



Máster en Data Science e Inteligencia Artificial

TRABAJO DE FIN DE MÁSTER

Role-Play DSMarket

Autores

José Bau

Víctor Barria

Pablo Guimerà

Hector Maninatt

Simone Solieri

Tutora

Raquel Revilla

Director

Isidre Royo

Curso académico 2024

1. Introducción	3
2. Descripción del Problema	4
Impacto del Problema en la Empresa	5
3. Metodología	7
4. Unificación y Limpieza de Datos	9
5. Creación Dashboard y análisis de datos	12
Estructura del Dashboard	12
Sección Global	12
Sección Stores & Regions (Tiendas y Regiones)	13
Sección Marketing	15
Sección Forecasting (Predicción de Ventas)	15
Navegación del Dashboard	16
6. Segmentación de Productos y Tiendas	19
7. Modelo de Predicción de Ventas	24
Sales Forecasting: Proceso y Resultados	24
Predicciones de Ventas por Tienda y Categoría	27
Comparación con el Año Anterior	29
Cálculo del RMSE	30
Comportamiento del Error a lo Largo del Tiempo	30
Razón para el Aumento del Error	31
Ventajas del Uso del RMSE	32
8. Caso de Uso y Piloto	33
9. Discusión	38
10. Conclusiones	42

1. Introducción

- Objetivo del proyecto

- **Análisis y Unificación de Datos:** Unificar diversas tablas de datos (ventas, precios, calendario) en un único dataset cohesivo para facilitar su análisis.
- **Aplicación de Clustering:** Identificar patrones de comportamiento en tiendas y productos a través de técnicas de clustering.
- **Desarrollo de Modelos Predictivos:** Construir y validar modelos predictivos de ventas basados en algoritmos de machine learning.
- **Modelo de implementación:** Desplegar un modelo predictivo que se integre en la operación diaria de DSMarket, optimizando el proceso de reabastecimiento.
- **Evaluación de Impacto:** Medir el impacto del modelo en términos de reducción del stock remanente y mejora en la gestión de inventarios.

- Contexto y relevancia del estudio

En el competitivo entorno del retail, la capacidad para anticipar la demanda y gestionar inventarios de manera eficiente se ha vuelto fundamental para garantizar el éxito de una empresa. DSMarket, una cadena de centros comerciales en Estados Unidos, se encuentra inmersa en un proceso de transformación digital para mejorar su eficiencia operativa y su competitividad en el mercado. Este proyecto de fin de máster se centra en el uso de la analítica avanzada de datos, clustering y modelos predictivos para optimizar el reabastecimiento de tiendas y mejorar la rentabilidad.

2. Descripción del Problema

El principal problema de negocio identificado en DSMarket está relacionado con la ineficiencia en la gestión del inventario y el reabastecimiento de productos, lo que genera importantes consecuencias en términos de costos operativos y satisfacción del cliente.

Actualmente, DSMarket enfrenta dificultades para anticipar con precisión la demanda de productos en sus tiendas, lo que deriva en dos problemas recurrentes:

1. Exceso de Stock:

Al no tener una predicción precisa de la demanda futura, DSMarket tiende a sobre aprovisionarse, lo que genera un exceso de inventario en productos perecederos y no perecederos. Esto no solo ocupa espacio de almacenamiento valioso, sino que también resulta en productos que se quedan en los estantes demasiado tiempo, lo que en el caso de productos perecederos, lleva a pérdidas directas debido a la expiración.

2. Faltante de Stock (Out-of-Stock):

Por otro lado, en los momentos en que la demanda supera las expectativas, DSMarket se enfrenta a la escasez de productos en algunas tiendas. Esto provoca situaciones en las que los clientes no encuentran los productos que buscan, afectando directamente las ventas. La falta de disponibilidad de productos clave, especialmente en los días y eventos importantes, resulta en una pérdida de oportunidades de ventas.

Además, el impacto de eventos estacionales y promociones también es difícil de prever. Durante días clave, como el Superbowl o el Black

Friday, las ventas tienden a experimentar picos que no siempre son manejados de manera eficiente por la empresa. El problema radica en que, sin un modelo predictivo adecuado, las tiendas no están preparadas con el stock adecuado en los momentos precisos, lo que agrava los problemas tanto de exceso como de faltante de inventario.

Causas Principales del Problema:

- Falta de un modelo predictivo robusto que permita anticipar la demanda de manera precisa, tanto para productos de venta diaria como para aquellos afectados por eventos especiales o promociones.
- Ineficiencias en la cadena de suministro, que no se adapta de manera ágil a los cambios en la demanda, especialmente en momentos de picos estacionales.
- Gestión inconsistente de inventarios entre las diferentes tiendas, lo que genera desigualdad en la disponibilidad de productos en las diversas regiones.

Impacto del Problema en la Empresa

El problema de gestión ineficiente de inventarios tiene varios impactos significativos en DSMarket, tanto en términos operativos como financieros:

1. Pérdidas Financieras por Exceso de Stock

- Costos de Almacenamiento: Los productos que no se venden rápidamente ocupan espacio de almacenamiento, lo que genera costos adicionales para la empresa. El almacenamiento prolongado de productos, especialmente en el caso de productos perecederos, puede generar desperdicios, lo que se traduce en pérdidas directas de capital.
- Desperdicio de Productos Perecederos: La acumulación de productos perecederos que no se venden a tiempo se traduce

en una pérdida directa de inventario, ya que los productos deben ser descartados. Este es un problema recurrente en la categoría de Supermercado, que representa el 56% de la facturación de DSMarket.

2. Pérdida de Oportunidades de Ventas

Oportunidades de venta Perdidas: Cuando DSMarket no tiene suficiente stock de productos clave, especialmente durante eventos de alto impacto como el Superbowl, pierde oportunidades significativas de venta. Esto no solo afecta los ingresos inmediatos, sino que también disminuye la confianza del cliente en la empresa, lo que puede llevar a una menor fidelización a largo plazo.

3. Metodología

Gestión ágil del proyecto

Para asegurar una ejecución eficiente y organizada del proyecto, adoptamos la metodología ágil Scrum. Esta metodología nos permitió trabajar de manera iterativa, respondiendo rápidamente a los cambios y mejorando continuamente el desarrollo del proyecto en cada ciclo o sprint. A continuación, se detalla cómo estructuramos y gestionamos el proyecto bajo este enfoque.

1. Sprints Semanales

Cada semana organizamos un sprint de una duración de 7 días, en los cuales establecimos metas específicas y dividimos las tareas a realizar. Al final de cada sprint, revisamos los avances y hacíamos ajustes según fuera necesario. Esto nos permitió avanzar progresivamente hacia los objetivos finales del proyecto, manteniendo la flexibilidad para hacer cambios cuando surgían nuevas ideas o requisitos.

2. Rotación del Scrum Master

Cada sprint fue liderado por un Scrum Master que rotaba semanalmente entre los cinco miembros del equipo: Víctor Barria, José Bau, Simone Solieri, Pablo Guimera y Hector Maninatt. El Scrum Master de la semana tenía las siguientes responsabilidades:

- Cumplimiento de la metodología: Garantizar que el equipo siguiera correctamente las pautas de Scrum, manteniendo una buena comunicación y flujo de trabajo.
- Daily Standups: Organizar reuniones diarias para que cada miembro informará sobre en qué estaba trabajando, qué impedimentos enfrentaba, y los avances logrados hasta ese momento.

