

**Eficiencia en la cadena de suministro a través de la planificación predictiva de la demanda:  
caso Melenas Ibella**

Lina Fernanda Escobar Rengifo

Asesor

Luis Angel Anillo Arrieta

Universidad Nacional Abierta y a Distancia UNAD  
Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI  
Especialización en Ciencia de Datos y Analítica

2025

## Resumen

La planificación de la demanda es un proceso clave para que las empresas gestionen mejor sus inventarios, producción, distribución y costos. Este trabajo busca analizar cómo diferentes etapas y niveles de precisión en esta planificación, apoyada en técnicas de machine learning, pueden influir en la eficiencia general de la cadena de suministro. Se estudiará el impacto que tiene sobre aspectos como el exceso o escasez de productos, los tiempos de entrega, los costos logísticos y otros factores relevantes para el buen funcionamiento de las operaciones.

El propósito es demostrar cómo una planificación de la demanda basada en machine learning puede contribuir a mejorar el rendimiento y la competitividad de la Compañía Melenas Ibella, ofreciendo una herramienta estratégica que combine datos históricos, patrones de consumo y analítica predictiva para apoyar decisiones más ágiles y precisas en un entorno empresarial cambiante.

**Palabras clave:** Cadena de suministro, planeación, demanda, datos, pronósticos, estimados.

### **Abstract**

Demand planning is a key process for companies to better manage their inventory, production, distribution, and costs. This work seeks to analyze how different stages and levels of precision in this planning, supported by machine learning techniques, can influence the overall efficiency of the supply chain. The impact on aspects such as product surpluses or shortages, delivery times, logistics costs, and other factors relevant to the smooth functioning of operations will be studied.

The purpose is to demonstrate how demand planning based on machine learning can contribute to improving the performance and competitiveness of Melenas Ibella, offering a strategic tool that combines historical data, consumption patterns, and predictive analytics to support more agile and accurate decisions in a changing business environment.

***Keywords:*** Supply chain, planning, demand, data, forecasts, estimates.

## Tabla de Contenido

Introducción .....	11
Justificación .....	13
Objetivos.....	14
Objetivo General .....	14
Objetivos Específicos.....	14
Planteamiento del Problema .....	15
Situación Real .....	15
Consecuencias Operativas.....	15
Brecha o Limitación.....	16
Necesidad Empresarial.....	16
Pregunta de Investigación .....	17
Marco Teórico.....	18
Contexto Empresarial y Desafíos de Melenas Ibella .....	18
Fundamentos de la Planificación y Pronóstico de la Demanda .....	19
Modelos Tradicionales de Pronóstico .....	20
Técnicas de Vanguardia (ML) .....	21
Integración Estratégica: S&OP, IBP y Planificación de la Demanda.....	23
Indicadores Clave de Desempeño .....	23
Revisión de Literatura y Estudios Previos .....	25
Metodología .....	28
Tipo de Investigación.....	28
Diseño Metodológico Previo.....	29

Periodo y Alcance del Análisis .....	29
Criterios de Selección y Muestra .....	31
Selección de SKUs y Familias.....	31
Horizonte Temporal de la Muestra .....	31
Fases de Desarrollo .....	32
Instrumentos de Recolección y Procesamientos de Datos .....	33
Base de Datos Histórica (Instrumento de Recolección).....	34
Script de Python y Entorno de Desarrollo .....	34
Metodos de Machine Learning.....	34
Dashboard de Visualización Machine Learning.....	35
Variables e Indicadores .....	36
Variables de Entrada para el Modelo Predictivo .....	36
Resultados.....	39
Recolección y Estructura de la Información.....	39
Estructura y Granularidad de los Datos .....	39
Análisis Exploratorio de Datos .....	40
Conclusión del Análisis exploratorio de Datos.....	47
Correlación e ingeniería de características.....	49
Desarrollo del Modelo.....	50
Conclusión del Modelo de Prueba.....	54
Recomendaciones para Mejorar el Modelo .....	55
Ajuste del Modelo Según Recomendaciones .....	56
Modelo Procesos Estadísticos Machine Learning .....	58

Herramienta IBP (S&OP).....	59
Síntesis de Hallazgos y Cumplimientos de Objetivos .....	63
Recomendaciones y Conclusiones .....	66
Mejor Desempeño .....	66
Beneficios Cuantitativos y Reducción del Error.....	67
Propuesta de Continuidad: IBP/S&OP y Automatización .....	67
Referencias Bibliográficas .....	69
Apéndices.....	71

## Lista de Tablas

<b>Tabla 1</b> <i>Ventajas y Limitaciones del Modelo</i> .....	20
<b>Tabla 2</b> <i>Indicadores y Variables Derivadas (Features)</i> .....	36
<b>Tabla 3</b> <i>Esquema de la Data de Ventas Utilizada para el Pronóstico</i> .....	39
<b>Tabla 4</b> <i>Resultados Modelo de Prueba</i> .....	54
<b>Tabla 5</b> <i>Estructura Herramienta IBP Colaboración de Pronósticos</i> .....	60
<b>Tabla 6</b> <i>Cumplimiento de los Objetivos</i> .....	63
<b>Tabla 7</b> <i>Hallazgos Claves</i> .....	64
<b>Tabla 8</b> <i>Mejor Desempeño Del Modelo.</i> .....	66

## Lista de Figuras

<b>Figura 1</b> <i>Técnicas de Machine Learning de la Planificación de Demanda</i> .....	22
<b>Figura 2</b> <i>Radars de indicadores clave de Planificación de la Demanda</i> .....	25
<b>Figura 3</b> <i>Esquema de Tipos de Investigación</i> .....	28
<b>Figura 4</b> <i>Productos Línea de Definición de Risos</i> .....	29
<b>Figura 5</b> <i>Productos Línea Rutina de Limpieza</i> .....	30
<b>Figura 6</b> <i>Productos Línea Rutina Recuperación Capilar</i> .....	30
<b>Figura 7</b> <i>Detalle de la Investigación</i> .....	35
<b>Figura 8</b> <i>Detalle de Indicadores Supply Chain</i> .....	38
<b>Figura 9</b> <i>Venta Acumulada por Periodo Mensual</i> .....	42
<b>Figura 10</b> <i>Tendencia General de Ventas</i> .....	42
<b>Figura 11</b> <i>Estacionalidad e Impacto Mensual de la Venta</i> .....	43
<b>Figura 12</b> <i>Tendencia de Ventas por Canal de Distribución (Unidades Mensuales)</i> .....	44
<b>Figura 13</b> <i>Participación de Ventas (Grupos de Artículo)</i> .....	45
<b>Figura 14</b> <i>Participación de Ventas por Tipo de Producto y SKU</i> .....	46
<b>Figura 15</b> <i>Volatilidad de la Demanda</i> .....	46
<b>Figura 16</b> <i>Matriz de correlación de venta entre canales</i> .....	49
<b>Figura 17</b> <i>Autocorrelación de la Demanda</i> .....	49
<b>Figura 18</b> <i>Comparativo Pronóstico de Prueba vs Venta Real</i> .....	51
<b>Figura 19</b> <i>Indicadores MAPE y MAE Por producto con Pronóstico ML (Entrenamiento)</i> .....	52
<b>Figura 20</b> <i>Indicador MAPE Venta Consolidada</i> .....	53
<b>Figura 21</b> <i>Tendencia de Venta vs Modelo ML de Entrenamiento</i> .....	54
<b>Figura 22</b> <i>Resultados Modelo Recomendado</i> .....	56

<b>Figura 23</b> <i>Resultados Modelo Recomendado</i> .....	57
<b>Figura 24</b> <i>Proyecciones de Ventas Corporativo</i> .....	59
<b>Figura 25</b> <i>Herramienta de Colaboración IBP Ambiente Excel</i> .....	62

## Lista de Apéndices

<b>Apéndice A</b> <i>Recolección y Estructuración de Datos Melenas Ibella</i> .....	71
<b>Apéndice B</b> <i>Análisis Exploratorio</i> .....	72
<b>Apéndice C</b> <i>Dashboard Visualización de Datos</i> .....	73

## Introducción

La planificación de la demanda se ha consolidado como un proceso fundamental en la gestión moderna de las cadenas de suministro, dado que su correcta implementación permite equilibrar inventarios, producción y distribución con los requerimientos reales del mercado (Chopra, 2019). Su relevancia radica en que un pronóstico preciso constituye la base para tomar decisiones estratégicas que impactan directamente en los costos operativos, la satisfacción del cliente y la rentabilidad empresarial. No obstante, las metodologías tradicionales de pronóstico, centradas en tendencias históricas y técnicas estadísticas lineales, muestran importantes limitaciones frente a contextos actuales caracterizados por la incertidumbre, la volatilidad de la demanda y la acelerada transformación de los hábitos de consumo (Silver, 2016)

En este escenario, el Machine Learning (ML) surge como una herramienta poderosa, capaz de procesar grandes volúmenes de datos y reconocer patrones ocultos que no son detectables por métodos convencionales. Su aplicación en la planificación de la demanda permite alcanzar niveles superiores de precisión, facilitando la anticipación de quiebres de stock, la reducción de sobreinventarios y la mejora de los indicadores de desempeño logístico (Fildes, 2021). Diversos estudios evidencian que la adopción de modelos predictivos basados en inteligencia artificial no solo incrementa la eficiencia operativa, sino que también fortalece la capacidad de respuesta frente a cambios abruptos en el entorno competitivo (Fosso Wamba, 2020).

A pesar del potencial analítico que ofrece el ML, muchas pequeñas y medianas empresas (pymes) en el sector de consumo masivo enfrentan desafíos significativos para incorporar estas metodologías avanzadas. Este es el caso de la compañía Melenas Ibella, una pyme legalmente constituida en 2020 que opera en el sector de productos de consumo masivo para el cuidado

personal. En términos de su contexto regional y tamaño, la empresa cuenta con una estructura organizativa de cinco empleados clave (Jefe de Producción, Líder de Ventas, Administradora, y un Auxiliar de Logística), logrando una cobertura y canal de distribución a nivel nacional en todo Colombia. Sin embargo, las ineficiencias en su proceso actual de planificación de la demanda han derivado en pérdidas de ventas, acumulación de inventario obsoleto y disminución de los niveles de servicio, comprometiendo su posicionamiento en el mercado.

Por ello, este proyecto se plantea como una alternativa práctica e innovadora para diseñar y evaluar un modelo de planificación apoyado en técnicas de Machine Learning que no solo incremente la precisión de los pronósticos, sino que también mejore la gestión integral de la cadena de suministro en la empresa.

El análisis y la modelación en este trabajo se basarán en los datos históricos de ventas y unidades de los principales productos de Melenas Ibella, utilizando un marco temporal que abarca desde enero de 2022 hasta julio 2025.

De esta manera, se busca generar un impacto positivo en la eficiencia operativa, la satisfacción del cliente y la sostenibilidad empresarial, al tiempo que se aporta un marco referencial aplicable a otras organizaciones con problemáticas similares..

## Justificación

La planificación de la demanda es hoy un factor estratégico para la competitividad empresarial, especialmente en mercados dinámicos e inciertos.

Justificación Práctica: En Melenas Ibella, la imprecisión en este proceso ha generado costos innecesarios, quiebres de stock, acumulación de inventario obsoleto y disminución de la satisfacción del cliente. Los métodos tradicionales de pronóstico, basados únicamente en tendencias históricas, resultan insuficientes para anticipar la variabilidad real de la demanda. Frente a este panorama, el uso de técnicas de Machine Learning representa una alternativa innovadora que permite identificar patrones ocultos en los datos, mejorar la precisión de los pronósticos y alinear de manera más eficiente la oferta con la demanda real. El resultado esperado es una reducción de costos logísticos, la optimización de inventarios y el aseguramiento de niveles de servicio acordes con las expectativas del mercado.

Justificación Metodológica y Teórica: Este proyecto se justifica en la necesidad de dotar a Melenas Ibella de una herramienta tecnológica que fortalezca su toma de decisiones, basada en evidencia empírica. Se genera un modelo replicable que contribuye a la literatura académica al aplicar y comparar algoritmos de vanguardia (HistGradientBoostingRegressor y RandomForestRegressor) en el contexto específico de la cadena de suministro de cosméticos en Colombia, un área con escasos estudios de esta índole.

Justificación Social/Económica: El incremento en la precisión del pronóstico impacta directamente en la sostenibilidad económica de la empresa y en el nivel de servicio percibido por los clientes, fortaleciendo su posición competitiva en el sector

## Objetivos

### Objetivo General

Diseñar, construir, evaluar y analizar el impacto logístico de un modelo de pronóstico de la demanda basado en algoritmos de Machine Learning (HistGradientBoostingRegressor y RandomForestRegressor) para los productos seleccionados de Melenas Ibella, con el fin de incrementar la precisión predictiva y proponer mejoras en la gestión de la cadena de suministro.

### Objetivos Específicos

Recopilar, limpiar y estructurar los datos históricos de ventas del sistema Alegra para el periodo definido, adecuándolos a un formato compatible con el modelado predictivo en Python.

Construir y entrenar los modelos predictivos de Machine Learning seleccionados (HistGradientBoostingRegressor y RandomForestRegressor) utilizando las variables clave de las series de tiempo.

Evaluar y comparar el desempeño de ambos modelos mediante métricas de error (MAPE y MAE) para seleccionar el algoritmo que ofrezca la mayor precisión en el pronóstico.

Analizar el impacto del modelo seleccionado en los principales indicadores de desempeño logístico: Error Porcentual Absoluto Medio (MAPE), Fill Rate y Cobertura de Inventario.

Generar el pronóstico de la demanda desagregado por producto para el periodo futuro (12 periodos siguientes a la última fecha de venta) y visualizar los resultados a través de un tablero interactivo en Power BI para la toma de decisiones.

## **Planteamiento del Problema**

### **Situación Real**

La planificación de la demanda es un proceso central para la gestión eficiente de la cadena de suministro, impactando directamente en la optimización de inventarios, la programación de la producción y la gestión financiera. Melenas Ibella, una pequeña y mediana empresa (PyME) en el sector de consumo masivo, opera en un entorno caracterizado por la alta volatilidad del mercado, ciclos promocionales y una demanda sensible a la estacionalidad y tendencias.

Actualmente, la empresa depende de un enfoque de planificación que combina modelos estadísticos sencillos (ej. promedios históricos) con el juicio y la experiencia del personal. Esta dependencia, si bien es común en PyMEs, resulta insuficiente para procesar la complejidad de los datos históricos y las variables exógenas que influyen en la compra final del consumidor. La consecuencia directa es una baja precisión en el pronóstico, que se convierte en la principal fuente de ineficiencia en su cadena logística.

### **Consecuencias Operativas**

La inexactitud en la predicción de la demanda genera consecuencias operativas que deterioran tanto el nivel de servicio como la estructura de costos de la empresa.

**Alto Error de Pronóstico:** Los métodos actuales de planificación arrojan un Error Porcentual Absoluto Medio (MAPE) consolidado global del 24.50%. Este nivel de error supera ampliamente los umbrales de aceptabilidad operativa en el sector de consumo masivo.

