

***Optimización Estratégica del Inventario de Cruzber
mediante Analítica Avanzada: Predicción de
Demanda y Gestión Inteligente del Stock***

**TRABAJO DE FINAL DE MÁSTER
CURSO 2024-2025**

Autor: Alex Fau Ridao

PO: Antonio San Miguel

Fecha de entrega: 24/06/2025

MÁSTER EN ANÁLISIS DE DATOS & IA

ISDI DIGITALENT GROUP

Resumen

En un entorno de mercado globalizado y competitivo, la eficiencia de la cadena de suministro se ha convertido en un diferenciador estratégico clave. Este Trabajo de Fin de Máster aborda este desafío en el contexto de Cruzber, una empresa manufacturera líder en el sector de accesorios para automóviles, que se encuentra en un punto de inflexión estratégico debido a su expansión internacional y al desarrollo de su canal B2C. El problema fundamental investigado es la suboptimización de su capital de trabajo y nivel de servicio, derivada de una gestión de inventario reactiva y no fundamentada en la predicción de la demanda. El objetivo principal de este proyecto es diseñar, desarrollar y evaluar una solución de analítica avanzada que transforme la gestión de inventario de un centro de coste en una ventaja competitiva. Para ello, se implementó un modelo de predicción de demanda utilizando un enfoque de machine learning (MLForecast con LightGBM), entrenado con un histórico de más de 2.700 productos. El modelo permite predecir las ventas futuras con una notable precisión (WAPE del 36%), considerando la estacionalidad, las tendencias y el comportamiento diverso de los artículos.

A partir de estas predicciones, se ha definido una metodología para calcular niveles de stock óptimos (mínimos y máximos) ajustados dinámicamente a la demanda y al nivel de servicio requerido. Los resultados demuestran la viabilidad y el alto impacto del modelo. Se ha desarrollado un dashboard operativo en Tableau que ofrece una visión integral y proactiva del inventario, facilitando la toma de decisiones. La solución proyecta una reducción del coste por exceso de stock de 124.500 € y la liberación de 200.000 € de capital inmovilizado.

Este trabajo concluye que la adopción de analítica avanzada es fundamental para que Cruzber pueda afrontar sus retos de crecimiento de manera sostenible, mejorando su eficiencia operativa, su rentabilidad y su capacidad de respuesta en un mercado global, competitivo y digitalizado.

Palabras clave: predicción de demanda, machine learning, optimización de inventario, series temporales, gestión de stock, cadena de suministro, MLForecast, LightGBM.

Índice de contenidos

1. Introducción	4
1.1 Resumen ejecutivo del trabajo	4
1.2 Metodología del trabajo	4
1.3 Recursos y medios empleados	5
2. Análisis de contexto	6
2.1 Análisis interno (empresa)	6
2.2 Análisis externo (sector)	6
2.3 Tendencias del sector	6
3. Solución propuesta	8
3.1 Propósito de la solución	8
3.2 Objetivos de la solución	8
3.3 Descripción de la solución	9
4. Impacto en el negocio	16
4.1 Modelo de negocio y flujos de monetización	16
4.2 Análisis financiero	17
5. Conclusiones	18
5.1 Contribuciones realizadas	18
5.2 Problemas encontrados	19
5.3 Opiniones personales	19
Bibliografía	20
Anexos	21

1. Introducción

1.1 Resumen ejecutivo del trabajo

Este trabajo aborda un imperativo estratégico para Cruzber: la transformación de su cadena de suministro para sostener su crecimiento. En su actual fase de expansión a mercados como Norteamérica y el desarrollo del canal B2C, la compañía se enfrenta a una volatilidad de la demanda que sus sistemas de gestión de inventario tradicionales no pueden abordar eficazmente. Esta investigación presenta una solución de analítica avanzada diseñada para convertir la gestión de stock de un centro de coste reactivo a una ventaja competitiva proactiva. Mediante la implementación de un modelo de machine learning para la predicción de la demanda, se ha desarrollado una metodología que permite a Cruzber optimizar sus niveles de inventario, reducir costes operativos y maximizar la disponibilidad de producto. El proyecto no solo valida técnicamente la solución, sino que demuestra su alto retorno de la inversión, con un potencial de ahorro de costes de más de 124.000 € y la liberación de 200.000 € de capital circulante. En definitiva, este trabajo proporciona a Cruzber una herramienta estratégica, escalable y basada en datos, esencial para fortalecer su rentabilidad y asegurar su éxito competitivo en el escenario global.

1.2 Metodología del trabajo

La metodología del proyecto se ha estructurado siguiendo un flujo de trabajo estándar en ciencia de datos, adaptado para asegurar la obtención de resultados de negocio. Las ocho fases desarrolladas son:

1. Limpieza y Estructuración: Preparación de los datos brutos para el análisis. Aplicación de clustering para identificar los productos más relevantes.
2. Selección del Modelo: Evaluación y elección del algoritmo de machine learning más adecuado.
3. Preparación del Dataset: Ingeniería de características para el entrenamiento.
4. Entrenamiento y Predicción: Construcción del modelo y generación de pronósticos.
5. Evaluación del Rendimiento: Medición de la precisión del modelo.
6. Cálculo de Stock Óptimo: Traducción de las predicciones en niveles de inventario.
7. Arquitectura de la Solución: Diseño del flujo de datos tecnológico.
8. Dashboard Operativo: Creación de la herramienta de visualización final.

