tuvo ventas en las semanas 31 y 32 en un almacén específico, se agregaron esos períodos y se asignaron un valor de 0 para que sean considerados en la evaluación de desempeño de ese producto en esa ubicación.

Una vez filtrados los SKUs, se decidió probar varios modelos para cada combinación de “Almacén-SKU”. Se utilizó la validación cruzada para determinar cuál modelo era el mejor en cada caso. Además, para garantizar la relevancia y precisión de la información, se seleccionaron sólo aquellos SKU con al menos 15 semanas de ventas. Esto permite tener suficientes datos para trabajar y hacer una validación cruzada efectiva⁴.

Fase 3: codificación de datos:

Finalmente, se determinaron cuáles eran las variables categóricas más relevantes. Esto porque hubo columnas que tenían solo una categoría repetida para todas las filas o otras que tenían datos repetitivos, por lo tanto, no agregaban valor, así que se eliminaron y se aplicó un *one-hot encoding*⁵ a las restantes. Con los datos listos, se separó el conjunto en entrenamiento y prueba, utilizando las últimas 3 semanas como conjunto de prueba y el resto como conjunto de entrenamiento. Así, en cada caso, había un mínimo de 12 semanas de datos para entrenamiento y 3 para prueba.

³ En lugar de los mejores 10 SKU's del total catálogo, para no generar interferencias entre los datos que se venden más en un almacén que en otro.

⁴ La validación cruzada es una técnica utilizada para medir el rendimiento de un modelo cuando se enfrenta a datos nuevos que no fueron utilizados durante el entrenamiento.

⁵ El *One Hot Encoding* es una técnica utilizada para convertir datos categóricos (como palabras o etiquetas) en un formato numérico que los algoritmos de aprendizaje automático puedan entender.

4. Modelos predictivos y resultados

Después de la transformación de los datos, se procedió a investigar diversos modelos de regresión que ayuden a predecir con mejor precisión las ventas futuras. Cabe destacar que Aditivo actualmente cuenta con un modelo de predicción de ventas futuras basado en el promedio de las últimas seis semanas, el cual se usó como *baseline* en este trabajo.

Los modelos seleccionados y aplicados en el presente trabajo fueron los siguientes⁶:

| Modelo | Breve descripción |
|-----------------------------|---|
| Aditivo Actual | Se basa en el promedio histórico de ventas de las últimas 6 semanas para predecir el futuro. |
| Aditivo Modificado | Se basa en el mejor promedio histórico de ventas de las n (hiperparámetro) semanas para predecir el futuro. Este fue un modelo de elaboración propia y se iteró entre 4 y 8 semanas. |
| Regresión de <i>Poisson</i> | Analiza datos que representan cuentas o frecuencias de eventos en un periodo de tiempo o espacio fijo y predice basándose en ellos aplicando una función logarítmica. |
| <i>Prophet</i> | Modelo desarrollado por <i>Facebook</i> . Predice series temporales incluso cuando se tienen datos incompletos o irregulares. Es especialmente útil para datos que tienen patrones estacionales (como fluctuaciones diarias, semanales o anuales) y tendencias a largo plazo. |
| <i>Random Forest (rf)</i> | Predice de manera automatizada basándose en múltiples árboles de decisión ⁷ . |
| <i>XG Boost</i> | Crea una serie de modelos simples (generalmente árboles de decisión) de manera secuencial en el que cada modelo intenta corregir los errores cometidos por el modelo anterior. Al final, todos estos modelos se combinan para hacer una predicción más precisa. |

La mayoría de los modelos descritos en el cuadro de arriba requieren de entrenamiento con datos históricos e inversión de tiempo en la ejecución de los códigos. Como se detalló anteriormente, debido a la gran cantidad de datos disponibles y para hacer el proceso más eficiente, se decidió seleccionar únicamente los 10 SKUs principales de cada uno de los 75 almacenes que tiene Aditivo a nivel nacional.

⁶ Los modelos *Sarima*, *KNN* y regresión lineal quedaron descartados dado que no ofrecieron resultados satisfactorios.

⁷ Un árbol de decisión es una estructura que se parece a un árbol con ramas. Cada nodo del árbol representa una pregunta o una condición sobre los datos, y cada rama es una posible respuesta a esa pregunta. Al final de cada rama, hay una "hoja" que representa la predicción o decisión final.

Con ello, se logró generar 750 combinaciones “Almacén – SKU” que fueron procesados por cada uno de los modelos y se definió el mejor para cada una de las combinaciones. Para ello se realizó una iteración para todos los periodos de tiempo (por semana) que se les llamó *lags*⁸ temporales. Por cada número de *lag* (4, 5, 6 o 7 semanas), se probaron todos los modelos seleccionados, se buscó la mejor combinación de hiperparámetros y la mejor versión de cada modelo para ese *lag* específico.