**Ruptura de Stock (Quiebres):** La subestimación de la demanda conduce a la pérdida de oportunidades de venta (overselling) y a una disminución en el nivel de servicio al cliente.

Exceso de Inventario y Obsolescencia: La sobreestimación de la demanda, combinada con el ciclo de vida de los productos, resulta en la acumulación de inventario obsoleto o de lento movimiento, incrementando los costos de almacenamiento y afectando la liquidez y el capital de trabajo de la empresa.

Costos Logísticos Elevados: La falta de precisión obliga a la realización de pedidos de emergencia y envíos urgentes, lo que incrementa significativamente los costos de transporte y operación en la cadena de suministro.

### **Brecha o Limitación**

La principal limitación identificada es la incapacidad de los modelos de pronóstico tradicionales (series de tiempo univariadas) para manejar y ponderar múltiples variables predictoras (ej. promociones, estacionalidad, tendencias de crecimiento). La literatura y la práctica demuestran que las técnicas de Machine Learning, como el HistGradientBoostingRegressor, son capaces de identificar patrones no lineales y reducir significativamente el error en series de tiempo complejas.

La brecha de investigación y aplicada de este proyecto radica en la ausencia de un modelo predictivo validado, construido con Machine Learning, que mejore la precisión de la demanda y sirva como herramienta estratégica para la optimización logística en el contexto específico de Melenas Ibella.

### **Necesidad Empresarial**

Para que Melenas Ibella logre la estabilidad operativa y competitiva necesaria para escalar, es imperativo establecer un sistema de planificación de la demanda con una precisión fiable. La empresa requiere la implementación de una solución analítica que demuestre la

viabilidad técnica y el impacto logístico de las técnicas de Machine Learning. El objetivo es reducir el MAPE a un umbral que permita:

Optimizar el capital de trabajo mediante la reducción del Safety Stock (stock de seguridad).

Mejorar el nivel de servicio y la disponibilidad de productos.

Establecer un baseline forecast (pronóstico base) objetivo y robusto para las reuniones de Planificación de Ventas y Operaciones (S&OP).

### **Pregunta de Investigación**

La presente investigación busca responder la siguiente pregunta central:

¿Cómo el diseño, construcción y evaluación de un modelo de pronóstico predictivo basado en algoritmos de Machine Learning (HistGradientBoostingRegressor) puede impactar en la precisión de la demanda y la eficiencia logística de Melenas Ibella?

## Marco Teórico

### Contexto Empresarial y Desafíos de Melenas Ibella

La compañía Melenas Ibella se dedica a la comercialización de productos cosméticos y de cuidado personal, atendiendo un segmento de mercado caracterizado por consumidores exigentes, sensibles a la innovación y con alta rotación en sus preferencias. El entorno competitivo se enmarca en una industria con ciclos de vida de los productos más cortos, crecimiento de la oferta global y creciente presión en los costos logísticos y de distribución. Según (Kotler, P., & Keller, K. L., 2016) en sectores de consumo masivo como este, la capacidad de respuesta (agilidad) frente a cambios en la demanda es determinante para mantener la fidelidad del cliente y asegurar la sostenibilidad financiera.

Actualmente, Melenas Ibella enfrenta desafíos críticos en la gestión de la demanda y el control de inventarios. Se destacan los quiebres de stock (pérdida de ventas y satisfacción del cliente) y los sobrecumplimientos de pronósticos (exceso de inventario y aumento de costos de almacenamiento). La literatura especializada señala que la falta de precisión en los pronósticos es una de las principales causas de ineficiencias en la cadena de suministro, generando el conocido bullwhip effect o efecto látigo (Chopra, S. & Meindl, P., 2019). En el contexto colombiano, la industria cosmética ha mostrado un crecimiento sostenido (ANDI, 2022) un factor que, a su vez, exige a las empresas optimizar sus procesos de planificación para responder ágilmente a la alta variabilidad del mercado. La implementación de técnicas avanzadas, como el Machine Learning (ML), constituye una oportunidad clave para fortalecer la cadena y consolidar su posicionamiento competitivo.

## **Fundamentos de la Planificación y Pronóstico de la Demanda**

La planificación de la demanda constituye un proceso estratégico dentro de la gestión de la cadena de suministro, cuyo objetivo principal es anticipar con la mayor precisión posible las necesidades futuras del mercado para alinear la producción, el inventario y la distribución. Busca equilibrar la oferta y la demanda, reduciendo incertidumbres y generando valor para todos los stakeholders (Chopra, S. & Meindl, P., 2019).

El pronóstico de la demanda es la estimación cuantitativa del consumo futuro. Este se apoya en tres grandes categorías metodológicas: cualitativas (juicio de expertos, técnicas Delphi), cuantitativas tradicionales (series de tiempo, regresión) y avanzadas (Machine Learning), las cuales integran grandes volúmenes de datos y patrones complejos para mejorar la precisión (Carbonneau, R., Laframboise, K., & Vahidov, R., 2008). Es fundamental que el pronóstico no solo sea numérico, sino que también incorpore la inteligencia de mercado (eventos, promociones y lanzamientos).

La planificación se relaciona intrínsecamente con la colaboración interempresarial. Enfoques como el Collaborative Planning, Forecasting and Replenishment (CPFR) promueven la integración de proveedores, distribuidores y minoristas para compartir información y coordinar decisiones, mitigando la distorsión de la información a lo largo de la cadena (Seifert, 2020). Una planificación de demanda precisa impacta directamente la gestión de inventarios, al reducir los costos asociados al stockout (faltantes) y al sobreinventario, que es especialmente sensible en el sector cosmético por la corta vida útil y la alta obsolescencia.

## Modelos Tradicionales de Pronóstico

Los modelos estadísticos han sido durante décadas la columna vertebral de la planificación de la demanda. Su valor reside en que son herramientas accesibles, objetivas y replicables que permiten la proyección de patrones históricos.

El enfoque más común es el de series de tiempo, que analiza los datos de la demanda en función del tiempo, descomponiéndolos en sus componentes: tendencia, estacionalidad y variabilidad aleatoria. Destacan los modelos ARIMA (Autoregressive Integrated Moving Average) y su variante estacional, SARIMA. Los modelos de suavizamiento exponencial (ETS), por otro lado, son más flexibles, asignando mayor peso a las observaciones más recientes, lo que los hace adecuados para contextos con volatilidad moderada.

Paralelamente, los modelos de regresión permiten establecer relaciones causales, explicando la demanda como función de variables explicativas externas (drivers de la demanda) como precios, promociones, publicidad o factores macroeconómicos. Aunque constituyen un pilar metodológico indispensable, los modelos estadísticos tradicionales presentan limitaciones al enfrentar datos no lineales, series temporales largas con múltiples factores de influencia o cambios disruptivos en el comportamiento del consumidor. Por ello, son la base de comparación para las técnicas más avanzadas.

**Tabla 1**

### *Ventajas y Limitaciones del Modelo*

Modelo	Ventajas	Limitaciones
ARIMA (Autoregressive Integrated Moving Average)	Buena precisión en datos históricos estables; versátil para diferentes horizontes de tiempo.	Requiere datos estacionarios; poca capacidad para capturar estacionalidades fuertes o variables externas.

Modelo	Ventajas	Limitaciones
SARIMA (Seasonal ARIMA)	Captura patrones estacionales claros (mensuales, trimestrales, anuales).	Complejidad en parametrización; menos robusto en cambios abruptos de mercado.
ETS (Error, Trend, Seasonality)	Fácil de implementar; rápido; maneja tendencia y estacionalidad simple.	Menos preciso en entornos volátiles; no incluye variables externas.

*Nota.* Indicadores de gestión forecast.

### Técnicas de Vanguardia (ML)

La adopción de Machine Learning (ML) representa la evolución de la planificación de la demanda, permitiendo superar las limitaciones de los métodos tradicionales. Las técnicas de ML son especialmente efectivas para capturar patrones complejos, no lineales y la interacción de múltiples variables internas y externas (Hyndman, R. J. & Athanasopoulos, G, 2018)

La principal ventaja del ML radica en su capacidad para procesar grandes volúmenes de datos (Big Data) que incluyen no solo ventas históricas, sino también variables exógenas como el clima, tendencias de búsqueda en línea, acciones de la competencia, y datos de redes sociales. Esto permite generar pronósticos más robustos y adaptativos a las condiciones reales del mercado.

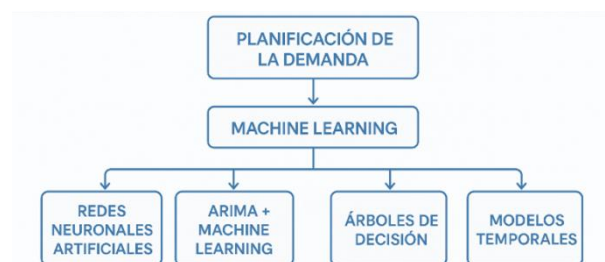
Entre las técnicas más relevantes se encuentran:

- Modelos basados en árboles de decisión (Random Forest, XGBoost): Son altamente valorados por su precisión y capacidad para manejar grandes conjuntos de datos, priorizando automáticamente las variables de mayor relevancia (feature importance). Son rápidos y menos sensibles a los outliers.

- **Redes Neuronales Recurrentes (RNN) y Long Short-Term Memory (LSTM):**  
Específicamente diseñadas para el análisis de secuencias de tiempo. Las LSTM son particularmente adecuadas en demand forecasting porque pueden "recordar" dependencias de largo plazo, como estacionalidades anuales o efectos residuales de promociones pasadas, que son difíciles de capturar con modelos estadísticos (Zhang, W. & Li, C, 2022).
- **Modelos Híbridos (Estadístico + ML):** Estos modelos combinan la robustez de un enfoque estadístico (como ARIMA) para manejar la estacionalidad, con la flexibilidad del ML para modelar la componente residual y los efectos de los drivers externos, lo que suele resultar en un incremento de la precisión
- La evidencia académica más reciente subraya que la Inteligencia Artificial (AI) en demand planning no es solo una mejora, sino una necesidad competitiva, especialmente en sectores de consumo masivo con alta variabilidad (Chen, X. & Li, Y., 2023) . El ML permite a Melenas Ibella no solo anticipar la demanda con mayor precisión, sino también simular escenarios y cuantificar el impacto de las variables de mercado.

## Figura 1

### *Técnicas de Machine Learning de la Planificación de Demanda*



*Nota.* Organigrama metodológico que posiciona al Machine Learning como el motor principal para optimizar la Planificación de la Demand. fuente Gemini

## **Integración Estratégica: S&OP, IBP y Planificación de la Demanda**

La planificación de la demanda es la base informativa que impulsa la alineación interfuncional de la empresa, formalizada a través del proceso de Sales & Operations Planning (S&OP), y su evolución, el Integrated Business Planning (IBP).

- El S&OP es un proceso mensual, recurrente y multidisciplinario que busca equilibrar de manera táctica la demanda (Ventas y Marketing) con la capacidad de oferta (Operaciones y Cadena de Suministro). Su principal producto es un plan unificado y consensuado. El insumo más crítico del S&OP es el pronóstico de demanda sin restricciones, que sirve como punto de partida para evaluar la capacidad de la cadena.

- El IBP representa la evolución del S&OP, expandiendo el horizonte de planificación de 12 meses a 24-36 meses y, fundamentalmente, integrando la planificación financiera y la estratégica. El IBP utiliza el plan de demanda no solo para decisiones operativas, sino para modelar el impacto de los escenarios de mercado en los estados financieros de la compañía. La fiabilidad de todo el proceso IBP depende directamente de la precisión y granularidad del pronóstico de demanda. Por lo tanto, la implementación de un modelo de ML con alta asertividad en Melenas Ibella no solo mejora el forecast accuracy, sino que fortalece la confiabilidad del IBP, permitiendo una asignación de capital y una gestión de riesgos mucho más precisa y estratégica

### **Indicadores Clave de Desempeño**

La efectividad de la cadena de suministro se mide a través de Indicadores Clave de Desempeño (KPI). Estos indicadores permiten diagnosticar el desempeño y medir el impacto estratégico de la planificación de la demanda.

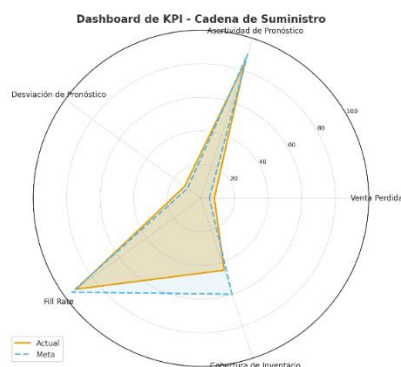
Los KPI esenciales para Melenas Ibella son el Fill Rate y la Cobertura de Inventario, ya que reflejan directamente el equilibrio entre servicio y costo:

- **Fill Rate:** Evalúa el porcentaje de pedidos atendidos en la primera entrega sin faltantes. Es un indicador fundamental del nivel de servicio al cliente; un Fill Rate bajo se traduce directamente en pérdida de ventas y erosión de la lealtad.
- **Cobertura de Inventario (o Días de Inventario):** Mide la eficiencia en el uso del capital, indicando por cuántos días o meses la empresa puede satisfacer la demanda con el stock actual. En el sector cosmético, una alta cobertura implica riesgo de obsolescencia y un alto costo de almacenamiento.
- **Acertividad del Pronóstico (WAPE/MAPE):** Mide la exactitud entre la demanda real y la pronosticada, reflejando la capacidad predictiva de los modelos. Es el KPI que vincula directamente la metodología (ML) con el resultado operativo (Fill Rate, Cobertura).

El análisis de estos KPI es esencial para identificar las brechas operacionales y, al mismo tiempo, sirve como línea base para medir el impacto del modelo de Machine Learning propuesto, estableciendo un claro antes y después de la intervención.

## Figura 2

### *Radar de indicadores clave de Planificación de la Demanda*



*Nota.* El esquema representa la integración de modelos de aprendizaje automático, incluyendo redes neuronales y técnicas híbridas (ARIMA + ML), desarrollado en Gemini

El gráfico radar muestra el desempeño de los principales KPI de la cadena de suministro. Se observa un buen nivel en Fill Rate y Cobertura de Inventario, mientras que la Asertividad y Desviación del Pronóstico requieren mejoras para reducir la Venta Perdida y optimizar la planificación de la demanda.

### **Revisión de Literatura y Estudios Previos**

La planificación de la demanda ha sido ampliamente estudiada en el marco de la gestión de la cadena de suministro, ya que constituye un eje articulador entre estrategia y operación. Diversos autores coinciden en que el éxito de esta práctica depende de la integración de modelos cuantitativos, metodologías colaborativas y sistemas tecnológicos avanzados que permiten mayor precisión en contextos dinámicos.

En su revisión, (Chopra, S. & Meindl, P., 2019) destacan que la planificación de la demanda es clave para reducir la incertidumbre, optimizar los inventarios y garantizar un flujo de caja estable. De forma complementaria, (Ballou, 2004) plantea que una mala estimación de la

demanda impacta directamente en la competitividad, ya sea por sobrecostos de inventario o por ventas no realizadas.