- **Sprint Planning:** Dirigir la reunión de planificación del sprint, donde priorizamos las tareas y distribuimos las responsabilidades entre los miembros del equipo.
- **Retrospectiva:** Al final de cada sprint, el Scrum Master lideraba la retrospectiva, una sesión para reflexionar sobre los desafíos y éxitos del sprint, con el fin de mejorar el proceso en el siguiente ciclo.

3. Herramientas y Tablero en Trello

El equipo utilizó Trello como herramienta principal de gestión del proyecto. En el tablero de Trello, creamos columnas para cada fase del sprint:

- **Backlog:** Todas las tareas pendientes o nuevas ideas que surgían se añadían aquí. Todos los miembros del equipo podrían agregar tareas al backlog.
- **En Proceso:** Las tareas que se estaban ejecutando en el sprint en curso se movían a esta columna.
- **Completadas:** Una vez que una tarea se finaliza, se movía a la columna de completadas.

Cada miembro del equipo actualizaba diariamente el estado de sus tareas, facilitando la visibilidad del progreso para todos y manteniendo la transparencia en el trabajo.

4. Flexibilidad y Adaptación

Una de las principales ventajas del enfoque ágil es la flexibilidad para adaptarse a los cambios. A lo largo del proyecto, surgieron nuevos desafíos y descubrimientos que requerían ajustes en la planificación inicial. Gracias a la metodología Scrum, pudimos modificar las prioridades en función de los nuevos insights obtenidos, sin perder el ritmo de trabajo ni afectar los plazos establecidos.

4. Unificación y Limpieza de Datos

Una parte fundamental para el éxito del proyecto fue la unificación y limpieza de los datos provenientes de múltiples fuentes, lo que permitió realizar un análisis cohesivo y construir modelos predictivos precisos. A continuación, se detalla cómo se realizó el proceso de join de las tablas y la preparación de los datos para su análisis.

1. Unificación de las Tablas (Join)

DSMarket manejaba varios conjuntos de datos que estaban dispersos entre diferentes tablas. Estas tablas incluían información sobre días de negocio y eventos, ventas, productos, y datos demográficos de las tiendas. Para poder extraer insights valiosos y crear modelos predictivos, fue necesario unificar todos estos datos en un único dataset.

A. Proceso de Join de las Tablas:

Identificación de las Claves Comunes:

Las tablas contenían datos que compartían claves comunes, como:

ID de producto: Identificador único de cada artículo en el inventario.

ID de tienda: Identificador de cada tienda en diferentes ubicaciones.

Día de negocio: Identificador único del día de negocio.

B. Transformación de los datos y ejecución de los Joins:

Primero, dado que inicialmente la tabla presentaba una columna por cada día de negocio con la cantidad vendida en ese día, se transformó el dataset de ventas en formato “long”. De esta forma se obtuvo una única columna que indique el día de negocio.

En segundo lugar, se generó la columna “yearweek” dentro de la tabla Calendario, para poderla utilizar como clave para el join con la

tabla de Precios. Este paso no fue trivial, dado que la columna parecía no seguir ninguna estandarización, como por ejemplo la ISO, y por eso fue necesario un profundo estudio de la lógica de esta variable para que los datos resultantes fueran coherentes.

Con las claves mencionadas anteriormente finalmente se realizó el join, obteniendo un dataset consolidado que contiene toda la información completa.

C. Limpieza de Datos

Una vez unificadas las tablas, el siguiente paso fue limpiar el dataset consolidado para asegurar la calidad de los datos y su integridad. Este proceso fue crucial para evitar errores en el análisis y garantizar que los modelos predictivos fueran confiables y precisos.

Proceso de Limpieza de Datos:

Control de Duplicados:

No se identificaron en ningún momento valores duplicados.

Manejo de Valores Faltantes:

Una vez obtenido el dataset unificado, los únicos valores faltantes pertenecían a los campos de Eventos y Precio de venta. Para abordar este problema, utilizamos diferentes estrategias según el caso:

Eventos: se imputó a los nulos la etiqueta “No Event”, dado que razonablemente la falta de valor en ese campo puede indicar la ausencia de eventos.

Precio de venta: Sabiendo que los nulos en Sell price corresponden a semanas sin ventas, se imputaron los valores faltantes con el último precio disponible de las semanas anteriores, o en su defecto, el primer precio disponible en las semanas sucesivas.

D. Tratamiento de Valores Atípicos (Outliers):

Identificamos y analizamos los outliers en variables clave como ventas y precio de venta. Aunque algunos de estos valores eran válidos (por ejemplo, picos de ventas durante eventos como el Superbowl), otros representaban errores o inconsistencias.

Para los outliers no justificados, reemplazamos los valores con datos coherentes, normalmente la media.

E. Enriquecimiento de Datos:

Además de la limpieza, también enriquecemos los datos con factores externos que podrían impactar en las ventas. En particular añadimos ulteriores eventos a la columna Event, agregando nueva información que mejoraron la precisión de los modelos predictivos. .

2. Resultados de la Unificación y Limpieza de Datos

El proceso de unificación y limpieza nos permitió tener un dataset sólido, coherente y de alta calidad, listo para ser utilizado en el análisis y desarrollo de modelos predictivos. Al contar con un dataset bien estructurado:

- Fuimos capaces de realizar un análisis profundo de los patrones de ventas y demanda.
- Garantizamos que los modelos de clustering y forecasting se basaran en datos precisos, lo que resultó en predicciones confiables.

5. Creación Dashboard y análisis de datos

El Dashboard desarrollado para DSMarket es una herramienta avanzada de análisis de datos que permite a los equipos de operaciones, marketing y finanzas tomar decisiones fundamentadas en datos en tiempo real. Este manual de uso tiene como objetivo proporcionar una guía detallada para los usuarios, asegurando una correcta interpretación y utilización de las funcionalidades clave, con el fin de maximizar los beneficios derivados de los insights generados por el sistema.

Estructura del Dashboard

El Dashboard está estructurado en cuatro secciones principales, cada una de ellas ofrece una visión exhaustiva de distintos aspectos del negocio, lo que permite un análisis detallado a múltiples niveles:

- **Global**
- **Stores & Regions (Tiendas y Regiones)**
- **Products (Productos)**
- **Marketing**
- **Forecasting (Predicción de Ventas)**

Sección Global

La Sección Global proporciona una visión integral de las métricas clave del negocio a nivel macro. Esta sección permite a los usuarios realizar un análisis de alto nivel del rendimiento general de la empresa.

- **Ingresos Totales:** Permite visualizar los ingresos generados en un periodo específico, brindando una visión clara de la evolución de las ventas.
- **Unidades Vendidas:** Muestra el total de unidades vendidas hasta la fecha o durante un periodo determinado.