Existen varias métricas para medir la precisión de un modelo predictivo. Para este trabajo, se decidió por el Error Absoluto Medio o MAE⁹ usado para medir cada una de las combinaciones.

En términos simples, el MAE dice en promedio, cuánto se equivocó el modelo en sus predicciones al compararlas con los datos reales. Si el MAE del modelo es más próximo a cero, quiere decir que es más preciso. Con esta métrica se pudo medir la validación cruzada de cada uno de los modelos.

Como ya se mencionó anteriormente, los entrenamientos de los modelos fueron con datos históricos desde enero 2022 a mayo 2024. Las predicciones de los modelos consistieron en las ventas de las semanas 1, 2 y 3 de junio 2024 y se compararon con las ventas reales de estos mismos periodos. El modelo “Aditivo Actual” fue el que se usó como línea base; se calculó su MAE y el objetivo fue reducirlo de manera significativa con el resto de los modelos descritos.

En la siguiente imagen, se muestra los resultados del MAE para cada uno de los modelos entrenados y evaluados, y el mejor para cada una de las combinaciones “Almacén – SKU”:



Resultados de los modelos predictivos analizados

| 75 | | 750 | | | 10 | | 32 | | 6 | | | |
|---------------------------|---------|---|--------------------|-----------------|-----------------------------------|----------------------|-------------------------------------|---------------------------------------|--------------------------------|----------------|----------------------|-----------------|
| Almacenes | | Combinaciones de Almacén - SKU analizados | | | SKUs más vendidos de cada Almacén | | SKUs distintos | | Modelos predictivos analizados | | | |
| Combinación Almacén - SKU | Línea | Sublínea | Año de lanzamiento | Estilo | Mejor Modelo | MAE del mejor modelo | Modelo Aditivo (promedio 6 semanas) | Modelo Aditivo 2 (promedio n semanas) | Modelo Poisson | Modelo Prophet | Modelo Random Forest | Modelo XG Boost |
| 4 CAMINOS - T23C2139CH | PLAYERA | MANGA LARGA | 2021 | CASUAL | xgboost | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 |
| 4 CAMINOS - T23C2139EG | PLAYERA | MANGA LARGA | 2021 | CASUAL | xgboost | 0.00 | 0.00 | 0.00 | 0.00 | 0.32 | 0.00 | 0.00 |
| 4 CAMINOS - T23C2139GD | PLAYERA | MANGA LARGA | 2021 | CASUAL | xgboost | 0.00 | 1.00 | 0.00 | 0.00 | 0.43 | 0.00 | 0.00 |
| 4 CAMINOS - T23C2139MD | PLAYERA | MANGA LARGA | 2021 | CASUAL | xgboost | 0.33 | 0.67 | 0.67 | 0.33 | 0.37 | 0.33 | 0.33 |
| 4 CAMINOS - T6C01007CH | PLAYERA | MANGA CORTA | 2013 | COOL | xgboost | 0.00 | 0.00 | 0.00 | 0.00 | 0.35 | 0.00 | 0.00 |
| 4 CAMINOS - T6C01007EG | PLAYERA | MANGA CORTA | 2013 | COOL | xgboost | 0.00 | 0.00 | 0.00 | 0.00 | 0.65 | 0.00 | 0.00 |
| 4 CAMINOS - T6C01007GD | PLAYERA | MANGA CORTA | 2013 | COOL | xgboost | 0.00 | 0.00 | 0.00 | 0.00 | 0.57 | 0.00 | 0.00 |
| 4 CAMINOS - T6C01007MD | PLAYERA | MANGA CORTA | 2013 | COOL | xgboost | 0.00 | 0.00 | 0.00 | 0.00 | 0.34 | 0.00 | 0.00 |
| 4 CAMINOS - TA4D3027MD | PLAYERA | MANGA CORTA | 2024 | SIN ESTILO_VIDA | prophet | 0.53 | 1.33 | 0.67 | 2.00 | 0.53 | 4.00 | 4.67 |
| 4 CAMINOS - TBAD3009MD | PLAYERA | MANGA CORTA | 2023 | BÁSICOS | aditivo_modificado | 0.00 | 0.00 | 0.00 | 2.00 | 1.28 | 2.00 | 2.00 |
| APIZACO - T23C2139CH | PLAYERA | MANGA LARGA | 2021 | CASUAL | poisson | 0.67 | 1.00 | 1.00 | 0.67 | 0.67 | 1.00 | 1.67 |
| APIZACO - T23C2139EG | PLAYERA | MANGA LARGA | 2021 | CASUAL | xgboost | 0.00 | 0.00 | 0.00 | 0.00 | 0.16 | 0.00 | 0.00 |
| APIZACO - T23C2139GD | PLAYERA | MANGA LARGA | 2021 | CASUAL | xgboost | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 |
| APIZACO - T23C2139MD | PLAYERA | MANGA LARGA | 2021 | CASUAL | rf | 0.00 | 0.67 | 0.67 | 0.33 | 0.47 | 0.00 | 0.33 |
| APIZACO - T23P5C01MD | PLAYERA | MANGA LARGA | 2023 | URBANO | aditivo_modificado | 0.33 | 0.33 | 0.33 | 1.33 | 0.73 | 1.00 | 1.00 |
| APIZACO - T6C01007EG | PLAYERA | MANGA CORTA | 2013 | COOL | xgboost | 0.00 | 0.00 | 0.00 | 0.00 | 0.13 | 0.00 | 0.00 |
| APIZACO - T6C01007GD | PLAYERA | MANGA CORTA | 2013 | COOL | xgboost | 0.00 | 0.00 | 0.00 | 0.00 | 0.06 | 0.00 | 0.00 |
| APIZACO - T6C01007MD | PLAYERA | MANGA CORTA | 2013 | COOL | xgboost | 0.00 | 0.00 | 0.00 | 0.00 | 0.06 | 0.00 | 0.00 |
| APIZACO - TBAD3008MD | PLAYERA | MANGA CORTA | 2023 | BÁSICOS | xgboost | 0.00 | 0.00 | 0.00 | 2.00 | 0.85 | 0.00 | 0.00 |