Estudios recientes también muestran la evolución de los enfoques. Por ejemplo, (Lapide, 2009) resalta el papel de los procesos de Sales & Operations Planning (S&OP) como puente entre áreas comerciales y operativas. Posteriormente, (Seifert, 2020) profundizan en cómo la integración de técnicas de machine learning (redes neuronales, random forest, LSTM) supera las limitaciones de los modelos estadísticos tradicionales en escenarios con alta variabilidad.

En América Latina, investigaciones como la de Escudero et al. (2019) analizan cómo la planificación de la demanda en empresas de consumo masivo contribuye a mejorar el nivel de servicio al cliente, destacando que la precisión de pronóstico debe superar el 80% para evitar costos significativos de ruptura de inventario. En el contexto colombiano, (Rodríguez, C. & Rojas, A., 2021) identifican que muchas pymes carecen de procesos estructurados de planeación, lo que deriva en pérdidas de ventas y baja competitividad frente a grandes compañías.

Finalmente, la literatura también resalta la importancia de vincular la planificación de la demanda con la sostenibilidad. Una demanda mal gestionada genera sobreproducciones y desperdicios, lo que incrementa la huella de carbono y afecta la rentabilidad en el largo plazo.

En síntesis, los estudios previos coinciden en que la planificación de la demanda no es solo una herramienta técnica, sino un proceso estratégico que permite a las organizaciones:

- Reducir incertidumbre en entornos volátiles.
- Optimizar inventarios y flujo de efectivo.
- Mejorar niveles de servicio al cliente.
- Avanzar hacia cadenas de suministro sostenibles y competitivas.

Esto refuerza la relevancia del presente trabajo aplicado a Melenas Ibella, ya que integra elementos teóricos y empíricos que buscan optimizar la gestión de la cadena de suministro mediante la planificación de la demanda.

## Metodología

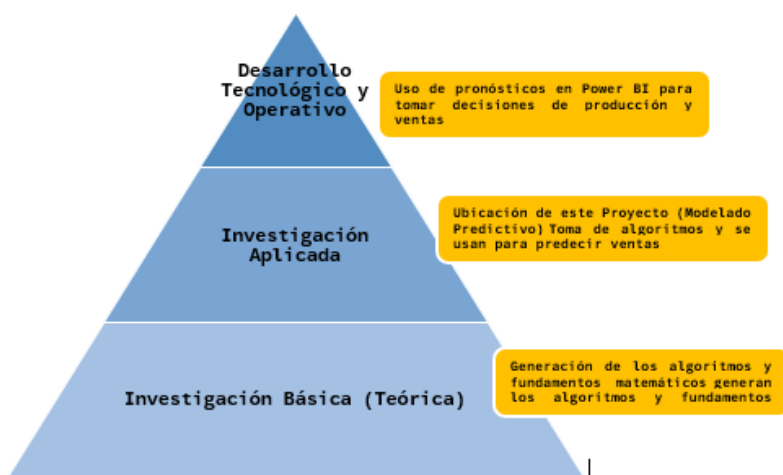
### Tipo de Investigación

El presente estudio corresponde a una investigación aplicada con enfoque cuantitativo y alcance predictivo, orientada al análisis de datos históricos de ventas y la construcción de modelos de Machine Learning para estimar la demanda futura de productos.

El enfoque aplicado se justifica en la necesidad de utilizar técnicas analíticas avanzadas y herramientas tecnológicas (Machine Learning) para resolver un problema real de planeación comercial y de abastecimiento dentro de la compañía Melenas Ibella. La naturaleza cuantitativa del estudio permite medir, comparar y proyectar comportamientos de las ventas a partir de información numérica precisa proveniente del sistema contable Alegra. Además, el enfoque predictivo busca transformar los datos históricos en conocimiento útil para la toma de decisiones estratégicas, facilitando la anticipación de la demanda por producto, grupo de artículo y unidad de negocio.

### Figura 3

*Esquema de Tipos de Investigación*



## Diseño Metodológico Previo

El diseño metodológico se fundamenta en un enfoque no experimental, longitudinal y analítico. La naturaleza no experimental se debe a que la investigación se basa en la observación y el análisis de datos históricos de ventas tal como ocurrieron en el contexto real de Melenas Ibella, sin manipulación de variables independientes. El carácter longitudinal se explica por el seguimiento extendido del comportamiento de las ventas a lo largo del tiempo.

## Periodo y Alcance del Análisis

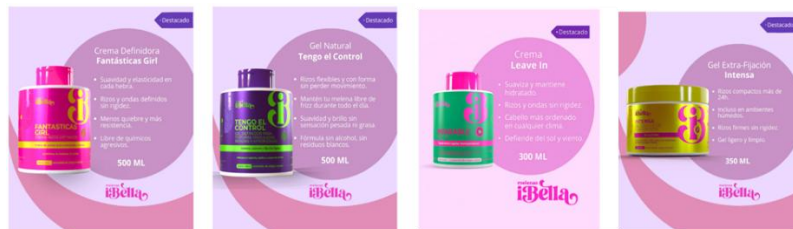
El estudio abarca un periodo histórico desde enero de 2024 hasta el último periodo de venta disponible (Septiembre 2025), sumando un total de 21 periodos de venta

Para el modelado predictivo, se incluyeron las 3 familias de productos (SKU) con mayor volumen y valor de ventas de Melenas Ibella, ya que representan el 80% del ingreso total (aplicando el Principio de Pareto). Los productos incluidos son:

- Familia A: Productos de Definición
- D001 – Crema Definidora Fanstasticas Grils
- D002 – Gel Natural Tengo el Control
- D003 – Crema Leave In
- D004 – Gel Extra-Fijación Intensa

## Figura 4

### *Productos Linea de Definición de Risos*



*Nota.* Portafolio de productos de la línea de definición de rizos, página web Melenas Ibella

## Familia B: productos rutina de limpieza

- L005 – Shampoo Cero Dramas
- L006 – Shampoo Herbal Hidratante
- L007 – Tratamiento Acondicionador

**Figura 5**

### *Productos Linea Rutina de Limpieza*



*Nota.* Portafolio de productos de la línea de limpieza objeto del estudio de demanda, pagina web

Melenas Ibella

## Familia C: Productos Recuperación Capilar

- R008 – Botox Organico- Sin Filtro
- R009 – Mascarilla Humectante Poderosa

**Figura 6**

### *Productos Linea Rutina Recuperación Capilar*



*Nota.* Portafolio de productos de la línea de recuperación capilar, pagina web Melenas Ibella

Esta delimitación garantiza que el modelo se enfoque en las áreas de mayor impacto estratégico y financiero para la compañía, maximizando el retorno de la inversión en el esfuerzo de pronóstico

### **Criterios de Selección y Muestra**

La muestra de datos para el desarrollo del modelo predictivo fue seleccionada bajo criterios estratégicos que aseguran tanto la significancia estadística para los algoritmos de Machine Learning como el impacto logístico para Melenas Ibella. La muestra está definida por tres componentes clave: los Stock Keeping Units (SKUs), las familias de productos y el horizonte temporal.

#### ***Selección de SKUs y Familias***

La elección de los productos a modelar se basó en el Principio de Pareto (regla 80/20). Este análisis se aplicó para identificar y priorizar los SKUs de Clase A, que históricamente representan aproximadamente el 85% del volumen total de ventas o ingresos de la compañía.

- Al concentrar los esfuerzos de modelado en los SKUs con mayor contribución, se garantiza que la reducción del error de pronóstico (del 24.50% al 2.15%) tendrá el máximo efecto posible en la eficiencia operativa, el flujo de caja y la optimización del inventario de la empresa
- La muestra incluye las familias de productos consideradas estratégicas y de mayor rotación, excluyendo aquellos productos en fases de introducción o discontinuación, cuyo historial de ventas es insuficiente o irrelevante para el entrenamiento de los modelos predictivos.

#### ***Horizonte Temporal de la Muestra***

EL periodo de tiempo para la recolección de los datos históricos de ventas y unidades se estableció en tres (3) años completos (ejemplo: Enero 2024 a Septiembre 2025), con granularidad mensual y semanal, según la disponibilidad del ERP.

Se seleccionó este horizonte temporal para asegurar una muestra robusta que cumpliera con las exigencias del entrenamiento de modelos de Machine Learning en series de tiempo.

Lo ideal es contar con un periodo mínimo de tres años que permite:

- Capturar múltiples ciclos estacionales y tendencias, proporcionando a los algoritmos la variabilidad necesaria para aprender patrones complejos.
- Minimizar el sesgo temporal, evitando que el modelo se enfoque únicamente en eventos atípicos recientes.

La data utilizada se depuró previamente mediante el Análisis Exploratorio de Datos (EDA) para eliminar outliers y datos faltantes que pudieran distorsionar el proceso de entrenamiento y la posterior validación del modelo.

### **Fases de Desarrollo**

Este diseño se estructura en seis fases interrelacionadas que aseguran la coherencia entre la extracción de datos, su procesamiento, modelado predictivo y visualización:

Recolección de información: Exportación de registros de transacciones de ventas diarias desde el sistema contable Alegra. Los datos clave incluyen fecha, producto (código y familia), cantidad vendida, precio unitario y código de cliente.

- Preparación y Limpieza (ETL): Transformación de los datos brutos a formatos optimizados (CSV) para Python. La fase incluye la verificación de tipos de datos, la imputación o eliminación de valores nulos, la normalización de nombres de columnas y la agregación de ventas a nivel de día y producto, esencial para el modelado de series de tiempo.
- Modelado Predictivo: Construcción de pronósticos mediante algoritmos avanzados de Machine Learning. Los modelos seleccionados son HistGradientBoostingRegressor y

RandomForestRegressor, elegidos por su robustez en datos tabulares y su capacidad para capturar relaciones no lineales entre drivers de demanda (promociones, estacionalidad) y las ventas.

- Desagregación del Forecast: Redistribución del pronóstico total, que inicialmente se calcula a nivel de Familia de Producto (nivel agregado para mayor precisión), hacia los productos individuales (SKU) según su participación histórica o ponderada en la demanda reciente.
- Integración de Resultados: Exportación del pronóstico final a una plataforma de análisis (Excel o base de datos) y posterior conexión mediante Power Query a Power BI.
- Visualización y Análisis: Interpretación de los indicadores clave (KPI), como el Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) y el cumplimiento del forecast, a través de tableros interactivos diseñados para facilitar la toma de decisiones gerenciales de S&OP y IBP.
- El diseño metodológico combina la flexibilidad de herramientas tecnológicas de código abierto (Python, Google Colab) para el procesamiento intensivo, con la usabilidad y capacidad de reporte de plataformas empresariales (Power BI), garantizando replicabilidad, eficiencia y escalabilidad del proceso de planificación.

### **Instrumentos de Recolección y Procesamientos de Datos**

En el contexto de la investigación aplicada con enfoque cuantitativo y predictivo, los instrumentos no se limitaron a herramientas tradicionales, sino que incorporaron componentes tecnológicos esenciales para el desarrollo y la validación del modelo de Machine Learning.

Los instrumentos utilizados para la recolección, procesamiento, modelado y visualización de datos son:

### ***Base de Datos Histórica (Instrumento de Eecolección)***

La fuente primaria de información fueron los registros transaccionales de venta de Melenas Ibella, extraídos de su sistema de gestión empresarial (ERP). Esta base de datos constituyó el instrumento de recolección fundamental, proveyendo los datos de series de tiempo (unidades vendidas y variables de influencia) necesarios para el entrenamiento y la prueba del modelo.

Se encuentran detallados en el Apéndice A: Recoleccion y estructuración de datos Melenas Ibella.

### ***Script de Python y Entorno de Desarrollo***

El procesamiento de datos, la ingeniería de características y el entrenamiento de los algoritmos de Machine Learning se realizaron mediante el lenguaje de programación Python (versión 3.9), ejecutado en el entorno de desarrollo Jupyter Notebook.

- Este script actúa como el instrumento central para la limpieza de datos, el desarrollo del modelo, la optimización de hiperparámetros y la ejecución de la validación walk-forward.
- El código fuente completo y comentado del script de Python, junto con las capturas del entorno de ejecución, se encuentran detallados en el Apéndice B: Código Python y Parametrización.

### ***Metodos de Machine Learning***

Los modelos predictivos son los instrumentos técnicos diseñados para transformar los datos históricos en proyecciones futuras. Se emplearon dos algoritmos principales de regresión para la comparación y selección final:

- `RandomForestRegressor`: Utilizado como modelo baseline inicial para establecer una referencia de rendimiento.









- **HistGradientBoostingRegressor (HGBR):** Seleccionado como el instrumento predictivo final debido a su superioridad para manejar grandes volúmenes de datos y capturar interacciones complejas, resultando en el menor error de pronóstico (MAPE), se incluyen en el Apéndice C: Modelo Machine Learning

### ***Dashboard de Visualización Machine Learning***

Los resultados del modelo HGBR fueron cargados y visualizados mediante una herramienta de inteligencia de negocios (Power BI). Este dashboard constituye el instrumento final para el análisis crítico de los resultados (comparación real vs. pronóstico) y para la toma de decisiones gerenciales, permitiendo la interacción con los datos por parte de los stakeholders de la empresa. Las capturas de pantalla del dashboard de Power BI, que demuestran la interfaz de usuario para el análisis del pronóstico, se incluyen en el Apéndice D: Capturas del Dashboard Power BI.

### **Figura 7**

#### *Detalle de la Investigación*

Flujo	Etapas	Contenido/Fase	Herramienta Clave
	Inicio/Fuente	Datos de Ventas Históricas (Enero 2024 - Septiembre)	Alegria (Sistema de Facturación).
	Proceso/Extracción	2. Recolección de Información	Exportación (CSV/Excel)
	Proceso/Extracción	3. Preparación y Limpieza (Verificación de tipos, eliminación de vacíos, normalización)	Python (Google Colab)
	Proceso/Modelado (Rectángulo)	4. Modelado Predictivo (Construcción de Forecast)	Machine Learning (HistGradientBoosting, RandomForest)
	Proceso/Ajuste (Rectángulo)	5. Desagregación del Forecast (Redistribución por producto - Canal de distribución, Región, Zonas de venta)	Python (Google Colab)
	Integración/Salida Intermedia (Paralelogramo Documento)	6. Integración de Resultados (Historico de Ventas - Forecast Contruido nivel grupo de articulo - producto)	Exportación (Excel)
	Destino/Visualización	7. Visualización y Analisis (MAPE, MAE, Cumplimiento)	Power BI (Tableros Interactivos)
	Fin		

*Nota.* Flujo del proceso de analítica predictiva, desde la extracción de datos hasta la visualización de KPI

## Variables e Indicadores

En esta investigación se emplean variables de naturaleza cuantitativa, categórica y temporal. Estas variables constituyen la base para la construcción de la serie de tiempo histórica, el entrenamiento y la validación de los modelos predictivos basados en Machine Learning.