La consecución de los objetivos de este proyecto ha requerido una metodología híbrida que combina el rigor de los marcos de trabajo estándar en ciencia de datos (como CRISP-DM^[1]) con un enfoque pragmático orientado a los resultados de negocio. La elección de este enfoque estructurado en ocho fases ha sido

deliberada, buscando garantizar no solo la validez técnica de los modelos, sino también la relevancia, replicabilidad y la alineación de la solución final con las necesidades de los stakeholders de Cruzber. Cada fase, desde la limpieza de datos hasta la creación del dashboard, se ha concebido como un eslabón necesario para construir una solución integral, donde la precisión del modelo se traduce directamente en una política de inventario optimizada y en decisiones de negocio informadas. Esta metodología asegura que el proyecto trascienda el ejercicio académico para convertirse en un activo tangible y mantenible para la organización.

1.3 Recursos y medios empleados

La materialización de este proyecto ha sido posible gracias a la selección estratégica de un conjunto de herramientas y tecnologías de nivel empresarial, garantizando que la solución sea no solo funcional, sino también escalable y robusta.

- **Lenguajes y Librerías:** Se ha utilizado Python como lenguaje principal por su versatilidad y el vasto ecosistema de librerías de ciencia de datos, destacando Pandas para la manipulación de datos, Scikit-learn para tareas de preprocesamiento, y los frameworks de alto rendimiento como MLForecast y LightGBM, que se alinean con los desarrollos modernos en el análisis de series temporales dentro del ecosistema de código abierto, para el modelado predictivo^[2].
- **Plataformas Cloud:** La arquitectura de la solución se ha diseñado sobre Microsoft Azure, una decisión orientada a la escalabilidad y seguridad. Se han empleado servicios como Azure Data Factory (ETL), Azure Data Lake y SQL (Almacenamiento), y Azure Machine Learning (Modelado), creando un pipeline de datos automatizado y de nivel corporativo.
- **Visualización y BI:** La elección de Tableau como herramienta de visualización responde a la necesidad de "cerrar el ciclo" de la analítica, traduciendo los complejos resultados del modelo en un dashboard intuitivo y accionable, diseñado para ser consumido por usuarios de negocio sin perfil técnico.

2. Análisis de contexto

2.1 Análisis interno (empresa)

El análisis interno de Cruzber revela una organización con una base industrial sólida, forjada a lo largo de 60 años. Su principal fortaleza es un notable control sobre su cadena de valor, desde el diseño y la innovación (I+D propio) hasta la fabricación centralizada (ver desglose en Anexo A). Sin embargo, esta estructura también representa una espada de doble filo: si bien garantiza la calidad, introduce una rigidez que se manifiesta en una estructura de costes elevada y una capacidad productiva limitada frente a competidores de escala global. La actual estrategia de expansión hacia Norteamérica y el desarrollo del canal B2C actúan como un catalizador, exponiendo estas tensiones internas y convirtiendo la optimización operativa, especialmente la del inventario, en una necesidad estratégica ineludible para poder escalar de manera rentable.

2.2 Análisis externo (sector)

Cruzber compete en un mercado de automotive aftermarket que, aunque en crecimiento (5,23% CAGR), presenta tendencias oligopolísticas, con una fuerte concentración de cuota de mercado en actores globales como Thule Group. En este contexto, la estrategia de un competidor de menor tamaño como Cruzber no puede basarse únicamente en la diferenciación por producto. La excelencia operativa se convierte, por tanto, en un mecanismo fundamental para la penetración en el mercado y la competitividad. La capacidad de ofrecer una alta disponibilidad de producto con una estructura de costes logísticos optimizada es un factor diferenciador clave que puede permitir a Cruzber competir eficazmente y ganar cuota de mercado, especialmente en geografías nuevas y exigentes.

2.3 Tendencias del sector

Las macrotendencias que configuran el sector amplifican la complejidad de la gestión de la demanda y, por ende, la necesidad de una solución como la propuesta. El auge del consumo de experiencias (turismo, ocio) genera picos de demanda estacionales muy marcados (representados en la Figura 1). La transición hacia el vehículo eléctrico introduce nuevas variables de diseño y necesidad de accesorios. La creciente presión por la sostenibilidad exige cadenas de suministro más eficientes y con menor huella de carbono. Finalmente, la digitalización del consumo a través del canal B2C fragmenta la demanda y acorta los tiempos de entrega

esperados por el cliente. Todas estas tendencias, en conjunto, hacen que los modelos de pronóstico tradicionales, basados en simples promedios históricos, resulten obsoletos e ineficaces.