- **Beneficios:** Calcula los beneficios obtenidos por cada venta realizada, disponible tanto en términos generales como para un rango de fechas seleccionado.
- **Crecimiento de Ventas:** Ofrece comparativas gráficas del crecimiento de las ventas año tras año, con análisis tanto a nivel global como a nivel de tienda o región.
- **Crecimiento Anual Compuesto (CAGR):** Evalúa el crecimiento compuesto anual, una métrica esencial para entender la expansión sostenida del negocio.
- **Aplicación Práctica:** Esta sección es crucial para la alta dirección y los responsables de estrategia, ya que les permite medir rápidamente si las ventas están alineadas con los objetivos estratégicos. Además, se incluyen funcionalidades como consultas ad-hoc a través de la sección de Q&A, donde una inteligencia artificial responde instantáneamente a preguntas clave.

Sección Stores & Regions (Tiendas y Regiones)

La sección Stores & Regions proporciona una segmentación granular del desempeño de cada tienda individual y las regiones donde DSMarket tiene presencia, permitiendo un análisis profundo tanto del rendimiento comercial como del comportamiento de los consumidores en diferentes ubicaciones.

- **Ventas por Tienda:** Muestra una comparación entre el rendimiento de las distintas tiendas, ayudando a identificar las más exitosas.
- **Crecimiento por Tienda:** Proporciona una visión del crecimiento interanual de cada tienda, permitiendo distinguir entre tiendas consolidadas y emergentes.
- **Análisis de Demanda:** Muestra los productos más y menos demandados en cada tienda, optimizando así las estrategias de inventario.

- **Store Cluster:** Filtra las búsquedas mediante la segmentación en clústeres, permitiendo la comparación entre grupos homogéneos de tiendas.
- **Comparación de Regiones:** Ofrece una comparativa visual entre el desempeño de ventas y crecimiento en las principales regiones (por ejemplo, Nueva York, Filadelfia, y Boston).
- **Estadísticas Demográficas:** Proporciona datos demográficos relevantes, fundamentales para ajustar las ofertas de productos y campañas de marketing.

Aplicación Práctica: Esta sección resulta fundamental para los gerentes de tienda y los equipos de logística. Permite ajustar estrategias de inventario, reabastecimiento y expansión en función de la demanda local y las tendencias regionales.

Sección Products (Productos)

La Sección Productos ofrece un análisis pormenorizado del rendimiento de los productos del portafolio de DSMarket, desglosando cómo factores como las regiones, categorías o eventos influyen en sus ventas.

- **Ventas por Producto:** Desglosa las ventas por producto, identificando tanto los productos más vendidos como aquellos con menor rendimiento.
- **Historia del Producto:** Realiza un análisis histórico del desempeño de cada producto, identificando cambios y tendencias en el consumo.
- **Tendencias de Consumo:** Muestra patrones de consumo que pueden ser aprovechados para ajustar las estrategias de inventario y promociones.

Aplicación Práctica: Los equipos de producto y compras pueden usar esta sección para optimizar el portafolio, identificando oportunidades de mejora y ajustando la oferta de productos según

las tendencias actuales. Este análisis permite también una planificación estratégica más efectiva sobre las promociones y el reabastecimiento.

Sección Marketing

La Sección Marketing está diseñada para evaluar el impacto de las campañas promocionales y para facilitar la planificación de futuras estrategias de marketing.

- **Días Estratégicos:** Muestra los días clave para lanzar campañas de marketing, basados en el análisis histórico de ventas (ej. primeros días del mes coincidiendo con la recepción de beneficios sociales).
- **Planificación de Campañas:** Permite planificar campañas basadas en el comportamiento de los consumidores y los datos de ventas.
- **Impacto de las Campañas:** Analiza el impacto de campañas pasadas sobre las ventas, ajustando futuras estrategias para maximizar su efectividad.

Aplicación práctica: Los equipos de marketing pueden aprovechar esta sección para maximizar el impacto de las campañas publicitarias, focalizando sus esfuerzos en los momentos y productos que generan mayores retornos, optimizando así las inversiones promocionales.

Sección Forecasting (Predicción de Ventas)

La Sección Forecasting proporciona una visión predictiva de las ventas futuras, permitiendo a los directivos tomar decisiones estratégicas basadas en estimaciones precisas.

- **Predicción:** Compara las predicciones de ventas con los datos del año anterior, permitiendo un análisis comparativo preciso.

- **Planificación de Stock:** Ofrece herramientas predictivas para la planificación de inventario, optimizando los pedidos a proveedores.
- **Impacto de las Categorías:** Identifica qué categorías de productos presentan mayor demanda esperada, facilitando la planificación estratégica.

Aplicación Práctica: Esta sección resulta vital para la planificación operativa, ya que permite a los distintos departamentos anticipar la demanda y preparar estrategias proactivas de ventas, inventario y distribución. Las predicciones, respaldadas por modelos predictivos avanzados, aseguran que las decisiones se basen en datos sólidos y tendencias emergentes del mercado.

Este Dashboard es una herramienta indispensable que centraliza la información crítica del negocio, facilitando el acceso a datos clave y permitiendo una toma de decisiones más precisa y fundamentada en toda la organización.

Navegación del Dashboard

El tablero de control de DSMarket ha sido diseñado para ofrecer una navegación sencilla e intuitiva, adaptándose tanto a usuarios con experiencia en análisis de datos como a aquellos sin formación en ciencia de datos. Esto lo convierte en una herramienta accesible y poderosa para la toma de decisiones basadas en datos en diversas áreas del negocio.

Filtros y Personalización

El dashboard integra una amplia gama de opciones de filtrado que permiten personalizar la vista de los datos según los intereses y necesidades del usuario. Los filtros incluyen:

- **Fechas:** Para ajustar el análisis a periodos de tiempo específicos, facilitando la comparación de datos históricos con el desempeño actual.

- **Regiones/Tiendas:** Permite comparar el rendimiento entre distintas localizaciones y filtrar por tiendas individuales.
- **Productos/Categorías:** Posibilita un análisis focalizado en productos o categorías concretas, permitiendo detectar tendencias específicas.

Esta capacidad de personalización mediante filtros no solo mejora la experiencia del usuario, sino que también optimiza el proceso de análisis al permitir profundizar en áreas específicas del negocio de manera dinámica y en tiempo real.

Exportación de Datos

El dashboard también permite la exportación de gráficos y tablas en formatos como Excel o PDF. Esta funcionalidad es especialmente útil para la preparación de reportes personalizados que pueden ser utilizados en presentaciones ejecutivas, reuniones internas o estrategias comerciales, facilitando la comunicación de los resultados y hallazgos.

Navegación General

El tablero está estructurado en secciones que facilitan la navegación entre distintos niveles de análisis:

Análisis Global del Mercado: Proporciona una visión general del desempeño del negocio a nivel macro, incluyendo métricas clave como ventas totales, crecimiento y factores que influyen en el desempeño.

Análisis por Regiones y Tiendas: Ofrece un desglose del comportamiento de ventas en distintas ubicaciones geográficas, permitiendo identificar tendencias y oportunidades de expansión.

Análisis por Departamentos y Productos: Facilita un análisis detallado de las categorías de productos, ayudando a identificar los principales impulsores de ventas y áreas que requieren ajuste o mejora.