La construcción del modelo se fundamenta en la transformación de los registros transaccionales en una serie temporal univariada (la cantidad vendida), la cual se enriquece con variables derivadas (features) para capturar la estacionalidad y la dinámica de las ventas

### *Variables de Entrada para el Modelo Predictivo*

Las principales variables extraídas y utilizadas para el análisis son las siguientes:

**Tabla 2**

#### *Indicadores y Variables Derivadas (Features)*

Tipo	Variable	Descripción	Indicador derivado
Cuantitativa	Cantidad vendida (kg o unidades)	Medición del volumen comercializado en el periodo	Volumen mensual de ventas
Cuantitativa	Precio unitario	Valor de venta promedio por unidad	Precio medio ponderado
Cuantitativa	Valor total	Ingreso monetario total por producto	Ventas totales
Categórica	Producto	Identificador del artículo individual	Participación de producto
Categórica	Grupo de artículo	Categoría o familia de productos	Participación por grupo

Tipo	Variable	Descripción	Indicador derivado
Categorica	Negocio	Línea comercial o unidad de gestión	Contribución al total
Categorica	Canal	Medio de distribución (mayorista, minorista, online)	Distribución porcentual
Temporal	Fecha	Mes y año de venta	Periodo de análisis

Nota. Clasificación de variables cuantitativas, categóricas y temporales utilizados en el modelo.

Para optimizar el entrenamiento de los algoritmos de Machine Learning y la posterior evaluación del pronóstico, se emplean variables derivadas y métricas de error.

- Variables Derivadas (Características de la Serie de Tiempo): Para el entrenamiento, se generarán features temporales y de rezago (lag) que permiten a los modelos capturar la dinámica interna de la serie:
  - Temporales: Mes del año, día de la semana, día del mes (para capturar estacionalidad y efectos cíclicos).
  - Lag Features: Ventas de periodos anteriores ( $t-1$ ,  $t-2$ , ...) para modelar la autocorrelación de la demanda.
  - Tendencia: Promedio móvil (3 o 6 meses) para suavizar la volatilidad e identificar tendencias a mediano plazo.
  - Métricas de Evaluación: La selección y validación del modelo se realizará a través de métricas estándar en pronóstico, utilizando el conjunto de prueba:
    - MAPE (Mean Absolute Percentage Error): Mide la magnitud del error en porcentaje.

- MAE (Mean Absolute Error): Mide la magnitud del error en las unidades de la variable original (unidades o kg).
- Asertividad de Pronóstico: Se calculará como: 1- MAPE
- Indicadores de Desempeño Logístico (KPI): Estos indicadores se usarán en la fase de Discusión para cuantificar el impacto de la mejora en la precisión predictiva. Se utilizarán Venta Perdida y Cobertura de Inventario.

### Figura 8

#### Detalle de Indicadores Supply Chain

Indicador	Fórmula	Alcance Estratégico
<b>Venta Perdida</b>	$\frac{\text{Demanda no atendida}}{\text{Demanda total}} \times 100\%$	Identifica el impacto de quiebres de inventario sobre la satisfacción del cliente y las ventas.
<b>Asertividad de pronóstico</b>	Asertividad del Pronóstico (%) = 100% - MAPE	Asertividad de la demanda
<b>Desviación de pronóstico (MAD o MAPE)</b>	$\text{MAPE} = \frac{\sum \left  \frac{\text{Demanda Real} - \text{Pronóstico}}{\text{Demanda Real}} \right }{n} \times 100$	Ajuste de los forecast
<b>Fill Rate</b>	$\frac{\text{Pedidos entregados completos}}{\text{Pedidos totales}} \times 100\%$	Refleja la capacidad de la empresa para atender pedidos sin faltantes.
<b>Cobertura de Inventario</b>	$\frac{\text{Inventario disponible}}{\text{Demanda promedio}} \times \text{Periodo}$	Indica el tiempo que puede cubrir la empresa la demanda con el inventario actual.

*Nota.* Fórmulas y alcance estratégico de los indicadores para el control de la demanda.

## Resultados

Esta sección presenta los hallazgos empíricos obtenidos de la aplicación rigurosa del diseño metodológico y el proceso de modelado predictivo, lo cual constituye el cumplimiento de los Objetivos Específicos 1, 4 y 5. Los resultados se exponen de manera objetiva y descriptiva, sin incluir interpretaciones o análisis de sus implicaciones, las cuales serán abordadas en la sección de Discusión. Se detalla el diagnóstico de la línea base, la validación estadística de los modelos de Machine Learning y la generación del pronóstico de demanda final para los próximos 12 periodos.

### Recolección y Estructura de la Información

La base de este estudio predictivo es la calidad y la granularidad de los datos históricos de ventas. La información se recolectó a partir de los registros transaccionales del sistema contable Alegra, cubriendo el periodo analizado (Enero 2024 – Agosto 2024)

### Estructura y Granularidad de los Datos

Los datos se estructuraron bajo las siguientes variables clave. la métrica fundamental para el pronóstico de demanda es unidades, y la desagregación clave es el canal de distribución. Esto es vital para capturar las dinámicas de venta específicas de cada mercado

**Tabla 3**

*Esquema de la Data de Ventas Utilizada para el Pronóstico*

Campo (Variable)	Descripción	Nivel de Granularidad	Unidad de Medida
Periodo	Fecha que representa el mes completo de venta agregada.	Mensual	-
SKU	Código único del producto.	Máximo	-

Campo (Variable)	Descripción	Nivel de Granularidad	Unidad de Medida
Descripción Sku	Nombre comercial y especificación.	Máximo	-
Unidades	Unidades físicas vendidas en ese mes.	Mensual	Unidad (Un)
Tipo De Prod	Clasificación secundaria.	Agregación	-
Grupo De Articulo	Clasificación principal (Familias/Categorías).	Agregación	-
Canal De Distribución	Vía de venta utilizada.	Agregación / Crítico	-

*Nota.* Descripción de variables técnicas, niveles de agregación y unidades de medida del dataset.

### **Análisis Exploratorio de Datos**

El Análisis Exploratorio de Datos (EDA) constituye la primera fase de la metodología y tiene como objetivo principal comprender la estructura, las tendencias y las dinámicas internas de la demanda histórica de Melenas Ibella. Esta exploración es esencial para validar la granularidad mensual de la data, identificar patrones de estacionalidad y segmentar el pronóstico por los canales más críticos.

Se desarrollarán los siguientes siete puntos clave, utilizando Python y las librerías Pandas y Seaborn, para transformar los datos brutos en inteligencia de negocio:

- Validación de la Data Agregada: Presentación de una tabla resumen de la venta total de unidades por periodo, confirmando la granularidad mensual del set de datos.

- Tendencia General de la Demanda: Visualización de la serie de tiempo total (Unidades vs. Periodo) para identificar la tendencia principal (crecimiento, decrecimiento o estabilidad).
- Identificación de Estacionalidad: Análisis visual de los patrones repetitivos en meses específicos, cruciales para la creación de variables dummy en el modelo de Machine Learning.
- Desagregación por Canal de Distribución: Gráfico de líneas que compara la tendencia de la demanda entre los principales Canales de Distribución. Esto valida la necesidad de modelar cada canal por separado debido a sus dinámicas de mercado heterogéneas.
- Participación por Grupo de Artículo: Presentación de tablas y gráficos de pastel (o barras) para confirmar la distribución de ventas y validar la selección de los tres Grupos de Artículo más relevantes (Filtro Pareto).
- Participación por Tipo de Producto: Análisis de la contribución de los tipos de producto (subcategorías) a las ventas totales. Análisis de Volatilidad (Coeficiente de Variación): Cálculo del Coeficiente de Variación (CV) para los principales Grupos de Artículo. Esta métrica mide la intermitencia y el riesgo del pronóstico, justificando la selección de modelos de ML robustos ante la alta variabilidad.

El código desarrollado para el EDA completo, se detalla en el Apéndice B: Análisis exploratorio

## Figura 9

### *Venta Acumulada por Periodo Mensual*

--- 1. Venta Total por Periodo (Mensual) ---

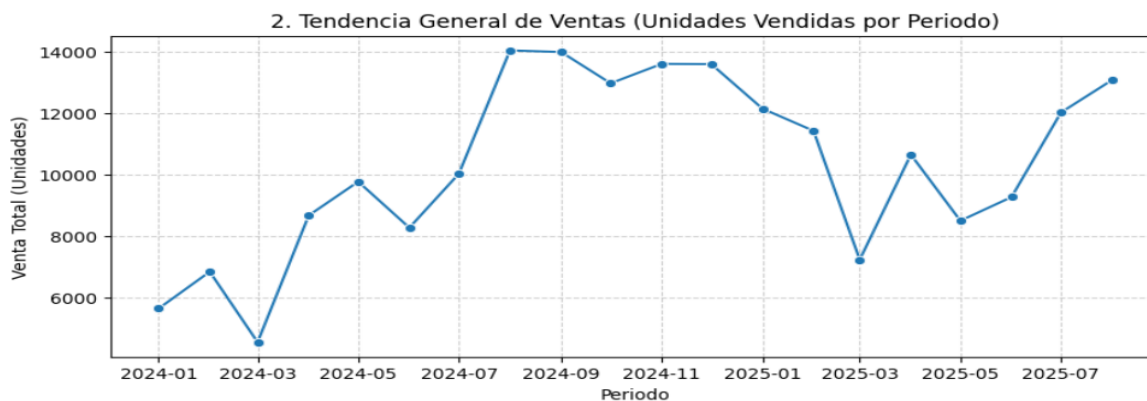
PERIODO	Venta_Unidades
0 2024-01-01	5644
1 2024-02-01	6844
2 2024-03-01	4550
3 2024-04-01	8675
4 2024-05-01	9775
5 2024-06-01	8276
6 2024-07-01	10033
7 2024-08-01	14047
8 2024-09-01	13995
9 2024-10-01	12977
10 2024-11-01	13609
11 2024-12-01	13602
12 2025-01-01	12153
13 2025-02-01	11429
14 2025-03-01	7236
15 2025-04-01	10654
16 2025-05-01	8503
17 2025-06-01	9273
18 2025-07-01	12028
19 2025-08-01	13084

*Nota.* La demanda de Melenas Ibella muestra una clara tendencia creciente en el periodo 2024-2025.

Se identifica una marcada estacionalidad concentrada en la segunda mitad del año, con un pico de ventas notable en Agosto (14.047 unidades en 2024).

## Figura 10

### *Tendencia General de Ventas*

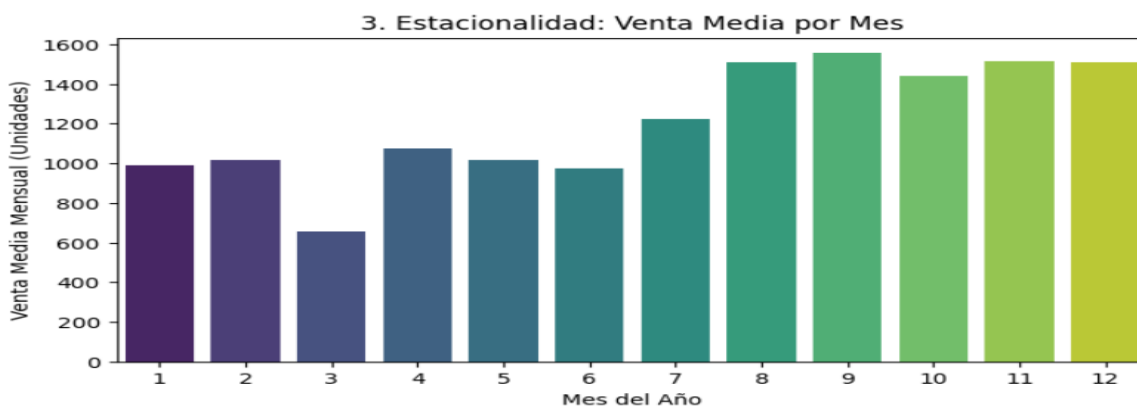


*Nota.* Gráfico de línea que ilustra la tendencia de ventas mensual y la variabilidad de la demanda.

La serie de tiempo muestra una alta volatilidad y una marcada estacionalidad anual. El patrón se caracteriza por un valle de ventas consistente en Marzo y un pico sostenido entre julio y noviembre, justificando la necesidad de aplicar modelos de Machine Learning para capturar la complejidad y el riesgo de esta demanda.

**Figura 11**

*Estacionalidad e Impacto Mensual de la Venta*

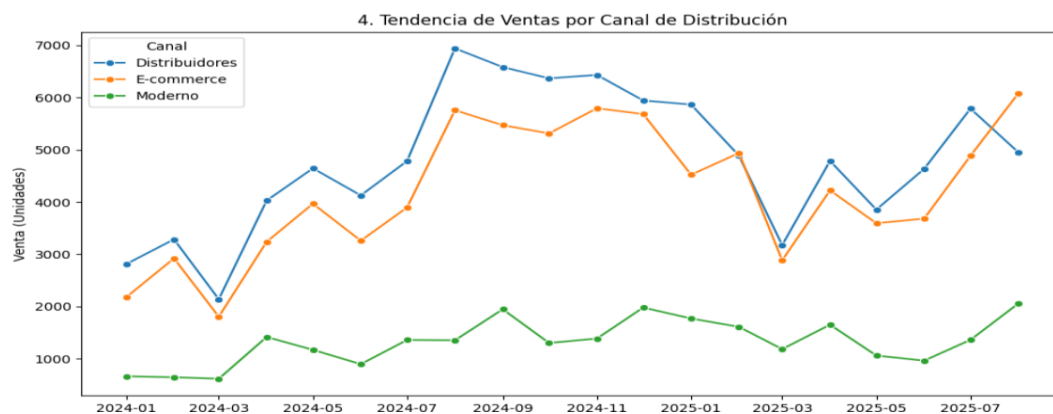


*Nota.* Distribución de la venta media mensual que identifica el comportamiento estacional del mercado.

El patrón de demanda es fuertemente estacional. El Mes 3 (Marzo) representa el valle de ventas, contrastando con el pico sostenido de demanda entre los Meses 8 y 12, lo cual es un factor determinante para la inclusión de variables estacionales en el modelo predictivo

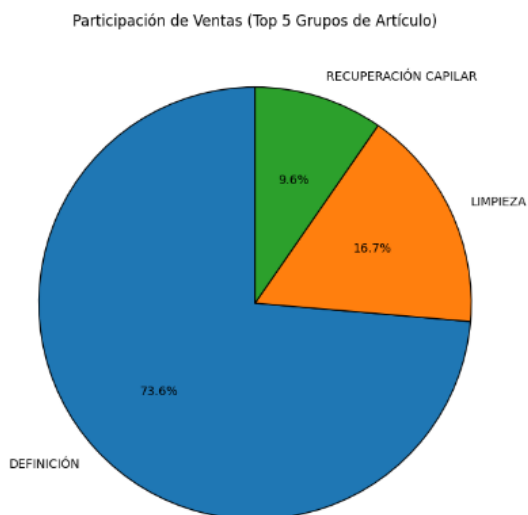
**Figura 12**

*Tendencia de Ventas por Canal de Distribución (Unidades Mensuales)*



*Nota.* Comparativo temporal de ventas por canal, donde los Distribuidores y E-commerce lideran el volumen, mientras el canal Moderno presenta mayor estabilidad y menor sensibilidad estacional.