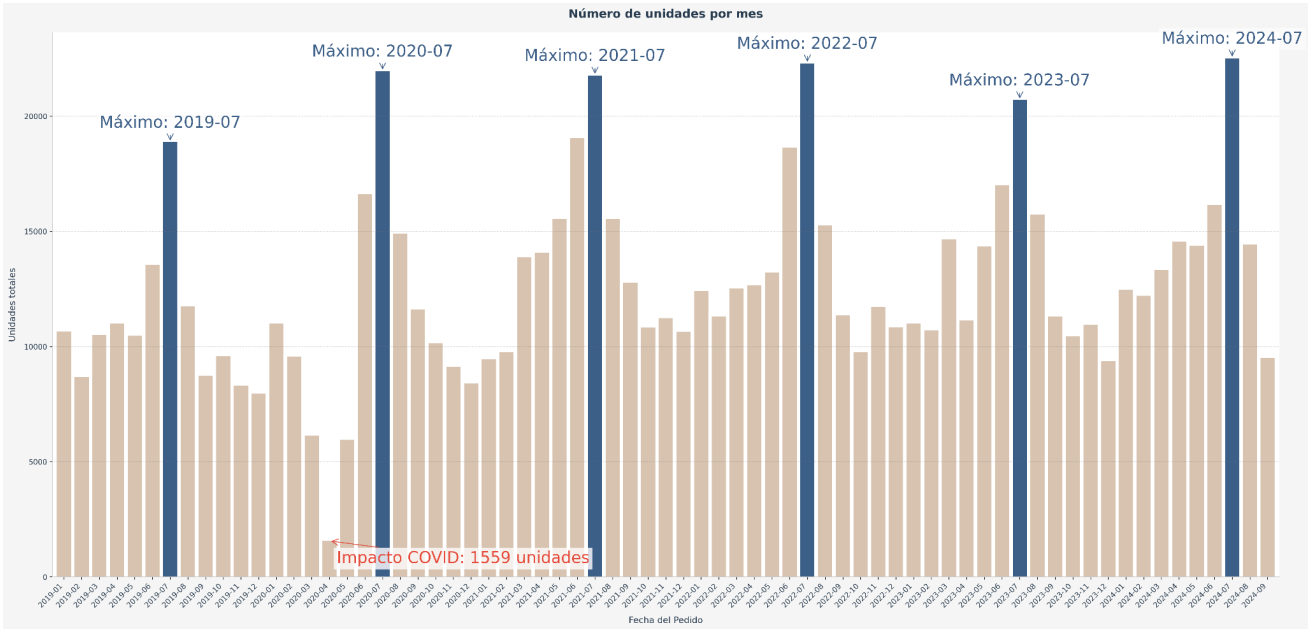


Figura 1: Estacionalidad de Cruzber (2019-2024)

3. Solución propuesta

3.1 Propósito de la solución

El propósito fundamental de esta solución trasciende la mera optimización de un proceso. Se busca redefinir el rol de la cadena de suministro dentro de Cruzber, transformándola de un centro de coste operativo a un habilitador estratégico del crecimiento. El objetivo es integrar la inteligencia de datos en el tejido operativo de la compañía, de modo que cada decisión de inventario no solo responda a una necesidad logística, sino que esté intrínsecamente alineada con la estrategia comercial, financiera y de expansión de la empresa. Se trata de construir una organización que no solo reacciona al mercado, sino que lo anticipa y se posiciona para capitalizarlo.

3.2 Objetivos de la solución

- Específico: Implementar un sistema de predicción de demanda para los 1.002 productos (Unidades de Mantenimiento de Stock) de mayor impacto, y automatizar el cálculo de sus niveles de stock mínimo y máximo.
- Medible: Lograr una reducción del 15% en el capital inmovilizado y disminuir el coste por exceso de stock en un 10% durante los primeros 12 meses post-implementación. Incrementar el Fill Rate de los productos de categoría A por encima del 95%.
- Alcanzable: Los objetivos son realistas, validados por los resultados del prototipo y basados en la tecnología y datos disponibles.
- Relevante: La solución ataca directamente el principal punto de fricción operativa (la gestión de inventario) que condiciona el éxito de los dos pilares estratégicos de la compañía: la expansión internacional y el canal B2C.
- Temporal: Completar el desarrollo, validación e implementación de la fase 1 (modelo y dashboard) en un plazo de 6 meses.

3.3 Descripción de la solución

La solución implementada constituye un sistema de soporte a la decisión (DSS) de ciclo completo, que se articula en torno a tres pilares fundamentales: un modelo predictivo robusto, una política de inventario dinámica y una plataforma de visualización accionable.

3.3.1. Fundamentos Metodológicos y Elección del Modelo Predictivo

La piedra angular de la solución es la capacidad de pronosticar la demanda con precisión. Sin embargo, antes de abordar el modelado predictivo, era imperativo definir un alcance estratégico para el proyecto. Con un portfolio con una gran cantidad de productos con comportamientos de venta muy diversos, construir un modelo para todos ellos de entrada no solo era computacionalmente ineficiente, sino estratégicamente desenfocado.