Buenas Prácticas de Uso

Para sacar el máximo provecho del dashboard, se recomienda seguir las siguientes buenas prácticas:

- **Usar filtros de manera estratégica:** Comenzar con una visión global y luego profundizar mediante filtros, como fechas, tiendas o productos, para obtener una mejor comprensión de áreas específicas.
- **Realizar análisis comparativos:** Aprovechar las funcionalidades de comparación entre periodos, regiones o productos para detectar tendencias clave y puntos de inflexión en el desempeño del negocio.
- **Exportar datos para informes personalizados:** La exportación de datos es útil para documentar los hallazgos y compartirlos con los distintos equipos de trabajo, facilitando la toma de decisiones basada en datos.

Ventajas del Dashboard para Usuarios no Especializados

Una de las mayores fortalezas del dashboard es su accesibilidad para usuarios sin experiencia en análisis de datos. A través de una interfaz amigable y una estructura lógica, los usuarios pueden obtener insights valiosos sin necesidad de interpretar complejas fórmulas o construir modelos estadísticos. El diseño interactivo permite a estos usuarios explorar las métricas de manera intuitiva, ajustando los filtros y visualizaciones según sus necesidades sin requerir conocimientos técnicos.

Además, la posibilidad de hacer consultas mediante la herramienta de **Q&A** (preguntas y respuestas) integrada en el dashboard, facilita el acceso a información clave de manera instantánea, convirtiéndolo en una herramienta indispensable para la toma de decisiones informadas en todas las áreas del negocio.

6. Segmentación de Productos y Tiendas

Estructura de los Segmentos de Productos

El análisis detallado de los datos de ventas y rendimiento de DSMarket ha permitido identificar patrones clave dentro de los productos y las tiendas. A través de la segmentación, hemos agrupado tanto los productos como las tiendas en categorías estratégicas que proporcionan insights valiosos para la toma de decisiones. A continuación, se presentan los detalles de los

Segmento	Cantidad de Productos	Ventas Diarias	Ingresos Diarios	Cambios de Precio	Variación en Ventas Durante Eventos	Precio Promedio	Variabilidad en Ventas
Top Sellers	504 (16%)	35	\$107	15	+0.6%	5	1253
Low Price	962 (32%)	8	\$18	12	-1.7%	3	76
High Price	583 (19%)	4	\$36	10	-0.8%	10	21
Event Loss	688 (23%)	7	\$29	17	-17.6%	5	72
Event Boosted	310 (10%)	5	\$31	14	+21.4%	7	74

segmentos identificados para productos y tiendas. El algoritmo de clusterización utilizado en ambos casos fue K-Means.

La segmentación de los productos de DSMarket está basada en variables como ventas diarias, ingresos, variación en ventas durante eventos, y cambios de precios. Esta clasificación ha revelado cinco segmentos clave que se describen en la siguiente tabla:

Descripción de los Segmentos:

1. Top Sellers:

- **Cantidad de Productos:** 504 (16%)
- **Descripción:** Estos productos representan la mayor fuente de ingresos, con ventas diarias elevadas y una alta variabilidad en las ventas. Son fundamentales para mantener un flujo constante de ingresos.
- **Estrategia:** Mantener inventario alto y asegurar su disponibilidad continua. Estos productos son ideales para promociones cruzadas y programas de fidelización.

2. Low Price:

- **Cantidad de Productos:** 962 (32%)
- **Descripción:** Productos de bajo precio y constante, con ventas diarias bajas. Aunque representan una gran parte del portafolio, generan ingresos bajos.
- **Estrategia:** Revisar la rentabilidad de estos productos y considerar promociones agresivas o eliminar productos con poco impacto en las ventas.

3. High Price:

- **Cantidad de Productos:** 583 (19%)
- **Descripción:** Productos de alto precio con ventas estables y clientes leales. Estos productos generan ingresos altos y bastante estables en el tiempo.
- **Estrategia:** Continuar con precios premium y dirigir campañas de marketing hacia consumidores de alto valor.

4. **Event Loss Products:**

- **Cantidad de Productos:** 688 (23%)
- **Descripción:** Estos productos sufren una caída significativa en ventas durante eventos clave, lo que afecta negativamente los ingresos. Son también los productos que más frecuentemente cambian de precio.
- **Estrategia:** Desarrollar promociones especiales o estrategias de mitigación para contrarrestar esta caída durante eventos importantes.

5. **Event-Boosted Products:**

- **Cantidad de Productos:** 310 (10%)
- **Descripción:** Estos productos experimentan un aumento en ventas durante eventos especiales, lo que los convierte en productos clave durante periodos de alta demanda. En su mayoría son productos de precio medio-alto.
- **Estrategia:** Aumentar inventario antes de los eventos clave y lanzar campañas de marketing específicas para maximizar su impacto.

Estructura de los Segmentos de Tiendas

La segmentación de las tiendas de DSMarket ha permitido agruparlas en tres categorías: **Tiendas Consolidadas**, **Tiendas Emergentes**, y **Tiendas Rezagadas**. Estos segmentos se obtuvieron utilizando métricas como crecimiento interanual, cuota de mercado regional y global, los detalles se describen en la siguiente tabla.

Segmento	Cantidad de Tiendas	Cuota de Mercado Regional	Cuota de Mercado Global	Crecimiento Interanual	Localización
Consolidadas	6 (60%)	33.4%	11.1%	1.3%	Philadelphia (Queen Village), New York (Greenwich Village, Tribeca), Boston (South End, Roxbury, Back Bay)
Emergentes	2 (20%)	33.1%	9.2%	13.7%	Philadelphia (Midtown Village, Yorktown)
Rezagadas	2 (20%)	17%	7.3%	4.7%	New York (Harlem, Brooklyn)

Descripción de los Segmentos:

- 1. **Tiendas Consolidadas:**
 - **Cantidad de Tiendas:** 6 (60%)
 - **Descripción:** Son las tiendas más establecidas, representando la mayor parte de la cuota de mercado

regional y global, aunque su crecimiento es muy modesto.

- **Estrategia:** Invertir en estrategias de revitalización para evitar la pérdida de cuota de mercado. Son tiendas críticas que requieren ajustes en promociones y manejo de inventario.

2. Tiendas Emergentes:

- **Cantidad de Tiendas:** 2 (20%)
- **Descripción:** Estas tiendas están experimentando un crecimiento interanual acelerado y tienen un gran potencial de expansión. Ubicadas en Filadelfia, están preparadas para capturar una mayor cuota de mercado.
- **Estrategia:** Focalizarse en la expansión y el soporte de estas tiendas para consolidar su crecimiento y aprovechar nuevas oportunidades.

3. Tiendas Rezagadas:

- **Cantidad de Tiendas:** 2 (20%)
- **Descripción:** Estas tiendas, ubicadas en Nueva York, están enfrentando un crecimiento más lento y menor participación en el mercado. Necesitan una reestructuración para mejorar su rendimiento.
- **Estrategia:** Realizar intervenciones estratégicas para revitalizar estas tiendas, como promociones agresivas y ajustes en la gestión de inventario

7. Modelo de Predicción de Ventas

DS Market quiere revolucionar su método de forecasting, que hasta este momento se ha basado en un enfoque muy rudimental. Para testear el modelo, el directivo quiere que realicemos una predicción a 28 días. El uso principal del modelo de predicción será para organizar el abastecimiento de cada tienda, que normalmente se hace con frecuencia semanal. Por eso, para el desarrollo del modelo utilizamos un dataset agrupado por semana.