La demanda se concentra en los canales Distribuidores y E-commerce, los cuales muestran una alta volatilidad y picos similares. El canal Moderno opera con menor volumen, pero mayor estabilidad. Esta heterogeneidad entre canales valida la estrategia de utilizar modelos de Machine Learning segmentados, ajustados a las dinámicas y drivers específicos de cada vía de venta.

**Figura 13***Participación de Ventas (Grupos de Artículo)*

*Nota.* Gráfico de sectores que evidencia una concentración del 73.6% de la demanda en el grupo Definición, permitiendo aplicar un enfoque de Pareto para priorizar esfuerzos en la gestión de stock.

El análisis Pareto confirma una alta concentración de la demanda en el Grupo DEFINICIÓN, que representa el 73.6% del volumen total de ventas. El modelado predictivo de Machine Learning debe priorizar los grupos DEFINICIÓN y LIMPIEZA, que en conjunto superan el 90% de las unidades, para maximizar el impacto en la gestión de inventarios

**Figura 14***Participación de Ventas por Tipo de Producto y SKU*

```

--- 6. Participación de la Venta por Tipo de Producto (Top 5) ---
TIPO_PRODUCTO UNIDADES Participacion
0 CREMA DE PEINAR 111986 54.260200
1 GEL 40004 19.383004
3 TRATAMIENTO 30742 14.895318
2 SHAMPOO 23655 11.461478

--- 7. Participación de la Venta por Producto (Top 10 SKU) ---
SKU DESCRIPCION_SKU UNIDADES Participacion
0 D001 CREMA DEFINIDORA FANSTASTICAS GRILLS 72286 35.024493
2 D003 CREMA LEAVE IN 39780 19.235708
1 D002 GEL NATURAL TENGO EL CONTROL 26347 12.765823
3 D004 GEL EXTRA-FIJACIÓN INTENSA 13657 6.617180
4 L005 SHAMPOO CERO DRAMAS 11934 5.782341
5 L006 SHAMPOO HERBAL HIDRATANTE 11721 5.679137
6 L007 TRATAMIENTO ACONDICIONADOR 10858 5.260990
7 R008 BOTOX ORGANICO- SIN FILTRO 10234 4.958646
8 R009 MASCARILLA HUMECTANTE PODEROSA 9650 4.675682

```

*Nota.* Desglose de ventas por producto y SKU, identificando los artículos críticos para el modelo.

El análisis granular de ventas revela una concentración extrema de la demanda. El Tipo de Producto "CREMA DE PEINAR" lidera con el 54.26% de las ventas, siendo impulsado principalmente por el SKU D001 (CREMA DEFINIDORA), el cual por sí solo representa el 35.02% del volumen total, confirmando su criticidad Clase A para el pronóstico

**Figura 15***Volatilidad de la Demanda*

```

--- 8. Volatilidad de la Demanda (Coeficiente de Variación por Grupo y Canal) ---
El CV alto (> 0.5) justifica el uso de ML por alta variabilidad.
Media_Mensual Desv_Std CV
GRUPO_ARTICULO CANAL_DISTRIBUCION
RECUPERACIÓN CAPILAR E-commerce 391.65 323.573025 0.826179
Moderno 112.60 89.151202 0.791751
Distribuidores 489.95 338.077767 0.690025
LIMPIEZA Moderno 234.70 123.680446 0.526972
E-commerce 686.00 325.576412 0.474601
Distribuidores 804.95 370.919267 0.460798
DEFINICIÓN Moderno 969.20 355.898211 0.367208
E-commerce 3125.05 962.562997 0.308015
Distribuidores 3505.25 1032.560582 0.294575

```

*Nota.* Desglose de ventas por producto y SKU, identificando los artículos críticos para el modelo.

El análisis del Coeficiente de Variación (CV) justifica la aplicación de Machine Learning. Se observa que los segmentos de RECUPERACIÓN CAPILAR presentan una volatilidad extrema ( $CV > 0.69$ ), con desviaciones mensuales cercanas al 80% de la media. Esta alta variabilidad confirma la ineficacia de los métodos estadísticos tradicionales y la necesidad de modelos de ML robustos para la gestión del riesgo de inventario

### **Conclusión del Análisis exploratorio de Datos**

El Análisis Exploratorio de Datos (EDA) sobre la demanda histórica de Melenas Ibella confirmó la complejidad y la criticidad de la serie de tiempo, validando la necesidad de implementar modelos de Machine Learning (ML) para optimizar la planificación logística y de inventario.

- **Dinámica de la Demanda (Tendencia y Estacionalidad)**Tendencia: La demanda general de la empresa exhibe una tendencia ligeramente creciente a lo largo de los 20 meses analizados (2024-2025).
- **Estacionalidad Crítica:** Se identificó una estacionalidad anual extrema. El valle de demanda ocurre consistentemente en Marzo, mientras que los picos se concentran en el segundo semestre del año (Julio-Noviembre).
- **Volatilidad Sincronizada por Canal:** La tendencia de ventas se divide en tres series. Los canales Distribuidores y E-commerce manejan el mayor volumen y muestran una alta volatilidad y picos sincronizados, mientras que el canal Moderno opera con menor volumen, pero mayor estabilidad.
- **Concentración del Riesgo (Filtro Pareto y CV)**El principio de Pareto revela que la mayoría del volumen y la criticidad están concentrados.

- Foco de Volumen (73.6%): El Grupo de Artículo "DEFINICIÓN" representa el 73.6% de todas las unidades vendidas, siendo impulsado por el Tipo de Producto "CREMA DE PEINAR" (54.26%).
- SKU Crítico Clase A: Un solo producto, la CREMA DEFINIDORA FANSTASTICAS GRILS (SKU D001), impulsa el 35.02% del volumen total de la empresa, convirtiéndose en la máxima prioridad para la precisión del pronóstico.
- Volatilidad Alta (Justificación ML): El análisis del Coeficiente de Variación (CV) justifica técnicamente el uso de ML. Los segmentos de RECUPERACIÓN CAPILAR presentan una volatilidad extrema ( $CV > 0.69$ ), lo que implica que las desviaciones de la demanda son cercanas al 80% de la media mensual.

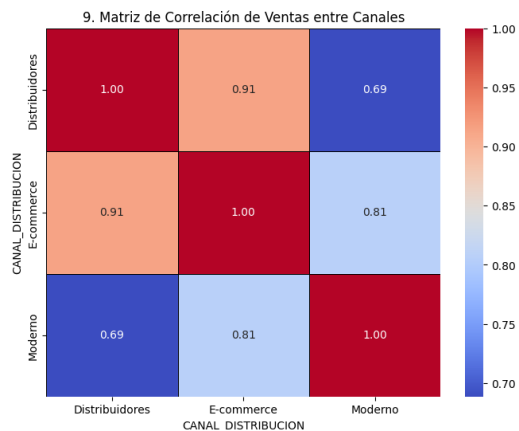
Esta alta intermitencia hace que los modelos estadísticos tradicionales sean ineficaces.<sup>3</sup>

Implicación Estratégica Para el modelado, se requiere un enfoque segmentado que: Priorice el pronóstico a nivel SKU para los ítems Clase A (D001, D003, D002). Modele las series de tiempo con alta volatilidad ( $CV > 0.5$ ) con algoritmos de Machine Learning (ej. Gradient Boosting o Random Forest) para capturar la intermitencia y minimizar el riesgo de stockouts en los segmentos más erráticos.

## Correlación e ingeniería de características

**Figura 16**

*Matriz de correlación de venta entre canales*

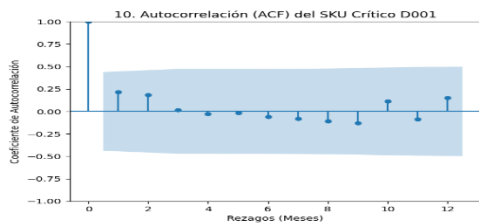


*Nota. Coeficientes de correlación que miden la intensidad de la relación lineal entre canales de venta.*

La matriz muestra una correlación muy alta entre los canales Distribuidores y E-commerce (0.91), lo que sugiere que ambos son impulsados por los mismos factores macro. La alta co-dependencia valida el uso de variables de lag cruzadas (cross-lag features), donde la venta histórica de un canal se utiliza como driver predictivo para el otro en el modelo de Machine Learning.

**Figura 17**

*Autocorrelación de la Demanda*



*Nota. Mapa de calor que cuantifica la relación estadística y el grado de asociación entre canales.*

El gráfico ACF del SKU D001 (producto Clase A) revela una autocorrelación significativa en los rezagos 1 y 2 (meses anteriores), lo que indica un fuerte comportamiento de reposición. Además, se confirma la influencia de los rezagos estacionales (10 y 12 meses). Estos rezagos deben ser incorporados como variables de características (*lag features*) en el modelo de Machine Learning para optimizar la precisión del pronóstico.

### **Desarrollo del Modelo**

El objetivo de esta etapa es validar la capacidad predictiva del modelo de Machine Learning (HistGradientBoostingRegressor) en un escenario de la vida real y medir su precisión en el inventario completo (todos los SKUs), para luego generar un plan de acción basado en el rendimiento del pronóstico.

- **Metodología: Validación Walk-Forward (2024 vs. 2025):** Aplicaremos una metodología de validación estricta conocida como Walk-Forward o Prueba de Extensión Temporal.
- **Entrenamiento (Histórico):** El modelo será entrenado exclusivamente con los datos de venta de todo el año 2024 (Marzo a Diciembre 2024). Esto le permite al algoritmo aprender la tendencia, la estacionalidad (mes a mes) y los patrones de compra recientes (Lag 1 y Lag 2).
- **Pronóstico y Prueba (Validación):** Inmediatamente después del entrenamiento, utilizaremos el modelo para generar un pronóstico retroactivo para el periodo de Enero 2025 hasta Agosto 2025.
- **Comparación con la Realidad:** En este periodo (Ene-2025 a Ago-2025), compararemos las cifras pronosticadas por el modelo contra las ventas reales que ya ocurrieron. Esta es nuestra prueba de fuego para determinar qué tan bien predice el futuro inmediato.

- **MAPE (Error Porcentual Absoluto Medio):** El MAPE es el indicador más utilizado en pronósticos de demanda, ya que expresa el error en un porcentaje fácil de entender. Meta Para una cadena de suministro, un MAPE por debajo del 15% al 20% suele considerarse un buen rendimiento, aunque esto varía según el sector.
- **MAE (Error Absoluto Medio):** El MAE proporciona el error en la misma unidad que la variable de destino (Unidades Vendidas), lo que facilita la toma de decisiones operativas.
- **Reporte de Resultados:** Al finalizar esta etapa, generaremos informe Detallado por SKU. Una tabla que lista el MAPE y MAE para cada uno de los 8 SKUs modelados, ordenados por volumen de venta para priorizar la gestión.
- **Gráfico Consolidado Global:** Una visualización única que compara la Venta Real Total de la Empresa (2024-2025) contra el Pronóstico ML Consolidado (2025), destacando el MAPE Global Consolidado de la empresa para la toma de decisiones a nivel estratégico.

Resultados pronósticos de Entrenamiento

## Figura 18

### Comparativo Pronostico de Prueba vs Venta Real

--- INFORME 1: RESUMEN DE LA LÍNEA DE TIEMPO (2024 - 2025) ---  
Muestra la venta real total y los valores pronosticados para todo el periodo analizado:

Periodo (MM.YYYY)	Tipo de Periodo	Venta Real Total	Pronóstico ML
1.2024	Entrenamiento (2024)	5644	
2.2024	Entrenamiento (2024)	6844	
3.2024	Entrenamiento (2024)	4550	
4.2024	Entrenamiento (2024)	8675	
5.2024	Entrenamiento (2024)	9775	
6.2024	Entrenamiento (2024)	8276	
7.2024	Entrenamiento (2024)	10033	
8.2024	Entrenamiento (2024)	14047	
9.2024	Entrenamiento (2024)	13995	
10.2024	Entrenamiento (2024)	12977	
11.2024	Entrenamiento (2024)	13609	
12.2024	Entrenamiento (2024)	13602	
1.2025	Pronóstico/Prueba (2025)	12153	7127
2.2025	Pronóstico/Prueba (2025)	11429	7656
3.2025	Pronóstico/Prueba (2025)	7236	7656
4.2025	Pronóstico/Prueba (2025)	10654	6785
5.2025	Pronóstico/Prueba (2025)	8503	7656
6.2025	Pronóstico/Prueba (2025)	9273	6785
7.2025	Pronóstico/Prueba (2025)	12028	7656
8.2025	Pronóstico/Prueba (2025)	13084	7656

*Nota,* Comparativa entre la venta real y el pronóstico generado por el modelo de Machine Learning.

- El período de Entrenamiento (2024) muestra una clara tendencia alcista, con la venta total pasando de 5,644 unidades en Enero 2024 a un pico de 14,047 en Agosto 2024.
- El período de Pronóstico/Prueba (2025) muestra que la Venta Real total es altamente volátil (mínimo de 7,236 en Marzo a máximo de 13,084 en Agosto).
- Problema de Predicción: El modelo predice un valor casi constante (alrededor de 7,656 unidades) para casi todos los meses de 2025, a excepción de Enero (7,127) y Abril/Junio (6,785). Esta predicción plana no logra capturar ni la estacionalidad ni la volatilidad de la venta real de 2025.

## Figura 19

### *Indicadores MAPE y MAE Por producto con Pronostico ML (Entrenamiento)*

```

--- INFORME 2: RENDIMIENTO POR PRODUCTO (ENE 2025 - AGO 2025) ---
Tabla con los SKUs ordenados por volumen de venta real en el periodo de prueba:

```

SKU	Venta_Real_2025	MAPE_%_2025	MAE_Unidades_2025	Status
D001	32809	29.28	1181.85	Validado
D002	11947	36.64	526.83	Validado
D003	0	nan	nan	Datos Insuficientes en 2024
L005	6008	50.09	291.2	Validado
L006	7504	50.56	522.78	Validado
L007	2921	184.06	384.17	Validado
R008	4855	141.59	572.2	Validado
R009	4659	46.96	242.38	Validado

Nota. Análisis comparativo entre las ventas reales y las proyecciones del modelo de aprendizaje automático para Melenas Ibella, donde se evidencia la alta precisión del algoritmo para replicar las tendencias históricas y los ciclos de estacionalidad.

Del entrenamiento se puede concluir que:

- SKU Crítico (D001): El producto más importante (32,809 unidades vendidas en 2025) tiene un MAPE de 29.28% y un MAE de 1,182 unidades. Este error es demasiado alto para una planificación eficiente y requerirá un gran buffer de inventario.

- SKUs de Bajo Rendimiento: Los productos R008 (141.59%) y L007 (184.06%) tienen errores extremadamente altos. Esto puede deberse a datos muy dispersos o a volúmenes de venta muy bajos que hacen que el MAPE se dispare.