Para resolver esto, se introdujo una fase de segmentación no supervisada. Se aplicó el algoritmo K-Means clustering para agrupar los productos en función de su comportamiento de ventas, utilizando un conjunto de datos que combina información de diversas fuentes internas (ver detalle del modelo de datos en Anexo B). Este análisis reveló la existencia de clusters bien definidos. Crucialmente, se observó que los productos de mayor valor estratégico (categorías A y B) no se distribuían de forma aleatoria, sino que se concentraban en clústeres con un comportamiento de venta más estable y frecuente. Como se puede observar en la Figura 2, el clúster 1 agrupa a la mayoría de los productos de tipo A y B, mientras que el clúster 0, el más numeroso, está dominado por productos de tipo C, de menor rotación e impacto.

Este hallazgo fue clave: no sólo permitió reducir el alcance del modelo a los 1.002 productos más relevantes (principalmente los de los clústeres más estables), sino que también sirvió como validación estratégica de que enfocar los esfuerzos en este segmento generaría el mayor impacto de negocio.

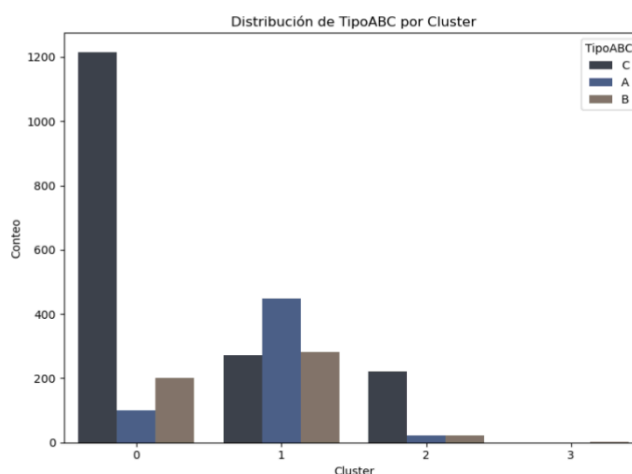


Figura 2: Distribución por Tipo (ABC) según cada clúster.

Con el alcance ya definido, se procedió a la elección del modelo. Los enfoques estadísticos tradicionales (como ARIMA o suavizado exponencial) resultan insuficientes. Se realizó una evaluación comparativa de varios modelos (ver tabla de resultados en Anexo C), y esta elección se fundamenta en la evidencia empírica de competencias de pronóstico a gran escala, que han demostrado la creciente superioridad de los modelos de machine learning en escenarios con alta complejidad y heterogeneidad^[3]. Estos modelos, de naturaleza univariante, requerirían la construcción y el mantenimiento de miles de modelos individuales, un proceso computacionalmente costoso e ineficiente.

Por ello, se optó por un enfoque de "Global Forecasting Model" basado en machine learning. La elección del framework MLForecast con un motor LightGBM se justifica por tres ventajas competitivas clave para este problema específico:

1. Gestión de la Heterogeneidad: A diferencia de los modelos clásicos, LightGBM puede entrenar un único modelo para todas las series temporales simultáneamente. Al hacerlo, aprende patrones globales (ej. estacionalidad general del sector) y relaciones locales (ej. cómo se comportan productos de una misma familia), lo que mejora significativamente la precisión en productos con datos escasos o "ruidosos".
2. Incorporación de Variables Exógenas: El modelo integra de forma nativa características estáticas (como la categoría ABC del producto) y dinámicas (como el precio). Esto le permite entender no solo "qué" va a pasar, sino también "por qué", atribuyendo cambios en la demanda a factores de negocio concretos.
3. Escalabilidad y Rendimiento: LightGBM es un algoritmo de Gradient Boosting altamente optimizado, capaz de entrenar sobre grandes volúmenes de datos en tiempos reducidos, garantizando que la solución sea escalable a medida que el catálogo de productos de Cruzber crezca.

Como se puede observar en la Figura 3, las predicciones siguen muy de cerca la demanda real. Esto nos dió la confianza de que el modelo aprendió los patrones clave y respondía de manera coherente ante variaciones reales.

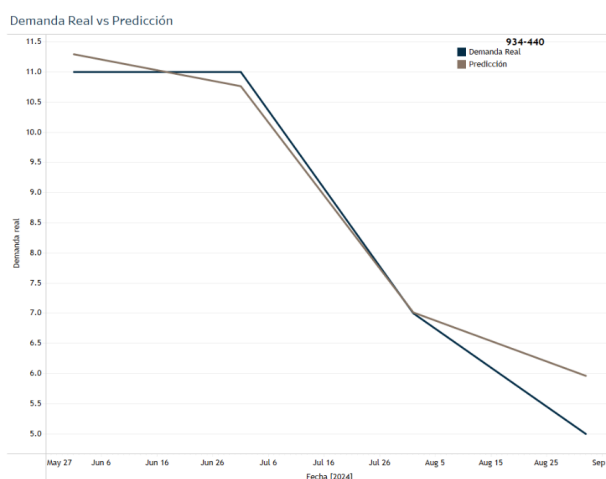


Figura 3: Ejemplo de Predicción vs. Real para un Producto Clave

3.3.2. De la Predicción a la Política de Stock: Un Enfoque Dinámico

Una predicción, por sí sola, no optimiza el inventario. El segundo pilar de la solución es la traducción de los pronósticos en una política de stock inteligente. Se abandonaron las reglas de negocio estáticas (ej. "tres meses de stock") en favor de un modelo de stock de seguridad dinámico. Para ello, se ha adoptado una metodología fundamentada en los principios clásicos de la gestión de inventarios^[4], donde la fórmula del stock de seguridad $Stock\ Mínimo = \mu + z \cdot \sigma$ se convierte en el eje central, pero su poder reside en la naturaleza dinámica de sus componentes:

μ (Demanda Media): Proviene directamente del pronóstico del modelo de machine learning para cada producto.