- Explicación del modelo predictivo utilizado

Después de limpiar el dataset de los artículos que ya no se venden desde hace al menos un año, tenemos un total de 30454 series temporales para predecir, correspondientes a las ventas de cada ítem en su propia tienda. Nos encontramos entonces delante de un problema multivariante.

Se decidió utilizar el modelo XGBoost Regressor. Al tener que predecir 4 semanas, si la previsión de la semana siguiente se hubiera basado en la previsión de la semana anterior, se habría producido una acumulación progresiva del error, dando lugar a resultados inverosímiles. Para evitar este problema decidimos entrenar 4 modelos diferentes, uno para cada semana que necesitamos predecir. De esta forma cada modelo se basa únicamente en los datos reales disponibles.

Sales Forecasting: Proceso y Resultados

El objetivo de esta fase fue desarrollar un modelo de predicción de ventas que permitiera anticipar la demanda de productos con una ventana temporal de hasta cuatro semanas en el futuro. El proceso incluyó la preparación de características temporales, la creación de series temporales a partir de las ventas históricas, y la implementación de modelos predictivos utilizando técnicas de machine learning avanzadas.

1. Preparación de los Datos para Forecasting

Para realizar la predicción de ventas, se creó un conjunto de características que capturara la tendencia histórica de las ventas. Se utilizaron tanto **lags** (valores previos de las ventas) como **medias móviles** (moving averages) y **varianzas móviles** (moving variances), lo que permitió capturar patrones cíclicos y estacionales en las ventas.

- **Media móvil de ventas:** Se calcularon medias móviles de 2, 4, 8, 12 y 24 semanas para observar la evolución de las ventas en diferentes ventanas temporales.
- **Varianza móvil de ventas:** Se calculó la varianza móvil de las ventas para capturar la variabilidad en las ventas de productos, lo que es especialmente útil para productos cuya demanda es volátil.
- **Lags de ventas:** Se crearon columnas adicionales para capturar las ventas en semanas anteriores (lags) desde la semana 1 hasta la semana 7.

Este enfoque permitió que el modelo tuviera acceso a información clave sobre el comportamiento pasado del producto, lo que le permitió hacer predicciones más precisas.

2. Creación de 4 Modelos de Predicción

Se generaron **cuatro modelos diferentes**, cada uno diseñado para predecir las ventas de una semana específica (Semana 1, Semana 2, Semana 3 y Semana 4 en el futuro). Cada modelo fue entrenado utilizando la misma metodología, pero con ventanas de predicción ajustadas:

1. **Semana 1:** Predicción para la semana inmediata posterior.
2. **Semana 2:** Predicción para la segunda semana en el futuro.
3. **Semana 3:** Predicción para la tercera semana en el futuro.
4. **Semana 4:** Predicción para la cuarta semana en el futuro.

3. Entrenamiento de los Modelos

Para entrenar los modelos, se utilizaron técnicas avanzadas de machine learning, específicamente el algoritmo **XGBoost Regressor**. Este algoritmo es robusto y capaz de manejar grandes volúmenes de datos, lo que lo hace ideal para este tipo de predicciones.

El proceso de entrenamiento incluyó los siguientes pasos:

- **División del dataset:** Se dividieron los datos en tres subconjuntos: **entrenamiento**, **validación** y **test**. Esto permitió entrenar el modelo, ajustarlo utilizando los datos de validación y finalmente evaluarlo con un conjunto de test.
- **Evaluación mediante RMSE:** Para cada modelo, se utilizó la métrica **Root Mean Squared Error (RMSE)** para medir la diferencia entre las ventas reales y las predicciones.

4. Importancia de las Características

Uno de los aspectos clave en el desarrollo de estos modelos fue analizar qué características resultaban más importantes para predecir las ventas. Se generaron gráficos de **importancia de las características** que muestran qué variables influyen más en las predicciones de cada modelo semanal.

- **Semana 1:** La característica más importante fue el **lag de ventas de la semana anterior** (`units_sold_lag1`), seguido por medias móviles de ventas de las últimas 2 y 4 semanas.
- **Semana 2, 3 y 4:** A medida que la ventana de predicción aumentaba, la importancia de las medias móviles de ventas a largo plazo (8 y 12 semanas) se hacía más evidente, indicando que las tendencias más prolongadas tenían mayor impacto para predicciones de largo plazo.

5. Evaluación del Rendimiento de los Modelos

Después de entrenar los modelos, se evaluó su rendimiento en el conjunto de test:

- **RMSE promedio:** Se calculó el error cuadrático medio para cada semana de predicción, con un promedio final de RMSE que osciló entre valores aceptables, indicando una buena capacidad predictiva del modelo.

Los resultados muestran que, aunque los modelos pueden predecir con precisión las ventas en la primera semana (con un RMSE bajo), la precisión disminuye ligeramente en las semanas posteriores, lo que es un comportamiento esperado debido a la naturaleza incierta de las predicciones a largo plazo.

6. Proceso de Predicción

Finalmente, se generaron las predicciones para las cuatro semanas siguientes. Los datos resultantes contenían las predicciones de ventas para cada producto y tienda, organizadas por las semanas futuras 1 a 4, listos para ser utilizados en la planificación de inventarios y reabastecimientos. Estas predicciones fueron generadas utilizando un modelo basado en series temporales con enfoques de Machine Learning y análisis predictivo. El modelo fue ajustado para capturar tanto patrones estacionales como eventos específicos, lo que permitió realizar estimaciones con mayor precisión.

Predicciones de Ventas por Tienda y Categoría

Desempeño por Tienda

El análisis de las predicciones de ventas por tienda mostró resultados variados, dependiendo tanto de las características de cada tienda como de la región en la que se encuentra. Se destacan los siguientes puntos clave:

1. **Tienda de Nueva York (Brooklyn):**

- **Semana 1:** Se predijeron **250,000 unidades vendidas** para la tienda de Brooklyn, lo que representa un incremento del **5.7%** en comparación con la misma semana del año anterior.
- **Semana 4:** Las predicciones para la cuarta semana mostraron una tendencia ligeramente a la baja con **230,000 unidades vendidas**, lo que supone un crecimiento interanual más moderado del **3.1%**.

2. **Tienda de Filadelfia (Yorktown):**

- **Semana 1:** Se estimaron **180,000 unidades vendidas**, un aumento del **8.2%** con respecto al año anterior.
- **Semana 4:** Las ventas proyectadas fueron de **165,000 unidades**, manteniendo un crecimiento positivo del **6.0%**, aunque reflejando una ligera desaceleración respecto a las primeras semanas.

3. **Tienda de Boston:**

- **Semana 1:** Las predicciones para Boston indicaron **130,000 unidades vendidas**, un **4.5%** por encima del mismo periodo del año pasado.
- **Semana 4:** Para la última semana, se prevé una venta de **120,000 unidades**, con un crecimiento interanual de **2.8%**.

Desempeño por Categoría

En cuanto a las **categorías de productos**, los resultados de predicción mostraron las siguientes tendencias destacadas:

1. **Supermercado (Supermarket):**

- Las predicciones sugieren que se venderán **760,000 unidades**, lo que representa un incremento del **12%** en comparación con la misma semana del año pasado.