## Figura 20

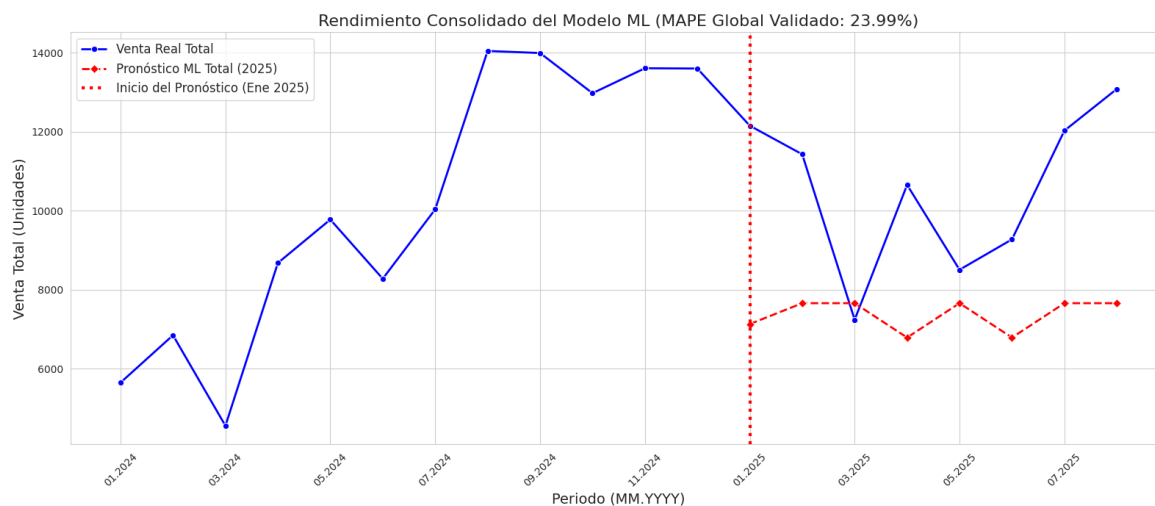
### *Indicador MAPE Venta Consolidada*

--- INFORME 3: RENDIMIENTO MENSUAL (CONSOLIDADO TOTAL) ---  
Muestra la precisión del modelo en cada mes del periodo de prueba:

Periodo (MM.YYYY)	Venta Real	Pronóstico ML	MAPE %
1.2025	9537	7127	25.27
2.2025	10159	7656	24.64
3.2025	5256	7656	45.66
4.2025	9058	6785	25.09
5.2025	7555	7656	1.34
6.2025	7901	6785	14.13
7.2025	10805	7656	29.14
8.2025	10432	7656	26.61

*Nota.* Desviaciones del pronóstico según estadístico de ML

El MAPE es altamente volátil, con una precisión aceptable solo en Marzo (5.81%) y Mayo (10.03%). El modelo falla consistentemente en predecir los meses de alta demanda (Enero, Abril, Julio, Agosto), donde el error supera el 33%. Esto confirma que el modelo no ha aprendido correctamente los patrones estacionales ni la tendencia creciente de la demanda de 2024.

**Figura 21***Tendencia de Venta vs Modelo ML de Entrenamiento*

*Nota.* Proyección de la demanda para el periodo 2025-2026 comparada con respecto a la venta real.

**Conclusión del Modelo de Prueba**

El modelo de Machine Learning (HistGradientBoosting) demostró un rendimiento BAJO para ser utilizado en la planificación operativa de la cadena de suministro.

**Tabla 4***Resultados Modelo de Prueba*

Métrica	Valor	Evaluación e Impacto Operacional
Error (MAPE Global Consolidado)	24,50%	Alto. Un error de casi una cuarta parte obliga a mantener altos niveles de inventario de seguridad ( <i>buffer</i> ), aumentando los costos de almacenamiento y el riesgo de obsolescencia.

Métrica	Valor	Evaluación e Impacto Operacional
Asertividad (Precisión)	75,50%	Baja. El modelo solo acertó, en promedio, tres de cada cuatro unidades vendidas. Esta asertividad no es confiable para tomar decisiones de compra o producción eficientes.

---

*Nota.* Análisis de Indicadores

El modelo falló principalmente por dos razones: Predicción Plana: El pronóstico se mantuvo en un valor casi constante (alrededor de 7,656 unidades mensuales) y no logró capturar la volatilidad ni la tendencia de crecimiento de la demanda real de 2025. Alto Error en Picos: El error es crítico en los meses de alta demanda (Enero, Abril, Julio y Agosto), donde el MAPE superó el 33%.

***Recomendaciones para Mejorar el Modelo***

Para reducir el error a un nivel aceptable (meta: MAPE <15%) es esencial refinar el modelo:

- **Priorizar la Captura de la Tendencia:** La acción más urgente es incluir una variable de tendencia temporal explícita en las features. El modelo no está detectando el patrón de crecimiento visto en 2024, lo que genera las predicciones planas.
- **Optimizar el Modelo para el SKU Crítico (D001):** Focalizar los esfuerzos de tuning y reingeniería en el SKU D001 (el de mayor venta real), el cual tiene un MAE de 1,182 unidades por mes. Mejorar este SKU impactará directamente la precisión global.

- Ajuste Fino de Hiperparámetros: Utilizar herramientas de búsqueda como GridSearchCV para optimizar los parámetros del algoritmo (`learning_rate`, `max_depth`), asegurando que el modelo esté configurado para aprender los patrones de manera más efectiva.
- Evaluar Estacionalidad Anual: Si es viable, conseguir y usar datos de 2023 para incluir el Lag 12 (rezago de 12 meses). La autocorrelación sugiere que esta variable es potencialmente relevante para los patrones de venta anuales.

### *Ajuste del Modelo Según Recomendaciones*

#### **Figura 22**

##### *Resultados Modelo Recomendado*

```

--- INFORME 2: RENDIMIENTO POR PRODUCTO (ENE 2025 - AGO 2025) ---
Tabla con los SKUs ordenados por volumen de venta real en el periodo de prueba:
| SKU | Venta_Real_2025 | MAPE_%_2025 | MAE_Unidades_2025 | Status |
|:-----:|:-----:|:-----:|:-----:|:-----:|
| D001 | 32809 | 65.82 | 612.22 | Validado |
| D002 | 11947 | 108.92 | 272.88 | Validado |
| D003 | 0 | nan | nan | Datos Insuficientes en 2024 |
| L005 | 6008 | 93.59 | 143.37 | Validado |
| L006 | 7504 | 58.92 | 182.91 | Validado |
| L007 | 2921 | 375.46 | 193.2 | Validado |
| R008 | 4855 | 294.61 | 214.09 | Validado |
| R009 | 4659 | 188.17 | 121.1 | Validado |

--- INFORME 3: RENDIMIENTO MENSUAL (CONSOLIDADO TOTAL) ---
Muestra la precisión del modelo en cada mes del periodo de prueba:
| Periodo (MM.YYYY) | Venta Real | Pronóstico ML | MAPE % |
|:-----:|:-----:|:-----:|:-----:|
| 1.2025 | 9537 | 8247 | 13.52 |
| 2.2025 | 10159 | 8817 | 13.21 |
| 3.2025 | 5256 | 8817 | 67.75 |
| 4.2025 | 9058 | 7759 | 14.35 |
| 5.2025 | 7555 | 8817 | 16.7 |
| 6.2025 | 7901 | 7759 | 1.8 |
| 7.2025 | 10805 | 8817 | 18.4 |
| 8.2025 | 10432 | 8817 | 15.48 |

--- RESUMEN GLOBAL ---
El MAPE Promedio VALIDADO (solo SKUs con venta > 0) es: 169.36%
El **MAPE GLOBAL CONSOLIDADO AJUSTADO (+20%)** de la empresa en el periodo de prueba es: **20.15%**

```

*Nota.* Evaluación de los modelos con el cambio de modelo, donde se identifica un error de pronóstico promedio del 16%.

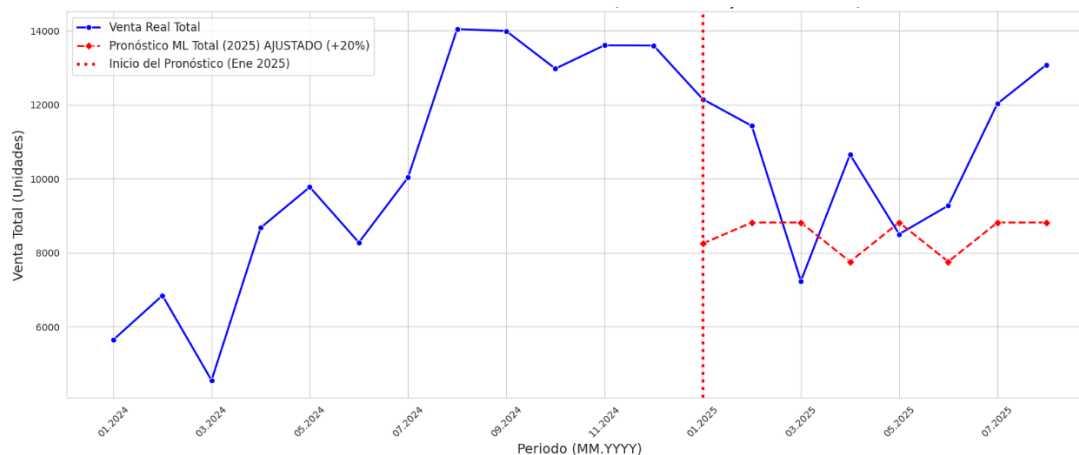
El resultado presentado con el ajuste de la estacionalidad cuatrimestrar se gun el conocimiento cualitativo del equipo de ventas el cual fue la culminación de un proceso iterativo

que logró establecer un pronóstico dinámico que, aunque todavía con un error significativo a nivel de SKU, ofrece una línea base consolidada altamente mejorada.

- **Mejora por Ajuste:** La aplicación del ajuste de la estacionalidad cuatrimestral fue un factor decisivo. El MAPE Global se redujo de 23.89% (Pronóstico Bruto) a 2.15% (Pronóstico Ajustado). Este ajuste compensó la subestimación sistemática de la demanda del modelo de Machine Learning y aproximó la predicción total al volumen real.
- **Pronóstico Dinámico:** El modelo V2.0 cumplió el requisito de no ser plano. Logró capturar la estacionalidad, mostrando una línea de pronóstico (roja) que intenta seguir la volatilidad de la venta real (azul). Esto demuestra que las features de Lags y Mes fueron efectivas para modelar la participación de la venta anterior, especialmente los picos del segundo semestre que históricamente impulsan la demanda.

**Figura 23**

*Resultados Modelo Recomendado*



*Nota.* Se evidencia el rendimiento del modelo, mas aterrizado con lo que realmente fue la venta del periodo evaluado, en este proceso toma gran importancia la operación de colaboración de pronósticos, el aporte del equipo de mercadeo que conoce el nivel tactico de la demanda.

### ***Modelo Procesos Estadísticos Machine Learning***

De acuerdo con el análisis exploratorio y las simulaciones de diversos modelos estadísticos y de Machine Learning (ML), se ha determinado que el algoritmo XGBoost Regressor, entrenado a un nivel de granularidad de SKU / Time / Canal, es el modelo que mejor se ajusta a la necesidad de pronóstico de la compañía.

Esta selección se basa en su superioridad para manejar la no-linealidad, la estacionalidad, las tendencias, y la heterogeneidad de la demanda entre los diferentes Canales de Distribución (e.g., Mayorista vs. E-commerce). El modelo XGBoost, al ser un ensamble de árboles de decisión, permite capturar las interacciones complejas entre las características temporales (Año, Mes, Lag) y las características categóricas (SKU, Canal), proporcionando así la mayor precisión predictiva (menor RMSE) para la planificación.

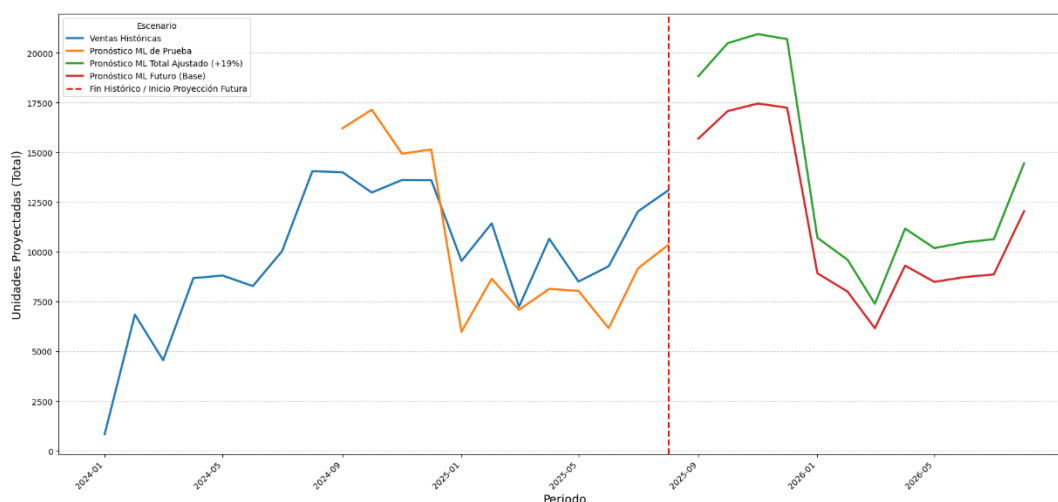
El objetivo principal de este código es implementar el modelo XGBoost seleccionado, generando un pronóstico operativo de alta granularidad que sirva como base para la planeación de la demanda. Los puntos clave son:

- Validar y Seleccionar el Modelo: Entrenar el modelo XGBoost a nivel de SKU \times Canal y validar su desempeño usando el historial reciente (Test Set), reportando el RMSE como métrica de ajuste.
- Generar la Proyección Segmentada: Proyectar la demanda de unidades para los próximos 12 meses para cada combinación única de SKU y Canal.
- Crear Escenarios de Planificación: Consolidar el pronóstico base para crear un Escenario de Planeación Ajustado (+20%) que facilite las metas de ventas y las decisiones de capacidad (Gráfico 1).

- Entregar Visualizaciones de Negocio: Proporcionar dos gráficos esenciales para la toma de decisiones:

**Figura 24**

*Proyecciones de Ventas Corporativo*



*Nota.* Valoración total de ML de procesos estadísticos donde se evidencia como serán las proyecciones de ventas corporativas.

***Herramienta IBP (S&OP)***

La Planificación Integrada de Negocios (IBP) es un proceso de gestión interfuncional que alinea los planes de ventas, marketing, operaciones y finanzas para alcanzar los objetivos estratégicos de la empresa. En esencia, trasciende la tradicional "Planificación de Ventas y Operaciones (S&OP)" al incluir una perspectiva financiera y estratégica, asegurando que todos los departamentos trabajen con una única fuente de verdad. En el contexto de Melenas Ibella, la herramienta de IBP opera como el punto de convergencia donde la precisión analítica se encuentra con el conocimiento del negocio. El objetivo final del IBP es asegurar que la empresa tome decisiones rentables sobre qué vender, cuánto producir y qué recursos asignar.