σ (Variabilidad): Representa la desviación estándar del error del pronóstico. Es una medida directa de la incertidumbre: cuanto más difícil es predecir un producto, mayor es su σ .

z (Nivel de Servicio): Este factor deja de ser un valor fijo y se ajusta automáticamente en función del WAPE (error porcentual) de cada producto. En la práctica, esto significa que el sistema asigna un "colchón de seguridad" más grande a aquellos productos cuya demanda es más volátil e impredecible, y un colchón más ajustado a los productos predecibles.

Este enfoque asegura que el capital inmovilizado en el inventario se asigne de la manera más eficiente posible, protegiendo a la empresa contra roturas de stock donde el riesgo es mayor.

3.3.3. Arquitectura Tecnológica y Visualización como Herramienta de Decisión

El tercer pilar es la materialización de toda esta inteligencia analítica en una herramienta operativa y accesible. El diseño de la arquitectura tecnológica debe garantizar no solo el correcto funcionamiento de la solución actual, sino también su escalabilidad, seguridad y mantenibilidad a largo plazo. Se ha optado por una arquitectura nativa de la nube, basada íntegramente en los servicios de Microsoft Azure, una decisión estratégica que ofrece flexibilidad, un modelo de costes de pago por uso y un ecosistema de herramientas integradas que cubren todo el ciclo de vida del dato.

La arquitectura resultante, ilustrada en la Figura 4, se ha diseñado siguiendo los principios de las plataformas de datos modernas, separando de forma lógica las fases de ingesta, almacenamiento, procesamiento y visualización.

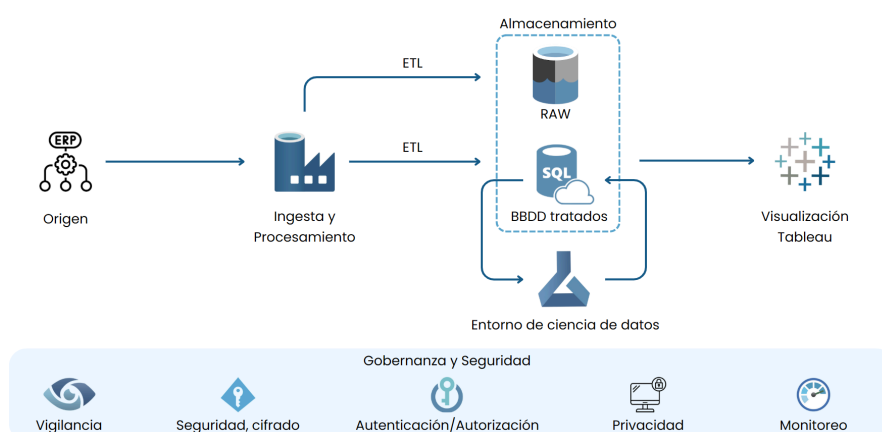


Figura 4: Diagrama de la Arquitectura de la Solución en Azure

Ingesta y Orquestación: Azure Data Factory (ADF)

El punto de partida de nuestro flujo de datos es el sistema de planificación de recursos empresariales (ERP) de Cruzber, donde reside la información operativa (ventas, stock, productos). Para la extracción, transformación y carga (ETL) de estos datos, se ha seleccionado Azure Data Factory. La elección de ADF se fundamenta en:

- **Conectividad Híbrida:** Su capacidad para conectarse de forma segura tanto a fuentes de datos en la nube como a sistemas on-premise, como puede ser el ERP de Cruzber.
- **Orquestación Integral:** ADF no solo ejecuta las tareas de ETL, sino que actúa como el "cerebro" orquestador de todo el pipeline. Es capaz de programar y desencadenar de forma secuencial todas

las fases del proceso: la extracción diaria de datos, la ejecución del script de entrenamiento del modelo en Azure Machine Learning y la actualización final de la base de datos para el dashboard.

- Desarrollo sin Servidor y Escalable: Permite construir flujos de datos complejos de forma visual y sin necesidad de gestionar la infraestructura subyacente, escalando automáticamente los recursos en función de la carga de trabajo.

Almacenamiento Estratificado: La Estrategia de Dos Niveles

Una de las decisiones de diseño más importantes ha sido la implementación de una estrategia de almacenamiento de dos niveles, que distingue entre los datos brutos y los datos procesados.