2. Home & Garden:

- Para esta categoría, se predijeron **277,000 unidades vendidas**, un aumento del **9%** en comparación con el año anterior.

3. Accessories:

- Las ventas estimadas fueron de **111,000 unidades**, mostrando un incremento del **4.8%** en relación con el mismo periodo del año anterior.

Comparación con el Año Anterior

En términos generales, las predicciones reflejan una **tendencia positiva** en la mayoría de las tiendas y categorías, con crecimientos interanuales que oscilan entre el **3% y el 8%**. Las tiendas en regiones más consolidadas, como Nueva York y Filadelfia, mostraron los mayores incrementos, mientras que las tiendas en Boston, aunque todavía crecientes, reflejan una ligera desaceleración, especialmente en las semanas más alejadas en el tiempo.

Por categorías, el **Supermercado** sigue siendo el motor principal de ventas, con un crecimiento sostenido a lo largo de las semanas. En cambio, las categorías de **Home & Garden** y **Accessories** presentan mayores fluctuaciones, mostrando una mayor sensibilidad a las condiciones externas y los eventos específicos.

7. Evaluación de los Resultados

El rendimiento del modelo se evaluó mediante el uso del **Root Mean Square Error (RMSE)**, una métrica ampliamente utilizada en

modelos de predicción de series temporales debido a su capacidad para medir la magnitud de los errores de predicción sin eliminar la dirección del error (sobreestimación o subestimación). El RMSE es particularmente útil en este contexto, ya que penaliza los errores grandes más que los pequeños, lo que es crucial en la planificación de inventarios donde las desviaciones significativas pueden resultar costosas.

Cálculo del RMSE

El **RMSE** se calculó con la siguiente fórmula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - x_i)^2}{n}}$$

Donde:

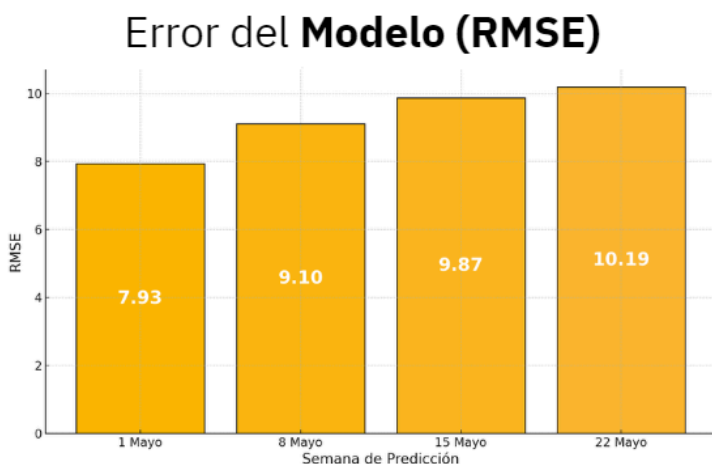
- Y_i representa las ventas reales.
- X_i representa las ventas predichas.
- n es el número total de predicciones realizadas.

Este método fue elegido porque nos permite obtener una medida precisa de la magnitud del error en las predicciones, y al ser una métrica basada en la raíz cuadrada, es más intuitiva en términos de unidades de predicción (en este caso, unidades de ventas). Además, es una métrica sensible a los errores grandes, lo que nos ayuda a identificar cuando el modelo falla significativamente en prever la demanda.

Comportamiento del Error a lo Largo del Tiempo

El análisis de **RMSE** a lo largo de las cuatro semanas mostró una tendencia clara: el error aumenta a medida que nos alejamos en el tiempo como se observa en el gráfico. Este fenómeno es típico en modelos de predicción de series temporales debido a la creciente incertidumbre sobre los factores que influyen en las ventas futuras.

- **Semana 1:** El **RMSE** para las predicciones de la primera semana fue de **7,93**, indicando un margen de error relativamente bajo, lo cual es esperable dado que las condiciones recientes aún son predominantes en la predicción.
- **Semana 2:** El error aumentó a **9,10**, reflejando una mayor incertidumbre en las condiciones externas y fluctuaciones de demanda.
- **Semana 3:** El **RMSE** fue de **9,86**, lo que refleja cómo las inexactitudes en las condiciones de mercado se amplifican con el tiempo.
- **Semana 4:** El error alcanzó su punto máximo en **10,19**, lo que coincide con un descenso en la precisión de las predicciones a medida que se proyectan más semanas en el futuro.



Razón para el Aumento del Error

El incremento del error con el tiempo se debe principalmente a dos factores:

1. **Falta de datos recientes:** la experiencia demuestra que en la previsión de ventas, las tendencias más recientes son más indicativas que las tendencias a largo plazo. Cuanto más lejos se intenta predecir el futuro, menos datos recientes reales están disponibles. En otras palabras, el forecasting a largo plazo es menos preciso porque necesariamente debe basarse en tendencias de largo plazo observadas en el pasado, y no puede basarse en tendencias recientes, ya que estas últimas son sólo predicciones.
2. **Incertidumbre en los factores externos:** Las condiciones de mercado, eventos promocionales y otras influencias externas que afectan las ventas son difíciles de prever con exactitud en plazos largos, lo que introduce una mayor variabilidad en las predicciones.

Ventajas del Uso del RMSE

El uso del **RMSE** es especialmente relevante en este caso porque nos proporciona una visión clara de la dispersión de los errores y su impacto en las decisiones estratégicas. Al utilizar esta métrica, podemos ajustar los inventarios y las estrategias de marketing teniendo en cuenta no solo las predicciones, sino también la **incertidumbre** inherente a ellas. Este enfoque permite una planificación más robusta, donde los márgenes de error son contemplados al optimizar la gestión de inventarios y campañas promocionales.

En resumen, las predicciones de ventas generadas para DSMarket permiten una planificación detallada del inventario y un ajuste

estratégico en las campañas de marketing para las semanas futuras. Si bien los resultados muestran una tendencia positiva en el crecimiento de las ventas en comparación con el año anterior, también se observa un aumento en la incertidumbre conforme nos alejamos en el tiempo, reflejado en el incremento del **RMSE**. El uso de esta métrica nos ha permitido medir y mitigar el riesgo asociado a las predicciones, contribuyendo a la toma de decisiones informadas y al manejo eficiente de los recursos de la empresa.

8. Caso de Uso y Piloto

- Implementación del caso de uso en términos de stock no vendido y oportunidades perdidas

Una vez se tuvo el modelo confeccionado, era hora de ponerlo a prueba. Para ello se decidió aplicarlo a una parte del negocio que no tuviese demasiado impacto sobre el negocio completo, para que en caso de que no fuese del todo bien, no afectase demasiado a los ingresos de la empresa.

Se obtuvieron 3 clusters de tiendas (Emergentes, Consolidadas y Rezagadas). El cluster que menos riesgo y menos volumen del negocio abarcaba era el cluster de tiendas rezagadas. Es por ello que se decidió implementar el modelo piloto en las tiendas rezagadas (piloto conservador), ya que las emergentes y consolidadas asumen más riesgo y volumen del negocio. En caso de que funcionase bien, se implementaría en las consolidadas (piloto neutral), y finalmente en las emergentes (piloto arriesgado).