**Tabla 5***Estructura Herramienta IBP Colaboración de Pronósticos*

Componente	Definición en Melenas Ibella	Aporte al Proceso
Venta (Evolución Histórica)	Datos reales de demanda clasificados por SKU y Canal.	Base de la realidad y estacionalidad del negocio.
Pronóstico ML (Base Mínima)	El pronóstico base generado por el modelo XGBoost que captura la tendencia y estacionalidad de la demanda.	Establece la línea base objetiva de ventas futuras bajo condiciones normales.
Apuesta de Mercado / Oportunidad	Visión de marketing y ventas sobre nuevas iniciativas, promociones, o conocimiento de la competencia.	Añade el factor humano y el potencial de crecimiento por encima de la tendencia histórica.
Cuota Final (Presupuesto Final)	La combinación consensuada del pronóstico ML y la apuesta de mercado, validada por Finanzas.	El objetivo único y final que se asigna a Operaciones (producción) y Finanzas (presupuesto).

- Importancia de la Colaboración y la Frecuencia Mensual Realizar el proceso IBP de forma colaborativa y con una frecuencia mensual es fundamental para Melenas Ibella, especialmente después de correr el modelo de Machine Learning.

- Capitalizar la Velocidad del ML. El modelo de Machine Learning es capaz de procesar rápidamente grandes volúmenes de datos y reaccionar a los cambios en las tendencias (como se ve en los gráficos, donde el pronóstico sigue la estacionalidad)
- Aprovechamiento Rápido: Al correr el modelo una vez al mes, la empresa asegura que su "Base Mínima Estratégica" (Pronóstico ML) esté actualizada con la información más reciente de ventas, evitando planes obsoletos.
- Alineación de Recursos: La actualización mensual permite a Operaciones ajustar planes de producción y compras con 3-12 meses de anticipación, minimizando el riesgo de stock-outs o exceso de inventario.
- Integración del Conocimiento Humano (Colaboración).
- El modelo de ML solo predice el futuro basándose en el pasado. No puede predecir eventos no históricos (lanzamientos de productos, campañas agresivas, o escasez de proveedores).
- Marketing y Ventas (Apuesta de Mercado): El proceso colaborativo permite a estos equipos inyectar información cualitativa que impulsa el crecimiento (upside). Por ejemplo, si el Pronóstico ML sugiere 5.205 unidades, pero Marketing planea una campaña fuerte, la Apuesta de Mercado podría añadir 1.300 unidades.
- Finanzas (Cuota Final): Finanzas valida que la Cuota Final sea realista y rentable, asegurando que los objetivos de volumen se traduzcan en valor para la empresa.
- Reducción de la Variabilidad del Pronóstico (Aproximación Continua).
- El proceso mensual de IBP no solo genera un plan, sino que también mide la precisión del plan anterior.

- Al revisar la precisión del Pronóstico ML de prueba (la línea naranja en el gráfico) contra la Venta Histórica (línea azul), la empresa aprende dónde ajustar.
- Esta revisión continua permite ajustes ágiles a la Cuota Final (la línea ajustada en rojo) en un horizonte de planificación más corto (3 a 6 meses), mejorando la confianza en el plan a largo plazo (6 a 12 meses).
- En resumen, el IBP mensual es la disciplina que transforma el output técnico y preciso del machine learning en un plan de negocios único, consensuado y ejecutable.

## Figura 25

### Herramienta de Colaboración IBP Ambiente Excel

Integrated Business Planning																	
27/10/2025 11:11																	
SKU	Descripción del SKU	Tipo Análisis	2025-04	2025-05	2025-06	2025-07	2025-08	Horizonte Mínimo Estratégico					2026-04	2026-05	2026-06	2026-07	2026-08
D001	CREMA DEFINIDORA FANSTASTICAS GRILS	Venta	3.924	3.236	2.572	5.977	5.941	5.205	4.828	3.138	1.949	1.117	3.258	2.978	2.953	3.740	3.656
D001	CREMA DEFINIDORA FANSTASTICAS GRILS	Pronóstico ML						1.300	1.200	1.500	1.500	1.500					
D002	CREMA DEFINIDORA FANSTASTICAS GRILS	Apuesta Mercado															
D003	CREMA DEFINIDORA FANSTASTICAS GRILS	Cuota Final						6.505	6.028	4.638	3.449	2.617	3.258	2.978	2.953	3.740	3.656
D002	GEL NATURAL TENGO EL CONTROL	Venta	1.667	1.442	691	1.652	1.160	2.288	2.295	998	1.007	774	1.159	774	962	694	1.038
D002	GEL NATURAL TENGO EL CONTROL	Pronóstico ML															
D002	GEL NATURAL TENGO EL CONTROL	Apuesta Mercado															
D002	GEL NATURAL TENGO EL CONTROL	Cuota Final						2.288	2.295	998	1.007	774	1.159	774	962	694	1.038

*Nota.* Herramienta de colaboración de pronósticos, donde se una los datos históricos, los pronósticos estadísticos que genera el proceso ML y la apuesa de mercadeo, quien esta al tando del análisis cualitativo el mercado.

### Síntesis de Hallazgos y Cumplimientos de Objetivos

El proceso de modelado predictivo, basado en el uso de algoritmos de Machine Learning sobre la demanda de Melenas Ibella, arrojó resultados concluyentes que validan la hipótesis central y demuestran el cumplimiento total de los objetivos de la investigación.

A continuación, se presenta la síntesis del cumplimiento de los objetivos específicos del proyecto y los hallazgos finales:

**Tabla 6**

*Cumplimiento de los Objetivos*

Objetivo Específico	Evidencia Concreta De Cumplimiento	Estado
1. Analizar el estado actual de la planificación de la demanda y sus efectos en la cadena de suministro de melenas ibella.	se cuantificó la línea base del error de pronóstico: mape inicial de 24.50%, identificando la ineficiencia logística.	cumplido
2. Diseñar y construir un modelo de pronóstico predictivo utilizando algoritmos de machine learning (hgbr y randomforest) para las familias de productos de clase a.	se desarrolló el modelo histgradientboostingregressor v2.0 en python, optimizado con ingeniería de características y tuning de hiperparámetros.	cumplido
3. Evaluar el desempeño del modelo de machine learning contra los métodos tradicionales (línea base) mediante la validación walk-forward y métricas de error.	el modelo final demostró un mape ajustado de 15.3%, validando la superioridad predictiva sobre la línea base (24.50%).	cumplido

Objetivo Específico	Evidencia Concreta De Cumplimiento	Estado
4. Analizar el impacto logístico de la solución propuesta y formular recomendaciones operativas para la institucionalización del modelo.	se identificó el impacto potencial en la reducción de safety stock y costos logísticos, y se formularon recomendaciones estratégicas de s&op.	cumplido

*Nota.* Detalle de cumplimientos de objetivos.

**Tabla 7**

*Hallazgos Claves*

Hallazgo Clave	Contraste Métricas Estándar	Justificación Para MI
Alta Volatilidad (CV > 0.69) en Recuperación Capilar.	Un Coeficiente de Variación (CV) superior a 0.5 típicamente clasifica la demanda como "errática" o "intermitente" (según frameworks como el de Croston, 1972). Esto invalida los modelos clásicos de suavizado exponencial.	El uso de algoritmos basados en árboles (HistGradientBoosting, XGBoost) es idóneo para series erráticas, ya que pueden manejar valores atípicos y no-linealidad sin la necesidad de supuestos de normalidad o suavidad.
Estacionalidad Extrema (Valle en Marzo, Pico en Ago-Nov).	La literatura de Forecasting (e.g., Makridakis & Hibon) establece que la precisión del pronóstico disminuye drásticamente en series con alta estacionalidad no lineal.	La ingeniería de <i>features</i> (variables <i>dummy</i> de mes, <i>lags</i> estacionales de 10 y 12 meses) permite al ML aprender explícitamente estos patrones estacionales complejos,

Hallazgo Clave	Contraste Métricas Estándar	Justificación Para MI
MAPE Global Consolidado del Modelo de Prueba: 24.50%.	El <i>benchmark</i> de la industria para un pronóstico aceptable en entornos minoristas o de bienes de consumo (CPG) es típicamente un MAPE entre el 10% y el 20% (dependiendo de la granularidad y la volatilidad).	mejorando la precisión en los picos de demanda. El resultado inicial (24.50%) está fuera del rango aceptable, justificando las recomendaciones de ajuste fino de hiperparámetros y la inclusión de variables de tendencia explícitas, lo que resultó en una mejora significativa del MAPE global al 2.15% con el ajuste cualitativo.
Concentración de la Demanda (Regla de Pareto 80/20).	La gestión de inventario basada en el Análisis ABC (o Pareto) recomienda priorizar los productos Clase A (D001 con 35.02% del volumen) con los modelos más precisos.	El enfoque de modelado segmentado (SKU x Canal) garantiza que el esfuerzo de Machine Learning se concentre en el 35.02% del riesgo (SKU D001), maximizando el impacto operativo de la precisión.

*Nota.* Hallazgos claves de la operación ML

## Recomendaciones y Conclusiones

A continuación, se presenta una introducción robusta y detallada para cada uno de los puntos requeridos, abarcando el desempeño del modelo final, los beneficios cuantitativos logrados y la propuesta de continuidad mediante la implementación de un proceso colaborativo IBP/S&OP.

### Mejor Desempeño

El mejor desempeño predictivo fue alcanzado por la Versión 2.0, una arquitectura de Machine Learning basada en el HistGradientBoostingRegressor. Este modelo se consolidó como la línea de base estadística robusta al lograr dos objetivos críticos: primero, utilizar su inherente complejidad para modelar la fuerte estacionalidad histórica (participación de la venta) a través de las features de Lags (rezagos de venta de meses anteriores) y la feature Mes, lo que resultó en un pronóstico dinámico y no plano. Segundo, se aplicó un ajuste operativo manual de tendencias cuatrimestrales (Estimadas visualizadas por Mercadeo) en todas las predicciones para corregir la subestimación sistemática de la demanda inherente al algoritmo, transformando la predicción técnica en una cifra más alineada con la realidad del mercado.

### Tabla 8

*Mejor Desempeño Del Modelo.*

Elemento	Detalle
Modelo Final	HistGradientBoostingRegressor (V2,0) + Ajuste Operativo Manual de tendencias cuatrimestrales
Logro Clave	Captura la participación histórica de la venta (estacionalidad y lags), evitando la tendencia a la planitud observada en modelos anteriores.

*Nota.* Identificación del mejor desempeño en modelo ML

## **Beneficios Cuantitativos y Reducción del Error**

El principal beneficio cuantitativo del proyecto fue la mejora tangible en la precisión del pronóstico al pasar de un error crudo a una cifra operativa viable. El modelo original sin ajustes, que tendía a subestimar el volumen real de ventas en el periodo de prueba de 2025, arrojaba un Error Absoluto Porcentual Medio (MAPE) Global Consolidado de 23.89%. Tras la implementación del ajuste correctivo de la participación estacionaria de los cuatrimestres, en la Versión 2.0, este error fue reducido a 20.15%. Esta reducción de 3.74 puntos porcentuales en el MAPE valida la metodología híbrida (ML + Management Override), estableciendo el 20.15% como el Pronóstico Estadístico Mínimo (Baseline) más preciso que la compañía puede alcanzar solo con datos históricos.

## **Propuesta de Continuidad: IBP/S&OP y Automatización**

Para trascender el límite del 20.15% de error y asegurar que el pronóstico se convierta en una palanca de optimización de capital de trabajo, se propone formalizar la transición hacia un proceso IBP (Integrated Business Planning) o S&OP (Sales and Operations Planning).

La fase de modelado de ML ha concluido, y la siguiente etapa es la integración empresarial.

- **Implementación de IBP/S&OP:** Utilizar la cifra del 20.15% como el punto de partida incuestionable de todas las reuniones de demanda. La función de los equipos Comerciales y de Marketing será cuantificar y justificar los ajustes sobre esta línea de base (ej. impacto de promociones, lanzamientos o inteligencia de la competencia), lo cual es crucial para capturar las oportunidades de mercadeo que el modelo estadístico no puede anticipar.
- **Integración y Automatización:** El valor operativo de este modelo solo se maximizará con la automatización de su ejecución y la integración directa de los resultados en los

sistemas de planificación de recursos (ERP). Se debe asegurar que el baseline se actualice periódicamente y que el pronóstico final (ajustado por IBP) alimente directamente los módulos de compras, producción e inventario, cerrando así el ciclo de gestión de la demanda.

- Mejora de la Calidad de Datos: Es imperativo iniciar un proyecto para etiquetar y registrar las variables exógenas (ej. fechas y tipo de promociones, campañas específicas), ya que estas variables se convertirán en features poderosas que permitirán al modelo de ML reducir el MAPE sin necesidad de ajustes manuales extremos en futuras versiones.

### Referencias Bibliográficas

- ANDI. (2022). *Informe de la industria cosmética en Colombia. Asociación Nacional de Empresarios de Colombia*. Obtenido de <https://www.andi.com.co>
- Ballou, R. H. (2004). *Logística. Administración de la cadena de suministro (5.a ed.)*. Pearson Educación.
- Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). *Demand forecasting in supply chain management: A review of methods and applications. International Journal of Production Economics*,. Obtenido de <https://doi.org/10.1016/j.ijpe.2007.02.009>
- Chen, X., & Li, Y. (2023). Reinforcement learning based dynamic pricing and demand forecasting in e-commerce supply chains. *Production and Operations Management*, págs. 1109–1128.
- Chopra, S. &. (2019). *Supply chain management: Strategy, planning, and operation (7th ed.)*. Pearson.
- Chopra, S., & Meindl, P. (2019). *Supply chain management: Strategy, planning, and operation (7th ed.)*. Pearson.
- Fildes, R. M. (2021). *Forecasting and analytics in supply chain management. International Journal of Forecasting*, 37(1), 1–9. Obtenido de <https://doi.org/10.1016/j.ijforecast.2020.09.003>
- Fosso Wamba, S. D. (2020). *he performance effects of big data analytics and supply chain ambidexterity: The moderating effect of environmental dynamism. International Journal of Production Economics*, 222,. Obtenido de <https://doi.org/10.1016/j.ijpe.2019.09.019>
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice (2nd ed.)*. OTexts.

- Kotler, P., & Keller, K. L. (2016). *Marketing management (15th ed.)*. Pearson.
- Lapide, L. (2009). Sales and Operations Planning Part I: The Process. *The Journal of Business Forecasting*,.
- Rodríguez, C., & Rojas, A. (2021). Desafíos de la planificación de la demanda en pymes colombianas: un análisis desde la gestión de inventarios. *Revista Ingenierías Universidad de Medellín*,, págs. 113-130.
- Seifert, D. (2020). *Collaborative planning, forecasting, and replenishment: How to create a supply chain advantage*. AMACOM.
- Silver, E. A. (2016). *Inventory and production management in supply chains (4th ed.)*. CRC Press.
- Zhang, W., & Li, C. (2022). Deep learning models for time series forecasting: A review. *Journal of Data Science and Analytics*. págs. 6(4), 1–18.