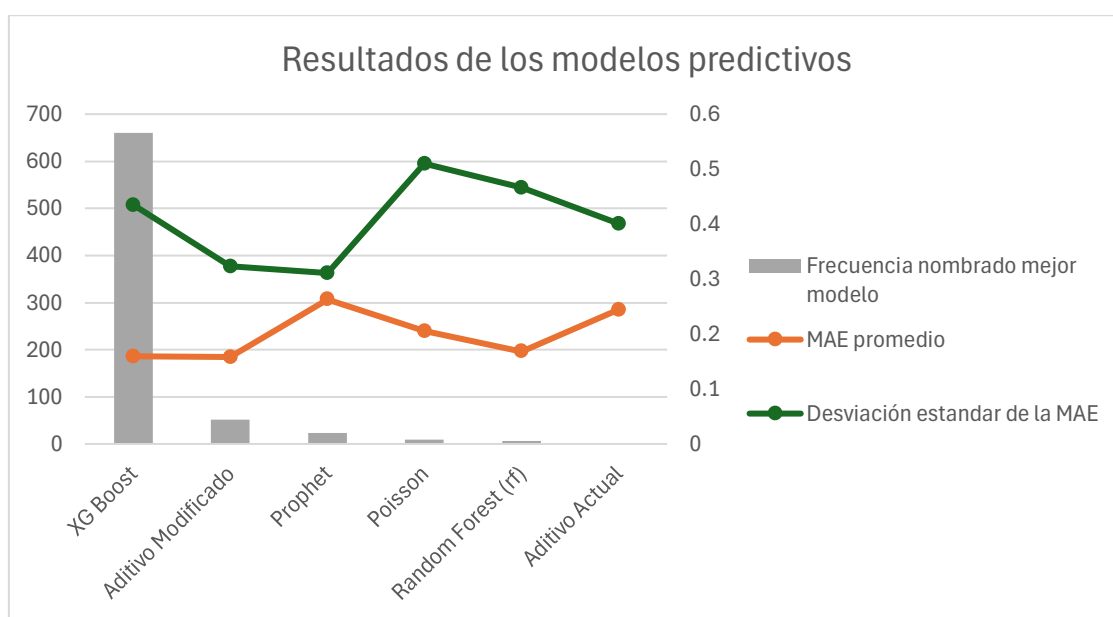
⁸Los *lags* temporales representan cuántas semanas de datos históricos se utilizan para hacer predicciones. Por ejemplo, un lag de 4 semanas significa que se usan datos de las últimas 4 semanas para predecir la siguiente. La selección del mejor lag ayuda a determinar cuánta historia pasada es más útil para hacer predicciones precisas en este contexto específico.

⁹ En inglés Mean Absolute Error (MAE).

Como se puede observar a modo de ejemplo, el SKU “T23C2139CH” del almacén “4 CAMINOS” cuyo año de lanzamiento fue en el 2021 tuvo como mejor modelo de predicción el *XG Boost*.

Esta visualización proviene de un reporte de *Power BI* en la que se conectó un archivo CSV generado por *Python* con los resultados de los modelos obtenidos.