Se implementó el modelo predictivo sobre las tiendas rezagadas y sus respectivos productos. Los resultados arrojaron cuánto stock sobraría (pérdidas) si utilizáramos nuestro modelo o cuánto habría faltado que se hubiese vendido al tenerlo (pérdida de oportunidad) para los últimos 4 meses.

Pero el análisis no se detuvo ahí. Se decidió que lo óptimo sería tener pérdidas y pérdidas de oportunidad en neto. Para ello se asumió que el COG (cost of goods), el coste de adquisición de los productos para la empresa, era de un 70% del precio de venta. Se utilizó este porcentaje porque es la media del COG en las empresas retail en EEUU (tras investigar cifras y artículos en internet). Por lo tanto, si el coste de adquisición es el 70% del precio de venta, eso quiere decir que el beneficio final de cada venta es del 30% de cada venta.

Teniendo este porcentaje en venta, se calcularon las pérdidas y pérdidas de oportunidad. El análisis se centró en Supermarket y también Supermarket 3, ya que son los productos más perecederos. Es decir, si sobrara stock de Supermarket y más en concreto Supermarket 3 (los productos con más ventas intradía, y por lo tanto con la fecha de caducidad más próxima) habría que desecharlos ya que no pueden revenderse, caducan pronto. Esto es muy común en productos de supermercado. Esto se traduce en pérdidas directas a la empresa.

Una vez se tuvieron las cifras de stock sobrante y stock que se podría haber vendido si se dispusiese de él, se pasó a neto. Se hablará de los resultados obtenidos más adelante.

- Descripción del plan piloto y su implementación en tiendas

Para la implementación del plan piloto se tuvo en cuenta la segmentación de tiendas creando distintos escenarios que se ordenaron según el riesgo que podrían suponer para la empresa hacer una primera aplicación del modelo, como se mencionó anteriormente. Tras su evaluación, se decidió por la opción sometida a menos riesgo, el cluster de tiendas rezagadas. Al tratarse de 2 tiendas con la menor actividad de las 10 y sin tener un crecimiento muy pronunciado, evaluamos que en caso de tener que realizar modificaciones eran las tiendas que permitían más margen de maniobra y que en caso de error, tenían menos en juego.

La idea del plan piloto es tener un primer contacto del modelo con la realidad y ver cómo éste responde ante las predicciones que está realizando frente a las ventas ocurridas durante un periodo de tiempo futuro. Al ver cómo este responde y poder evaluar su acierto contra los datos reales seremos capaces de realizar mejoras con nuevos datos y añadiendo comportamientos que se consideren pertinentes

para mejorar el rendimiento del modelo y así mejorar las métricas correspondientes: Pérdidas y Pérdidas de Oportunidad.

Para poder implementar el modelo dentro del plan piloto, se diseñaron los dos flujos de trabajo correspondientes para poder trabajar paso a paso de manera automática. Estos flujos de trabajo cuentan con APIs cuya función es recolectar la información tanto de las tiendas (ventas), departamento de compras (Stock comprado y distribuido a las tiendas) y del departamento de marketing (Promociones aplicadas a ciertos productos). Una vez se dispone de estos datos el flujo de trabajo se dispone a tratarlos para:

1. Realizar de forma anual la clusterización de productos y tiendas. Con los resultados de los productos se puede disponer de información relevante sobre el comportamiento de los productos para el departamento de marketing y es por ello que mediante una API se enviará la información de manera automática. En cuanto a la segmentación de las tiendas se recogerá la información en el siguiente flujo de trabajo y se usará para implementar el plan piloto.

2. Aplicar el modelo con una recurrencia semanal tras modelar los datos de las 3 fuentes mencionadas anteriormente y, en esta primera instancia, seleccionando aquellos que sean sólo de las tiendas donde se va a aplicar el plan piloto. Una vez se dispone de dichas predicciones el flujo de trabajo las envía al ERP mediante una API para que el departamento de compras pueda realizar las comandas. Es importante resaltar que dependiendo de qué tipo de producto sea, la predicción se redondeará mediante el error hacia arriba o abajo. Esto es debido al comportamiento que pueden tener los productos respecto a su fecha de caducidad.

Una vez el plan piloto se haya validado y esté aprobado por la junta directiva, se ha diseñado un “roadmap” de implementación en el resto de establecimientos de DS Market. En segunda instancia se aplicaría al resto de tiendas que se nos ha proporcionado para este trabajo (2 de Nueva York, 3 de Boston y 3 de Filadelfia) y al cabo de un año más (2 años desde el día 0 de implementación de modelo

piloto) en caso de haber sido exitoso se aplicaría en el resto de tiendas del país.

Para poder realizar esta implementación de manera escalonada sin perder tiempo ni información en el proceso, se ha diseñado un flujo de trabajo en el que los jueves se realizan las predicciones con el modelo para el que el viernes se pueda realizar la comanda desde el departamento de compras y el lunes haya una distribución a nivel nacional de cada uno de los supermercados.

- Evaluación del impacto económico

Se utilizó un notebook de google colab que está adjunto para calcular el neto de las pérdidas y pérdidas de oportunidad. Para ello se aplicó el modelo sobre las ventas de los últimos 4 meses. Si el resultado daba que iba a sobrar stock de un producto, se cogía la diferencia entre el stock real y el predicho y se multiplicaba por 0.7 que es el precio de adquisición. Esto nos da cuánto dinero hemos perdido en neto.

Si en el caso contrario, daba que se podría haber vendido más stock del adquirido, se cogía la diferencia entre el stock predicho y el real (esta vez está negativo, ya que es lo contrario de la diferencia anterior), y se multiplica por 0.3, que es el margen de beneficio unitario. Esto da cuánto dinero se ha dejado de ganar.

Los resultados arrojaron lo siguiente:

	SUPERMARKE T	SUPERMARKET_3
PÉRDIDAS	\$ 206.838	\$ 115.249

PÉRDIDA DE OPORTUNIDAD	\$ 125.249	\$ 74.366
PORCENTAJE DE PÉRDIDAS / BENEFICIOS	20%	11%
INCREMENTO DE BENEFICIOS DEBIDO A LA OPORTUNIDAD	12%	7%

En esta tabla se pueden apreciar las pérdidas y pérdidas de oportunidad para Supermarket y Supermarket 3 dentro del piloto conservador (tiendas rezagadas).

También se obtuvieron dos ratios, pérdidas sobre beneficios y porcentaje de beneficios gracias a la oportunidad. Esto lo se calculó teniendo los beneficios del piloto conservador (tiendas rezagadas) los últimos 4 meses, que nos daba 1,03 millones (multiplicando todas las ventas por producto por su precio unitario y por 0.3, y luego sumarlo todo). Así que luego era solo dividir las pérdidas entre estos beneficios (pérdidas / beneficios), y dividir la pérdida de oportunidad entre los beneficios (pérdida de oportunidad / beneficios).

Estos son los resultados que arroja el piloto conservador. No son para nada negativos, pero podrían ser mejores. Es por ello que durante la implementación del piloto en las infraestructuras de la empresa y los constantes testeos para optimizarlo, se intentará ir realizando ajustes para que los resultados sean todavía mejores.