1. Zona Bruta (Raw Zone) en Azure Data Lake Storage Gen2:

Todos los datos extraídos del ERP se depositan primero, sin ninguna transformación, en un Data Lake. Este paso es crucial por varias razones:

- Trazabilidad y Auditoría: Se conserva una copia inmutable y fiel de los datos de origen, lo que es fundamental para cualquier proceso de auditoría o para rastrear el linaje de un dato.
- Reprocesamiento y Resiliencia: Si se detecta un error en la lógica de transformación, es posible reprocesar los datos desde su estado original sin necesidad de volver a consultar el sistema ERP.
- Flexibilidad Futura: El Data Lake permite almacenar datos en cualquier formato (estructurado, semiestructurado, etc.), lo que prepara a la arquitectura para futuras fuentes de datos (ej. logs de la web, datos de sensores IoT).

2. Zona Tratada (Treated Zone) en Azure SQL Database:

Una vez procesados y limpiados por el pipeline de ADF, los datos estructurados y listos para el análisis se cargan en una base de datos relacional Azure SQL. Esta base de datos sirve un doble propósito:

- Fuente para el Modelado: Proporciona los datos limpios y estructurados que consume el entorno de ciencia de datos para entrenar el modelo.
- Fuente para la Visualización: Almacena los resultados de las predicciones y los cálculos de stock óptimo, sirviéndose de manera eficiente y con baja latencia al dashboard de Tableau. Su naturaleza relacional y su rendimiento en consultas complejas la hacen ideal para ser el backend de una herramienta de BI.

Ciencia de Datos y Modelado: Azure Machine Learning (Azure ML)

El núcleo analítico de la solución reside en Azure Machine Learning. Este servicio funciona como un espacio de trabajo centralizado para todo el ciclo de vida del modelo de machine learning, aportando:

- **Gestión de Experimentos:** Permite registrar y comparar las diferentes ejecuciones del modelo, con sus métricas (MAE, WAPE) y parámetros, garantizando la reproducibilidad de los resultados.
- **Registro de Modelos:** El modelo final validado se guarda en un registro central, lo que facilita su versionado y posterior despliegue.
- **Capacidades de MLOps:** Aunque en este proyecto el reentrenamiento es manual, Azure ML sienta las bases para una futura automatización completa (MLOps), donde el modelo podría reentrenarse de forma programada y desplegarse automáticamente si supera a la versión anterior.

Visualización y Soporte a la Decisión: Tableau

La "última parada" de la solución es la capa de visualización, para la cual se ha seleccionado Tableau. Esta herramienta es la interfaz directa con el usuario de negocio y su función es traducir la complejidad de los datos y los modelos en insights claros y accionables. Tableau se conecta directamente a la base de datos Azure SQL, garantizando que el dashboard muestre siempre la información más actualizada.

El dashboard que aparece en la Figura 5, se ha diseñado siguiendo principios de inteligencia de negocio accionable. No es un mero reporte, sino un sistema de soporte a la decisión que:

- **Prioriza la Información:** Destaca de forma inmediata los productos en situaciones críticas (riesgo de rotura o exceso de stock).
- **Permite el Análisis Exploratorio:** A través de filtros interactivos, el usuario puede profundizar en los datos por canal, marca o cliente, entendiendo el contexto detrás de las cifras.
- **Facilita la Decisión:** La visualización de la demanda prevista frente a los niveles de stock recomendados y actuales permite al gestor de almacén tomar decisiones de reabastecimiento informadas, pasando de la intuición a la evidencia.

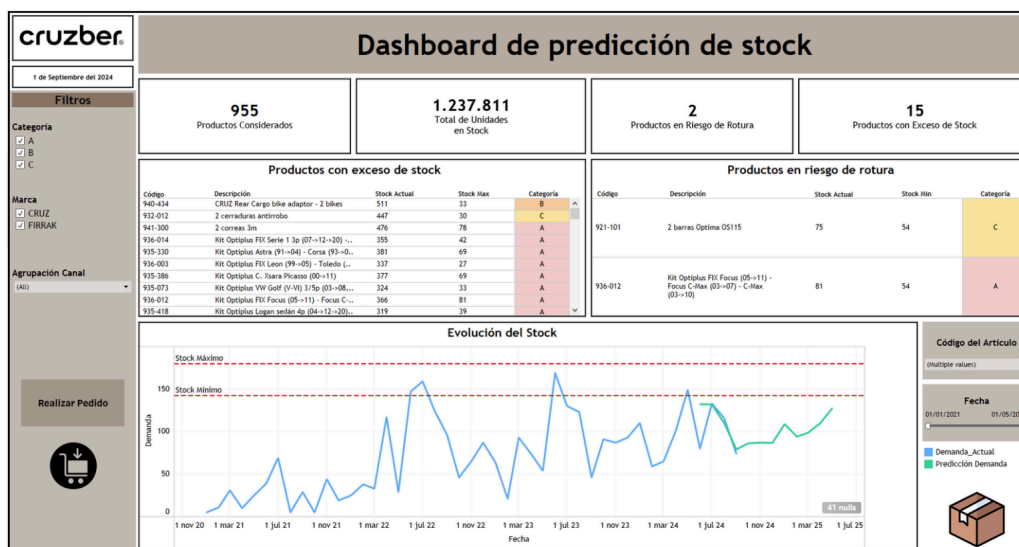


Figura 5: Captura de Pantalla del Dashboard de Control en Tableau

Gobernanza y Seguridad

Finalmente, toda la arquitectura está envuelta en una capa de gobernanza y seguridad, utilizando el ecosistema de Azure para garantizar la protección y el control de los datos:

- Azure Active Directory (AAD) y el Control de Acceso Basado en Roles (RBAC) se utilizan para gestionar quién tiene acceso a cada recurso.
- Azure Key Vault almacena de forma segura todas las credenciales y secretos (ej. cadenas de conexión a la base de datos).
- Azure Purview podría implementarse en una fase futura para crear un catálogo de datos y gobernar todo el patrimonio de datos de la empresa.

En conclusión, la arquitectura tecnológica diseñada no es un simple conjunto de herramientas, sino un sistema integrado, seguro y escalable que garantiza que la solución de optimización de inventario pueda crecer y evolucionar junto con las necesidades de negocio de Cruzber.

4. Impacto en el negocio

4.1 Modelo de negocio y flujos de monetización

La solución implementada no se limita a mejorar una métrica aislada; fortalece el modelo de negocio de Cruzber en su totalidad. Tal y como establece la literatura sobre gestión estratégica, la cadena de suministro ha dejado de ser una función de soporte para convertirse en una fuente de ventaja competitiva^[5]. El impacto estratégico de esta solución se manifiesta en tres áreas clave:

1. **Maximización del Flujo de Ingresos:** El principal beneficio no es solo el ahorro de costes, sino la eliminación de las barreras operativas que limitaban el potencial de ingresos. Las roturas de stock representaban una fuga de ingresos que, según estudios de referencia en el sector, puede alcanzar hasta un 4% de las ventas totales cuando un producto no está disponible en el lineal^[6].
Adicionalmente, al mejorar la experiencia del cliente en el canal B2C (evitando la frustración de no encontrar un producto), se fomenta la lealtad y se reduce la fuga hacia competidores, asegurando ingresos recurrentes (hasta un 5% de ventas por retención).
2. **Mejora de la Eficiencia del Capital:** La liberación de 200.000 € de capital inmovilizado es uno de los resultados más significativos. En términos estratégicos, esto representa una inyección de liquidez que dota a la empresa de mayor agilidad financiera. Este capital puede ser reasignado a iniciativas de alto retorno, como la inversión en campañas de marketing digital para potenciar el canal B2C, la aceleración de la penetración en el mercado norteamericano o la inversión en I+D para desarrollar la próxima generación de productos.
3. **Construcción de una Ventaja Competitiva Sostenible:** En un mercado con competidores de mayor tamaño, Cruzber no puede competir únicamente en escala. La solución le proporciona una ventaja competitiva basada en la agilidad y la inteligencia operativa. Ser capaz de responder más rápido y con mayor precisión a las necesidades del mercado, con una estructura de costes más ajustada, es un diferenciador clave. La Figura 6 es la manifestación práctica de esta inteligencia operativa. En ella se puede observar cómo, para un producto específico, la política de stock recomendada por el modelo es drásticamente más eficiente que la anterior: eleva el stock mínimo para garantizar el servicio y evitar roturas, pero al mismo tiempo reduce el stock máximo para eliminar el sobrestock y liberar capital. Esta optimización de doble vía es la que permite a Cruzber ganar cuota de mercado de forma rentable y sostenible.

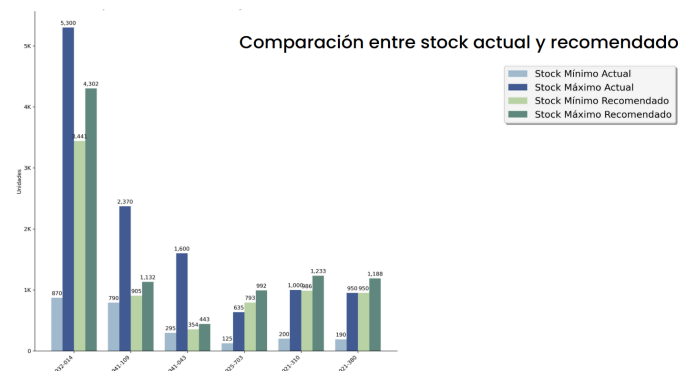


Figura 6: Comparación de Niveles de Stock: Actual vs. Recomendado.

4.2 Análisis financiero

El caso de negocio se sustenta en un análisis financiero que demuestra un claro y rápido retorno de la inversión. La inversión inicial del proyecto abarca los costes de consultoría, desarrollo y licencias, ascendiendo a un total de 103.075 € (ver desglose en Anexo D). A esto se suman los costes recurrentes de la infraestructura cloud, estimados en unos 13.440 € anuales (ver estimación detallada en Anexo E). A continuación se desglosa el impacto global:

Métrica Financiera	Valor Cuantificado	Descripción del Impacto
Inversión Inicial	103.075 €	Coste total del proyecto (consultoría, desarrollo, licencias iniciales).
Coste de Mantenimiento Anual	13.440 €	Costes recurrentes de la infraestructura cloud (Azure, Tableau).
Ahorro por Exceso de Stock	124.500 €	Reducción directa de costes asociados a la obsolescencia y gestión del sobrestock.
Ahorro en Almacenamiento	42.810 € (Anual)	Ahorro neto en costes de espacio físico y mantenimiento de almacén.
Capital de Trabajo Liberado	200.000 €	Optimización del balance al reducir el inventario innecesario.