Después de analizar los resultados para las 750 combinaciones, se concluye que *XG Boost* fue nombrado con mayor frecuencia como mejor modelo de predicción, seguido de Aditivo Modificado y *Prophet*. En el siguiente gráfico se puede observar la distribución de los mejores modelos por cada una de las combinaciones, su desviación estándar y MAE promedio.



| Modelo | MAE promedio | Frecuencia nombrado mejor modelo | Desviación estándar de la MAE |
|---------------------------|---------------|----------------------------------|-------------------------------|
| <i>XG Boost</i> | 0.1591 | 660 | 0.4346 |
| Aditivo Modificado | 0.1582 | 51 | 0.3229 |
| <i>Prophet</i> | 0.2627 | 23 | 0.3109 |
| <i>Poisson</i> | 0.2053 | 9 | 0.5100 |
| <i>Random Forest (rf)</i> | 0.1680 | 7 | 0.4663 |
| Aditivo Actual | 0.2444 | 0 | 0.4008 |
| Total Matriz | 0.0706 | 750 | 0.4076 |

Aunque inicialmente XGBoost aparece con mayor frecuencia como el mejor modelo (660 veces), presenta un MAE ligeramente superior (0.1591). Esto sugiere que, a pesar de su popularidad, el modelo Aditivo Modificado podría ser una opción más sencilla y fácil de interpretar, gracias a su equilibrio entre rendimiento y desviación.

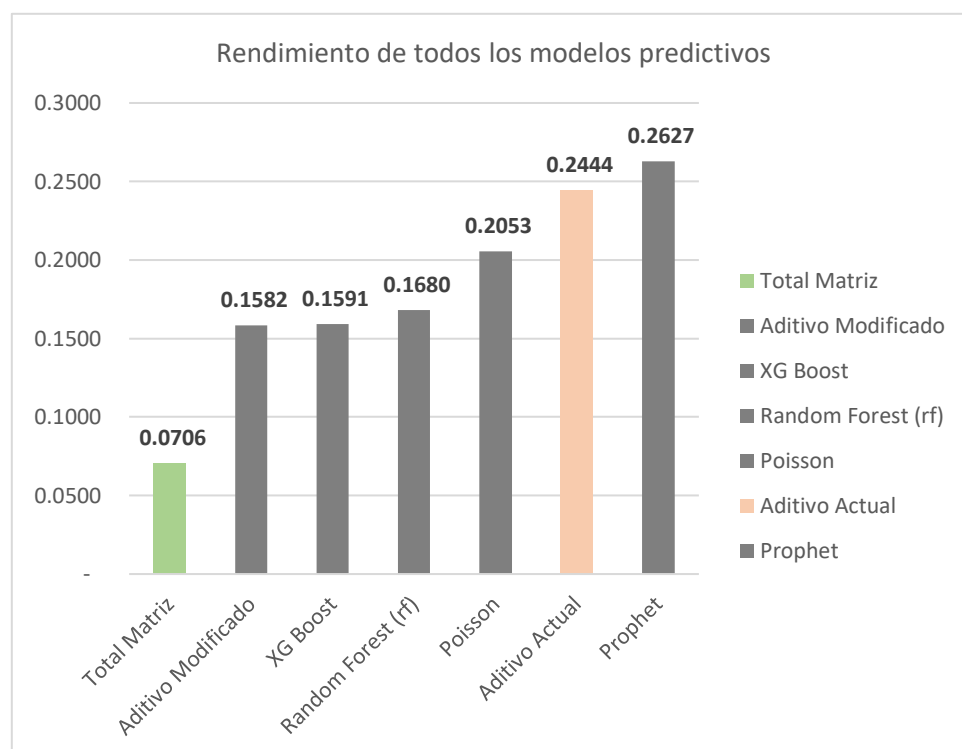
El modelo Aditivo Modificado destaca por tener el MAE promedio más bajo entre los modelos evaluados (0,1582), lo que indica que, en promedio, sus predicciones son las más precisas. Además, con una desviación estándar de 0,3229, una de las más bajas, se muestra que su rendimiento es consistentemente sólido en diversas situaciones.

En resumen, el modelo Aditivo Modificado ofrece un excelente equilibrio entre precisión, consistencia y simplicidad, haciéndolo una opción atractiva para quienes buscan un modelo robusto y fácil de manejar.