9. Discusión

El desarrollo de un modelo predictivo y su aplicación en la gestión de inventarios en DSMarket no solo resuelve los problemas operativos inmediatos de exceso y faltante de stock, sino que también tiene implicaciones estratégicas más amplias que pueden transformar la manera en que la empresa opera y compite en el mercado. A continuación, se exploran estas implicaciones y su posible impacto en las estrategias de la empresa.

1. Mejora en la Planificación y Agilidad Operativa

Uno de los principales beneficios de contar con un sistema predictivo preciso es la capacidad de planificación proactiva. Anteriormente, DSMarket reaccionaba ante los problemas de stock de manera retrospectiva, enfrentando pérdidas y costos elevados por decisiones mal informadas. Con el nuevo sistema de forecasting basado en datos históricos y modelos avanzados, DSMarket puede anticipar la demanda de forma mucho más precisa, lo que le permite ajustar su cadena de suministro de manera ágil.

Esto transforma la estrategia operativa de la empresa, ya que pasa de un enfoque reactivo a uno proactivo, permitiendo una planificación más ajustada y eficiente. La optimización de inventarios basada en predicciones precisas no solo reduce las pérdidas por productos perecederos y costosos almacenamientos, sino que también permite una asignación más inteligente de recursos. A largo plazo, esta agilidad operativa mejora la competitividad de DSMarket en el mercado, reduciendo tiempos de respuesta a las demandas del consumidor y mejorando la disponibilidad de productos clave en eventos de alto impacto como el Black Friday o el Superbowl.

2. Reducción de Costos y Mejora de Margen de Beneficio

El impacto del modelo predictivo en los costos operativos de DSMarket es significativo. La reducción del exceso de inventario no solo disminuye los costos de almacenamiento, sino que también

minimiza el desperdicio, especialmente en productos perecederos, lo que se traduce directamente en un incremento del margen de beneficio. En un sector con márgenes ajustados como el retail, esta mejora en la eficiencia puede tener un impacto considerable en la rentabilidad general de la empresa.

Adicionalmente, al reducir las situaciones de out-of-stock, DSMarket también incrementa sus ingresos al evitar perder ventas. Esto tiene un impacto positivo en la confianza del cliente y en la fidelización, ya que los consumidores perciben a DSMarket como una tienda confiable que siempre tiene el producto que buscan. Esta mejora en la experiencia del cliente puede traducirse en una mayor lealtad y un incremento en la participación de mercado a lo largo del tiempo.

3. Optimización de las Estrategias de Marketing y Promociones

Otro aspecto estratégico que puede verse potenciado con el uso de estos modelos predictivos es la optimización de las estrategias de marketing y promociones. Con la segmentación precisa de productos y la identificación de patrones de demanda a través del clustering, DSMarket puede personalizar sus campañas promocionales para que estén alineadas con la demanda anticipada de los clientes.

Por ejemplo, los productos que experimentan picos de demanda durante eventos específicos (“Event Boosted Products”) pueden beneficiarse de campañas dirigidas que maximicen su exposición justo antes de esos eventos. Esto no solo aumenta las ventas, sino que también mejora la eficiencia de las campañas de marketing al dirigir los recursos a las áreas que generarán un mayor retorno de la inversión. Por el contrario, los productos que sufren caídas durante eventos importantes (“Event Loss Products”) pueden recibir estrategias promocionales que busquen mitigar estas pérdidas.

A nivel de tiendas, la segmentación también permite ajustar las promociones según el comportamiento local. Las tiendas consolidadas pueden beneficiarse de campañas de retención de

clientes, mientras que las tiendas emergentes pueden priorizar estrategias de crecimiento, y las tiendas rezagadas pueden recibir intervenciones más agresivas para reactivar su desempeño.

4. Decisiones Basadas en Datos para Expansión y Optimización de Tiendas

El análisis de las tiendas a través del clustering también ofrece insights valiosos que pueden influir en las decisiones de expansión y optimización de la red de tiendas de DSMarket. Al identificar tiendas emergentes con alto potencial de crecimiento, la empresa puede asignar más recursos a esas ubicaciones, optimizando inventarios y enfocando las estrategias de expansión en las áreas geográficas más prometedoras.

De la misma forma, la identificación de tiendas rezagadas permite a DSMarket reorientar su estrategia en estas ubicaciones, ajustando el portafolio de productos o aplicando promociones específicas para revitalizar el rendimiento. Además, la comparación de la cuota de mercado y el crecimiento interanual entre las diferentes regiones permite que la empresa ajuste su estrategia de inversión de manera más precisa, maximizando el impacto de sus recursos.

5. Impacto en la Cultura de Toma de Decisiones de la Empresa

El uso de modelos predictivos avanzados, junto con dashboards que permiten visualizar datos en tiempo real, transforma la cultura de toma de decisiones en DSMarket. En lugar de depender únicamente de la intuición o experiencias pasadas, los directivos y gerentes de tiendas pueden basar sus decisiones en datos concretos y modelos validados. Esto no solo aumenta la confianza en las decisiones operativas y estratégicas, sino que también fomenta una cultura de mejora continua.

Con un acceso transparente a la información, los equipos de marketing, compras y operaciones pueden colaborar de manera más efectiva, alineando sus decisiones en torno a objetivos comunes,

como la optimización de inventarios o la maximización de ventas durante eventos clave. A nivel macro, este enfoque basado en datos tiene el potencial de mejorar la cohesión interna de la empresa y alinear a todos los departamentos hacia la misma visión estratégica.

6. Escalabilidad y Expansión Internacional

A largo plazo, el éxito de este modelo de predicción y su implementación podría sentar las bases para la expansión internacional de DSMarket. La capacidad de predecir con precisión la demanda en diferentes regiones y ajustar las estrategias de inventario y promociones es crucial para la entrada a nuevos mercados. Al aplicar estos mismos modelos en diferentes contextos geográficos, DSMarket podría adaptar su oferta y su gestión de inventarios a las necesidades específicas de cada mercado, lo que facilitaría una expansión exitosa.

El modelo también es escalable en términos de su implementación en otras áreas del negocio. Si bien el piloto inicial se centró en tiendas rezagadas y productos perecederos, el éxito de la prueba sugiere que, con ajustes y mejoras, el modelo puede aplicarse a toda la cadena de tiendas y categorías de productos, lo que generaría un impacto aún mayor en la rentabilidad de la empresa.

10. Conclusiones

El proyecto ha demostrado no solo el valor de la analítica avanzada y los modelos predictivos en la gestión operativa de DSMarket, sino también su potencial para influir en la estrategia general de la empresa. Al transformar la manera en que DSMarket gestiona su inventario, planifica promociones y toma decisiones, el proyecto se alinea con los objetivos a largo plazo de eficiencia y crecimiento de la empresa.

La implementación exitosa del modelo no solo tendrá un impacto inmediato en la reducción de costos y mejora de la satisfacción del cliente, sino que también fortalecerá la posición de DSMarket en el competitivo entorno del retail. A medida que se optimicen los modelos y se expandan a más áreas del negocio, DSMarket podrá capitalizar oportunidades de crecimiento, consolidando su liderazgo en la industria y fomentando una cultura de innovación impulsada por datos.