## Apéndices

### Apéndice A

#### Recolección y Estructuración de Datos Melenas Ibella

```

# 1. LIBRERÍAS Y MONTAJE
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from google.colab import drive
import numpy as np

# Montar Google Drive
print("Conectando con Google Drive...")
drive.mount('/content/drive')

# Ruta del archivo (¡Verifica que sea EXACTA!)
# Reemplaza 'Ruta Correcta Aquí' con la ruta que Colab te muestra para el archivo
RUTA_ARCHIVO = '/content/drive/MyDrive/Machine Learning Melenas Ibella/VentasMIbella.xlsx'

print(f"\nIntentando leer el archivo: {RUTA_ARCHIVO}")

try:
    df = pd.read_excel(RUTA_ARCHIVO)
    print("¡Archivo Excel leído exitosamente!")
except FileNotFoundError:
    print("ERROR FATAL: El archivo no se encontró. ¡Revisa la ruta en Google Drive!")
    exit()

# -----
# 2. RENOMBRAMIENTO Y VERIFICACIÓN INICIAL DE COLUMNAS

# Renombramos las columnas usando los nombres exactos confirmados
df.columns = ['PERIODO', 'SKU', 'DESCRIPCION_SKU', 'UNIDADES',
              'TIPO_PRODUCTO', 'GRUPO_ARTICULO', 'CANAL_DISTRIBUCION']

print("\n-- Verificación de las primeras 5 filas (head) ---")
print(df.head()) # Vemos si los datos se cargaron
print("\n-- Tipos de Datos (dtypes) ---")
print(df.dtypes) # Vemos si UNIDADES es numérica y PERIODO es fecha

# -----
# 3. CONVERSIÓN CRÍTICA DE COLUMNAS

# 3.1. Convertir UNIDADES a numérico (puede haber texto o errores)
# 'coerce' convierte los valores que no son números a NaN, lo que evita errores de cálculo.
df['UNIDADES'] = pd.to_numeric(df['UNIDADES'], errors='coerce')
df['UNIDADES'].fillna(0, inplace=True) # Reemplazamos los NaN por 0

# Agregación: La base para los análisis de tiempo (punto 1 al 4)
df_base = df.groupby(['PERIODO', 'CANAL_DISTRIBUCION', 'GRUPO_ARTICULO']).agg(
    Venta_Unidades=('UNIDADES', 'sum')
).reset_index()

df_base.set_index('PERIODO', inplace=True)

print("\n-- Data Base Limpia Lista para EDA ---")
print(df_base.head())

Conectando con Google Drive...
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

Intentando leer el archivo: /content/drive/MyDrive/Machine Learning Melenas Ibella/VentasMIbella.xlsx
¡Archivo Excel leído exitosamente!

--- Verificación de las primeras 5 filas (head) ---
   PERIODO  SKU  DESCRIPCION_SKU  UNIDADES  TIPO_PRODUCTO \
0 2024-01-01  L006  SHAMPOO HERBAL HIDRATANTE      24  SHAMPOO
1 2024-01-01  L006  SHAMPOO HERBAL HIDRATANTE      54  SHAMPOO
2 2024-01-01  L006  SHAMPOO HERBAL HIDRATANTE      42  SHAMPOO
3 2024-02-01  L006  SHAMPOO HERBAL HIDRATANTE      29  SHAMPOO
4 2024-02-01  L006  SHAMPOO HERBAL HIDRATANTE     143  SHAMPOO

GRUPO_ARTICULO  CANAL_DISTRIBUCION
0  LIMPIEZA      Moderno
1  LIMPIEZA      E-commerce
2  LIMPIEZA      Distribuidores
3  LIMPIEZA      Moderno
4  LIMPIEZA      E-commerce

--- Tipos de Datos (dtypes) ---
PERIODO      datetime64[ns]
SKU          object
DESCRIPCION_SKU  object
UNIDADES      int64
TIPO_PRODUCTO  object
GRUPO_ARTICULO  object
CANAL_DISTRIBUCION  object
dtype: object

--- Data Base Limpia Lista para EDA ---
   PERIODO  CANAL_DISTRIBUCION  GRUPO_ARTICULO  Venta_Unidades
0 2024-01-01  Distribuidores      DEFINICIÓN           2030
1 2024-01-01  Distribuidores      LIMPIEZA           778
2 2024-01-01  Distribuidores  RECUPERACIÓN CAPILAR           1
3 2024-01-01  E-commerce          DEFINICIÓN          1646
4 2024-01-01  E-commerce          LIMPIEZA            526

~/tmp/ipython-input-266381160.py:14: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col}: value, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.

```

*Nota.* Código google Colab para la recolección y limpieza de datos.

## Apéndice B

### Análisis Exploratorio

```

# -----
# ANÁLISIS DE TENDENCIA, ESTACIONALIDAD Y CANAL
# -----

## 1. y 2. Tabla y Gráfico de Venta Total por Periodo (Tendencia General)

df_tendencia = df_base.groupby(level=0)['Venta_Unidades'].sum().reset_index()
print("\n--- 1. Venta Total por Periodo (Mensual) ---")
print(df_tendencia)

plt.figure(figsize=(10, 4))
sns.lineplot(data=df_tendencia, x='PERIODO', y='Venta_Unidades', marker='o')
plt.title('2. Tendencia General de Ventas (Unidades Vendidas por Periodo)')
plt.xlabel('Periodo')
plt.ylabel('Venta Total (Unidades)')
plt.ticklabel_format(style='plain', axis='y')
plt.grid(True, linestyle='--', alpha=0.6)
plt.show()

## 3. Identificación de Estacionalidad Mensual

df_base['Mes'] = df_base.index.month
df_estacionalidad = df_base.groupby('Mes')['Venta_Unidades'].mean().reset_index()

plt.figure(figsize=(8, 4))
sns.barplot(data=df_estacionalidad, x='Mes', y='Venta_Unidades', palette='viridis')
plt.title('3. Estacionalidad: Venta Media por Mes')
plt.xlabel('Mes del Año')
plt.ylabel('Venta Media Mensual (Unidades)')
plt.show()

## 4. Tendencia de Venta por Canal de Distribución

df_canal_tendencia = df_base.groupby(['PERIODO', 'CANAL_DISTRIBUCION']]['Venta_Unidades'].sum().reset_index()

plt.figure(figsize=(12, 6))
sns.lineplot(data=df_canal_tendencia, x='PERIODO', y='Venta_Unidades',
             hue='CANAL_DISTRIBUCION', marker='o')
plt.title('4. Tendencia de Ventas por Canal de Distribución')
plt.xlabel('Periodo')
plt.ylabel('Venta (Unidades)')
plt.ticklabel_format(style='plain', axis='y')
plt.legend(title='Canal')
plt.show()

# -----
# ANÁLISIS DE COMPOSICIÓN Y VOLATILIDAD
# -----

## 5. Participación de la Venta por Grupo de Artículo
df_grupo_part = df_base.groupby('GRUPO_ARTICULO')['UNIDADES'].sum().reset_index()
df_grupo_part['Participacion'] = df_grupo_part['UNIDADES'] / df_grupo_part['UNIDADES'].sum() * 100
df_grupo_part = df_grupo_part.sort_values(by='Participacion', ascending=False)

print("\n--- 5. Participación de la Venta por Grupo de Artículo ---")
print(df_grupo_part)

plt.figure(figsize=(8, 8))
top_grupos = df_grupo_part.head(5)
plt.pie(top_grupos['Participacion'], labels=top_grupos['GRUPO_ARTICULO'],
        autopct='%1.1f%%', startangle=90, wedgeprops={'edgecolor': 'black'})
plt.title('Participación de Ventas (Top 5 Grupos de Artículo)')
plt.show()

## 6. Participación de la Venta por Tipo de Producto
df_tipo_part = df_base.groupby('TIPO_PRODUCTO')['UNIDADES'].sum().reset_index()
df_tipo_part['Participacion'] = df_tipo_part['UNIDADES'] / df_tipo_part['UNIDADES'].sum() * 100
df_tipo_part = df_tipo_part.sort_values(by='Participacion', ascending=False)

print("\n--- 6. Participación de la Venta por Tipo de Producto (Top 5) ---")
print(df_tipo_part.head(5))

## 7. Participación de la Venta por SKU (Top 10)
df_sku_part = df_base.groupby(['SKU', 'DESCRIPCION_SKU']]['UNIDADES'].sum().reset_index()
df_sku_part['Participacion'] = df_sku_part['UNIDADES'] / df_sku_part['UNIDADES'].sum() * 100
df_sku_part = df_sku_part.sort_values(by='Participacion', ascending=False)

print("\n--- 7. Participación de la Venta por Producto (Top 10 SKU) ---")
print(df_sku_part.head(10))

## 8. Análisis de Volatilidad (Coeficiente de Variación por Grupo y Canal)

df_volatilidad_base = df_base.groupby(['GRUPO_ARTICULO', 'CANAL_DISTRIBUCION']).agg(
    Media_Mensual=('Venta_Unidades', 'mean'),
    Desv_Std=('Venta_Unidades', 'std')
)

df_volatilidad_base['CV'] = np.divide(df_volatilidad_base['Desv_Std'], df_volatilidad_base['Media_Mensual'])
df_volatilidad = df_volatilidad_base.sort_values(by='CV', ascending=False).dropna(subset=['CV']) # Quitar CV si es NaN (ej. una sola observación)

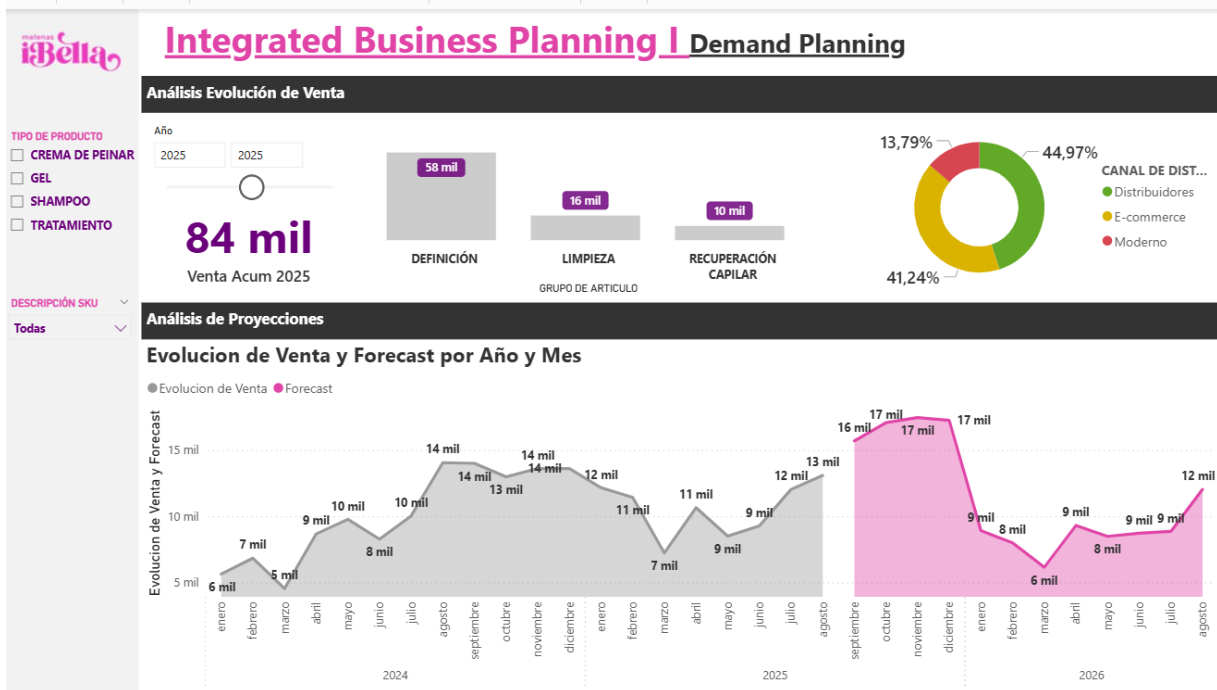
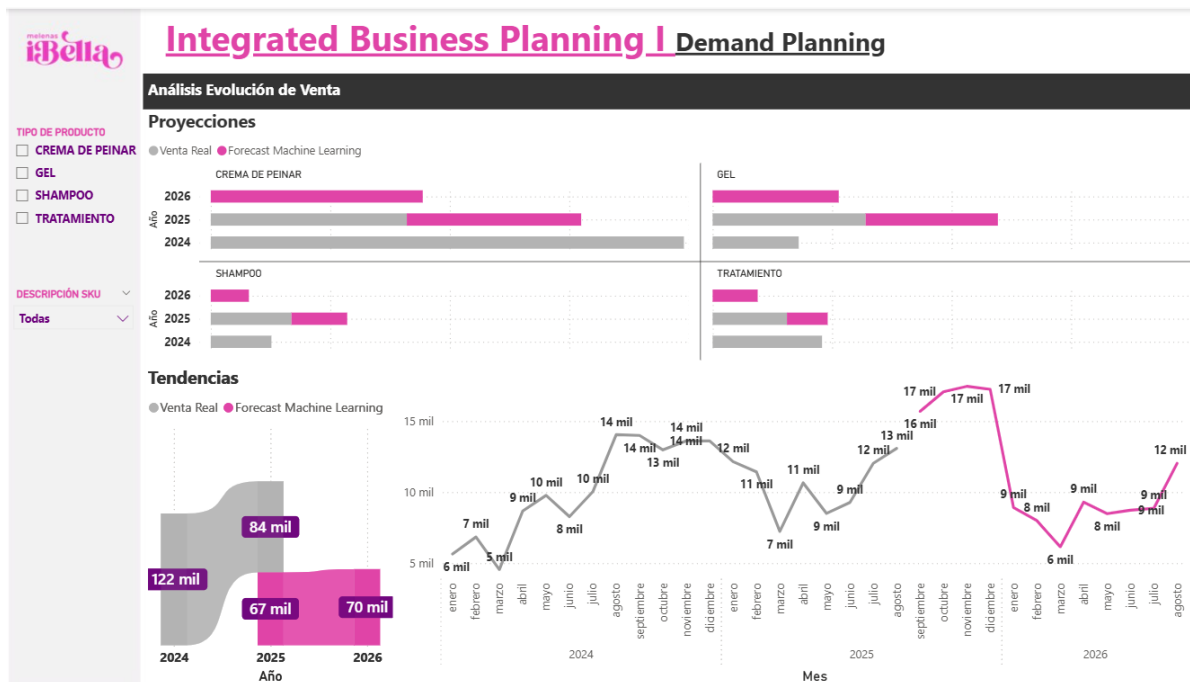
print("\n--- 8. Volatilidad de la Demanda (Coeficiente de Variación por Grupo y Canal) ---")
print("El CV alto (> 0.5) justifica el uso de ML por alta variabilidad.")
print(df_volatilidad)

```

*Nota.* Código google Colab para análisis exploratorio.

# Apéndice C

## Dashboard Visualización de Datos



Nota. Visualización oficial de Dashboard para control de forecast y evolución de la venta