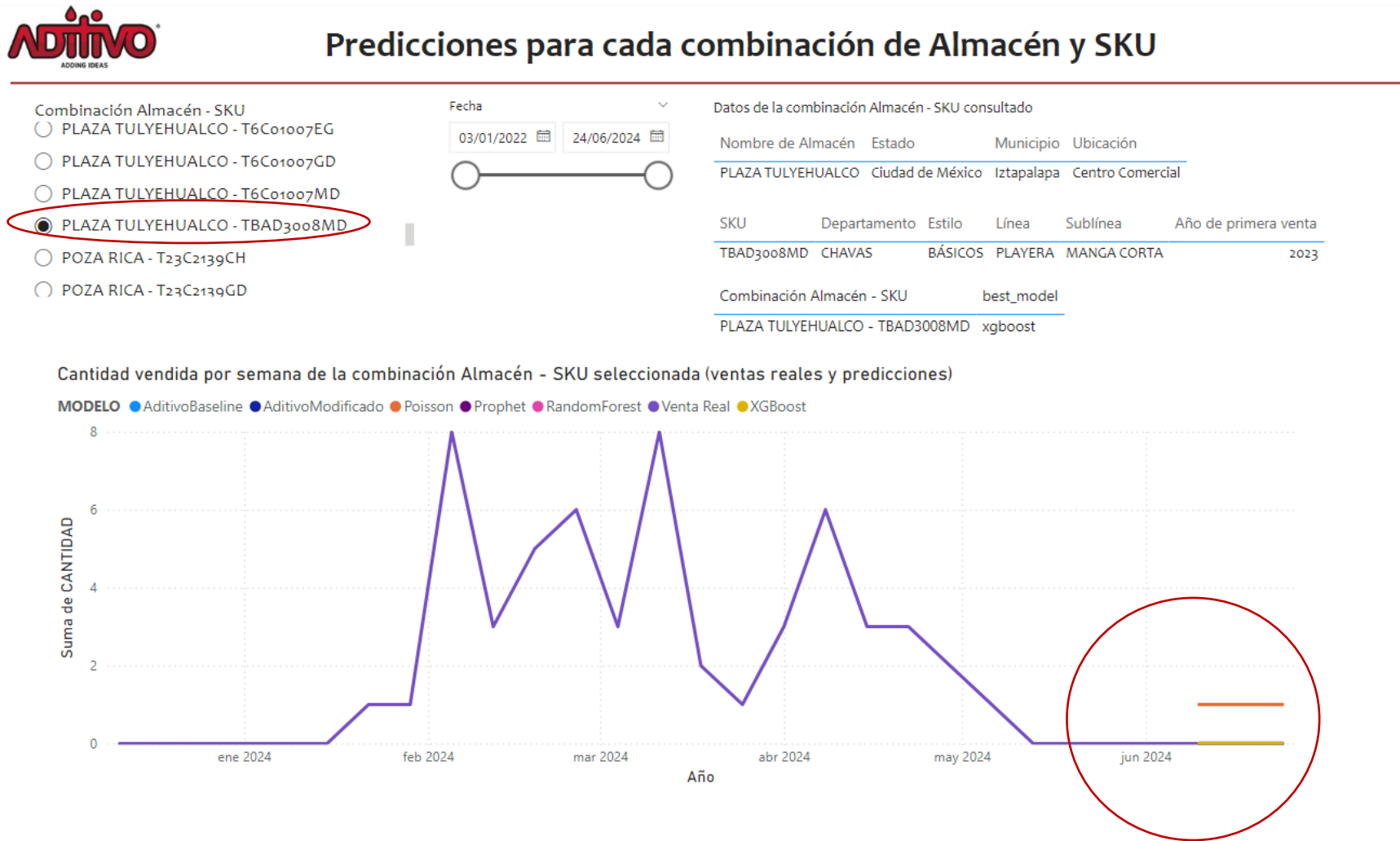
Sin embargo, al analizar los resultados de forma conjunta en lugar de individualmente, es posible crear una matriz combinada (Total Matriz) que utilice el mejor modelo para cada combinación. Esto da como resultado un MAE promedio total de 0,0706, significativamente inferior al de cualquier modelo individual.

Este valor es significativamente inferior al del modelo de predicción de ventas actual de la empresa (0,2444) y también la mitad del MAE del modelo Aditivo Modificado (0,1582).

Con el fin de visualizar los resultados y cuantificar las predicciones en unidades de ventas, se generó un segundo archivo CSV (a través de *Python*) y conectó a *Power BI* con el fin de crear un *dashboard*. En las siguientes páginas se mostrarán diversas vistas del *dashboard* con ejemplos concretos.



En la primera vista, se puede observar a modo de ejemplo que, para el SKU “TBAD3008MD” del almacén “PLAZA TULYEHUALCO” la predicción de *XG Boost* (clasificado como mejor modelo para esta combinación), de línea color amarilla, es de cero ventas para primeras tres semanas de junio 2024, sin embargo, el modelo *Poisson* (línea color naranja) predice que se va a vender 1 camiseta por semana, lo cual generaría seguramente una orden de producción de estas 3 camisetas cuando en realidad no se vendieron camisetas en estas fechas.



En esta segunda vista se puede observar que para todos los SKUs que se venden en el almacén “4 CAMINOS” la mejor alternativa es usar el mejor modelo para cada combinación “Almacén – SKU” (línea verde) que genera la matriz propuesta dado que es la más cercana a la venta real (púrpura).



Predicciones de ventas por semana

Fecha: 01/01/2024 - 24/06/2024

Nombre de almacén: 4 CAMINOS

Sublínea: Todas

Combinación Almacén - SKU:

- ☐ 4 CAMINOS - T23C2139CH
- ☐ 4 CAMINOS - T23C2139EG
- ☐ 4 CAMINOS - T23C2139GD
- ☐ 4 CAMINOS - T23C2139MD
- ☐ 4 CAMINOS - T6C01007CH
- ☐ 4 CAMINOS - T6C01007EG

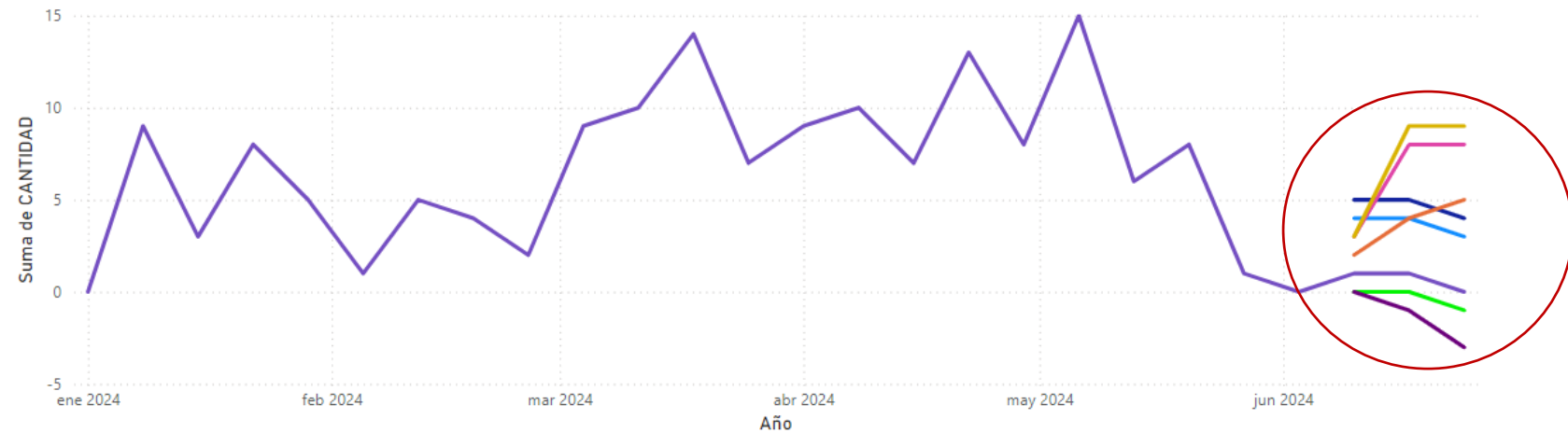
Población: Todas

Estilo de vida: Todas

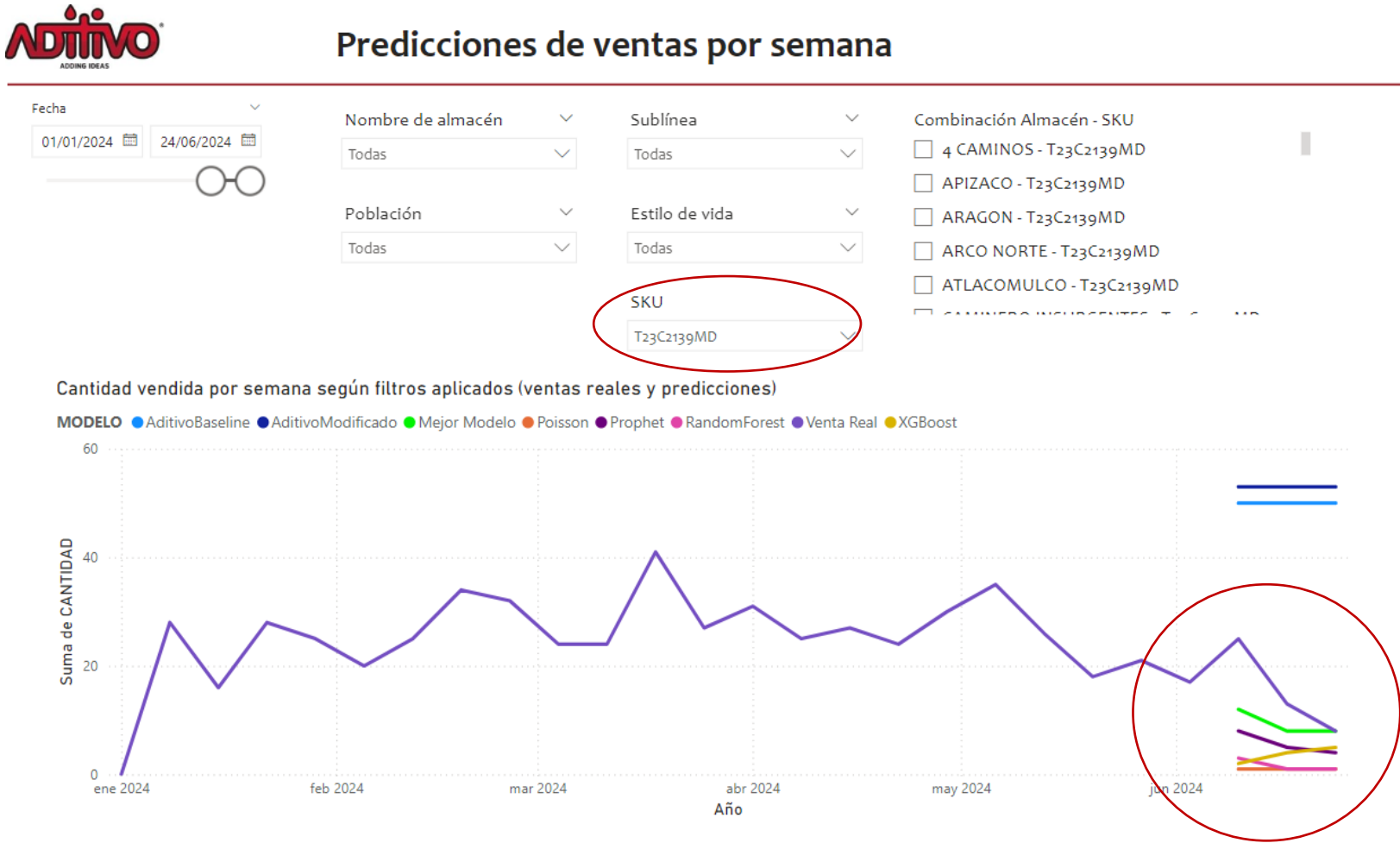
SKU: Todas

Cantidad vendida por semana según filtros aplicados (ventas reales y predicciones)

MODELO: AditivoBaseline AditivoModificado Mejor Modelo Poisson Prophet RandomForest Venta Real XGBoost



En esta tercera vista también se puede observar que para el SKU T23C2139MD que se venden en todos los almacenes, la mejor alternativa es usar el mejor modelo para cada combinación “Almacén – SKU” (línea verde) que genera la matriz propuesta dado que es la más cercana a la venta real (púrpura).



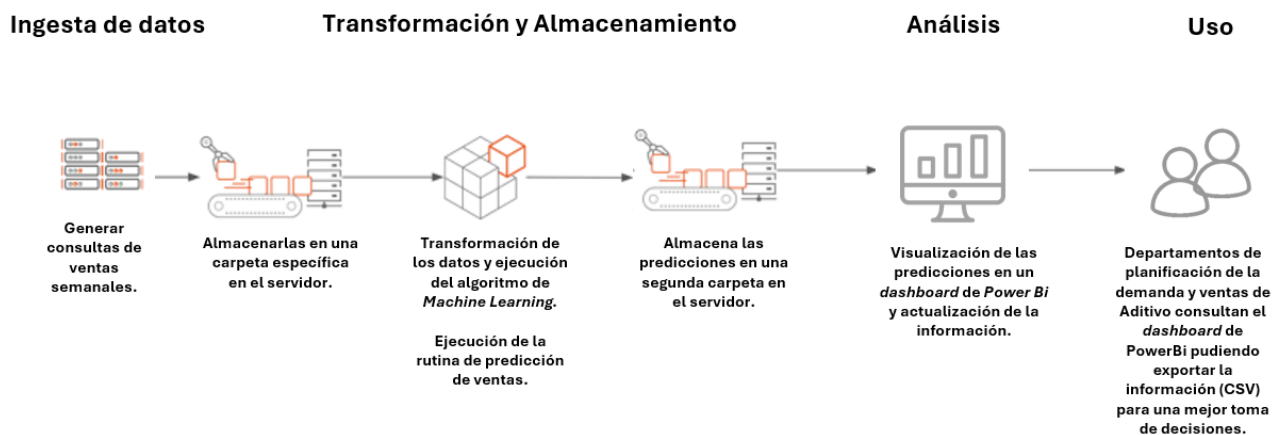
5. Uso del modelo predictivo a futuro

Los principales beneficiarios de esta solución serán los departamentos de planificación de la demanda y ventas de Aditivo. La implementación del modelo consistirá en la exportación de un archivo en formato MS Excel con predicciones para las próximas tres semanas de cada combinación “Almacén-SKU”, facilitando un control más eficiente de los inventarios y la producción.

Además, el dashboard creado permitirá al equipo visualizar las predicciones, compararlas con las ventas reales y analizar las curvas de demanda por SKU, sublínea de producto o segmento de clientes, lo que mejorará la toma de decisiones estratégicas.

Para asegurar la precisión continua del modelo predictivo, será esencial reentrenarlo con datos actualizados. El reentrenamiento constante, liderado por el equipo de TI, garantizará que el modelo se ajuste a las últimas tendencias del mercado, optimizando su precisión de manera sostenida.

A continuación, se presenta el *pipeline* de recolección de datos a proponer a Aditivo:



6. Conclusiones y recomendaciones

Continuar con este proyecto es una decisión acertada, ya que se ha logrado un avance significativo en la precisión de las predicciones, mejorando el MAE en un factor de 3x respecto al modelo actual. No obstante, ampliar el análisis a más SKUs y almacenes permitirá optimizar aún más los resultados obtenidos y fortalecer la capacidad predictiva del modelo.

Actualmente, cada modificación requiere aproximadamente 12 horas de procesamiento, lo que subraya la necesidad de mejorar la infraestructura tecnológica. La adopción de servicios en la nube agilizaría este proceso, reduciendo los tiempos de ejecución y facilitando el desarrollo continuo del proyecto.

Los resultados preliminares son prometedores, con una mejora notable respecto al modelo Aditivo actual. A pesar de esto, aún hay margen para seguir optimizando el modelo y asegurar un impacto sostenible a largo plazo. Por lo tanto, es esencial llevar a cabo una investigación más profunda que evalúe el impacto a largo plazo y el retorno de la inversión (ROI) para justificar el escalamiento del proyecto.

Plan de Acción Propuesto

1. **Optimización de la infraestructura tecnológica:** Evaluar la solución más adecuada, ya sea hardware dedicado o servicios en la nube, para garantizar que el proyecto sea escalable y eficiente. Esta mejora debe enfocarse en minimizar los costes y optimizar los tiempos de procesamiento sin comprometer la seguridad y confiabilidad del sistema.
2. **Ampliación del análisis de SKUs y almacenes:** Expandir progresivamente el número de almacenes y SKUs evaluados hasta abarcar todos los SKUs activos. Este paso permitirá que el modelo sea más representativo y preciso en diferentes entornos y mercados.
3. **Integración con los sistemas de inventario y producción:** Integrar el nuevo modelo predictivo con los sistemas de gestión de inventario y planificación de producción existentes en Aditivo, mejorando la toma de decisiones en tiempo real y reduciendo la posibilidad de desajustes de inventario.
4. **Análisis coste-beneficio:** Realizar un análisis exhaustivo que compare los costes de implementación y mantenimiento del nuevo modelo con los ahorros y beneficios potenciales, tanto en la gestión de inventario como en el incremento de ventas. Este análisis permitirá establecer una base sólida para futuras inversiones en tecnología predictiva.
5. **Sistema de mejora continua:** Implementar un sistema de mejora continua para aumentar la precisión del modelo. Esto debe incluir la experimentación con nuevos algoritmos de machine learning y diferentes combinaciones de hiperparámetros, adaptando el modelo a las dinámicas cambiantes del mercado.
6. **Incorporación de variables externas:** Investigar la inclusión de nuevas variables, como tendencias del mercado, datos económicos, festividades o eventos especiales, que puedan mejorar la precisión del modelo en previsiones más detalladas.

Gestión del Cambio y Capacitación

Para garantizar una implementación exitosa, es fundamental desarrollar un plan de gestión del cambio que facilite la adopción del nuevo sistema en toda la organización. Esto minimizará posibles resistencias por parte de los responsables actuales y asegurará una integración fluida.

Además, será crucial implementar un programa de capacitación para el personal de TI, ventas y planificación. Esto les permitirá interpretar y utilizar eficazmente las predicciones generadas por el modelo. Aunque algunos cambios puedan percibirse como disruptivos inicialmente, son esenciales para asegurar una adopción exitosa y sin contratiempos.

Propuesta de Inversión y Escalamiento

Proponemos un plan de inversión por fases, permitiendo que Aditivo incremente gradualmente su inversión en machine learning conforme se obtengan resultados positivos. Este enfoque reducirá riesgos y permitirá ajustar la estrategia de implementación a medida que se obtienen nuevos aprendizajes y avances.

Finalmente, de manifestar interés, se podría elaborar una propuesta económica detallada y un plan de trabajo estructurado. Este plan incluiría todas las fases necesarias, desde la inversión inicial hasta la capacitación y la integración con los sistemas actuales, asegurando una estrategia clara y eficiente para la implementación de un método más robusto de estimación de la demanda.