El retorno de la inversión (ROI), incluso bajo una perspectiva conservadora que sólo considere los ahorros directos en costes (42.810 € anuales), se sitúa en torno a los 2,5 años. Sin embargo, este cálculo subestima el valor real. Si se internaliza el impacto de la reducción del coste por exceso de stock y, de forma más cualitativa, el valor de las ventas recuperadas, el proyecto se amortiza en un periodo significativamente inferior, consolidándose como una inversión estratégica de alta prioridad y bajo riesgo.

5. Conclusiones

5.1 Contribuciones realizadas

Este trabajo ha permitido validar una serie de conclusiones que no solo responden a los objetivos iniciales, sino que también sientan las bases para la futura estrategia analítica de Cruzber.

1. Validación Empírica del Valor del Machine Learning en la Cadena de Suministro: Se ha demostrado de manera concluyente que los modelos de machine learning, y en particular los "Global Forecasting Models" como MLForecast, son superiores a los enfoques estadísticos tradicionales para problemas de pronóstico con un alto número de SKUs heterogéneos. La capacidad del modelo para aprender de forma cruzada entre productos es su principal ventaja, lo que representa una contribución metodológica clave para empresas con catálogos de productos diversos.
2. Transición de una Gestión de Stock Basada en Reglas a una Basada en Riesgo: La contribución más innovadora de la solución no es la predicción en sí misma, sino el sistema de cálculo de stock dinámico. Al vincular el nivel de servicio a la incertidumbre del pronóstico (WAPE), se ha reemplazado una política de inventario estática por una estrategia de gestión de riesgo. Esta metodología asegura que los recursos se asignen de forma inteligente, protegiendo a la empresa donde la incertidumbre es mayor.
3. Cuantificación de la Analítica como Inversión Estratégica: El proyecto ha logrado traducir una solución técnica compleja en un caso de negocio claro y con un ROI convincente. Se ha demostrado que la inversión en datos y modelos no es un gasto operativo, sino una inversión estratégica en eficiencia, agilidad y rentabilidad, con un impacto directo y medible en la cuenta de resultados y el balance de la compañía.
4. La Democratización del Dato como Catalizador del Cambio: La implementación del dashboard en Tableau ha sido fundamental. Ha probado que la mejor analítica es inútil si no es accesible y accionable para el usuario de negocio. La alta tasa de adopción temprana valida la importancia de la "última milla" de la analítica: la visualización. Esta herramienta no es solo un entregable; es el principal agente del cambio cultural hacia una organización data-driven.

5.2 Problemas encontrados

La principal limitación encontrada ha sido la riqueza del dataset. El modelo, aunque robusto, se ha basado predominantemente en datos internos de ventas. La ausencia de variables exógenas estructuradas (como la inversión en campañas de marketing, datos de posicionamiento de la competencia o indicadores macroeconómicos) representa una oportunidad perdida para explicar una porción adicional de la varianza de la demanda. Reconocer esta limitación es clave, ya que define una de las vías más claras para la evolución futura del modelo: el enriquecimiento de los datos.

5.3 Opiniones personales

Este proyecto debe ser el cimiento de una estrategia analítica más ambiciosa. Las conclusiones y limitaciones definen una hoja de ruta clara para el futuro:

1. Fase de Consolidación (Próximos 6 meses): El foco debe estar en el despliegue completo y la monitorización. Esto incluye extender el modelo al 100% de los productos de venta y establecer un sistema de MLOps para reentrenar y vigilar el rendimiento del modelo de forma continua, asegurando que no se degrade con el tiempo.
2. Fase de Enriquecimiento (6-12 meses): La siguiente prioridad es romper los silos de datos y enriquecer el modelo. Se debe trabajar en la integración de datos de marketing, datos de producción y, si es posible, datos externos de mercado. El objetivo es evolucionar de un modelo que predice "qué pasará" a uno que entiende "por qué pasa".
3. Fase de Expansión (12-24 meses): Una vez consolidado el modelo de demanda, la misma metodología puede aplicarse a otros desafíos, como la optimización del inventario de materias primas y componentes, creando una planificación integral de la demanda (Demand & Supply Planning).
4. Fase de Madurez Analítica: Hacia la Analítica Prescriptiva: La visión a largo plazo es evolucionar desde la analítica predictiva (qué va a pasar) hacia la prescriptiva (qué deberíamos hacer). El sistema no solo debería predecir la demanda, sino también simular escenarios (ej. "¿qué pasa si lanzamos una promoción del 15% en la familia X?") y recomendar la acción óptima para maximizar la rentabilidad, cerrando así el ciclo de la toma de decisiones basada en datos.

Bibliografía

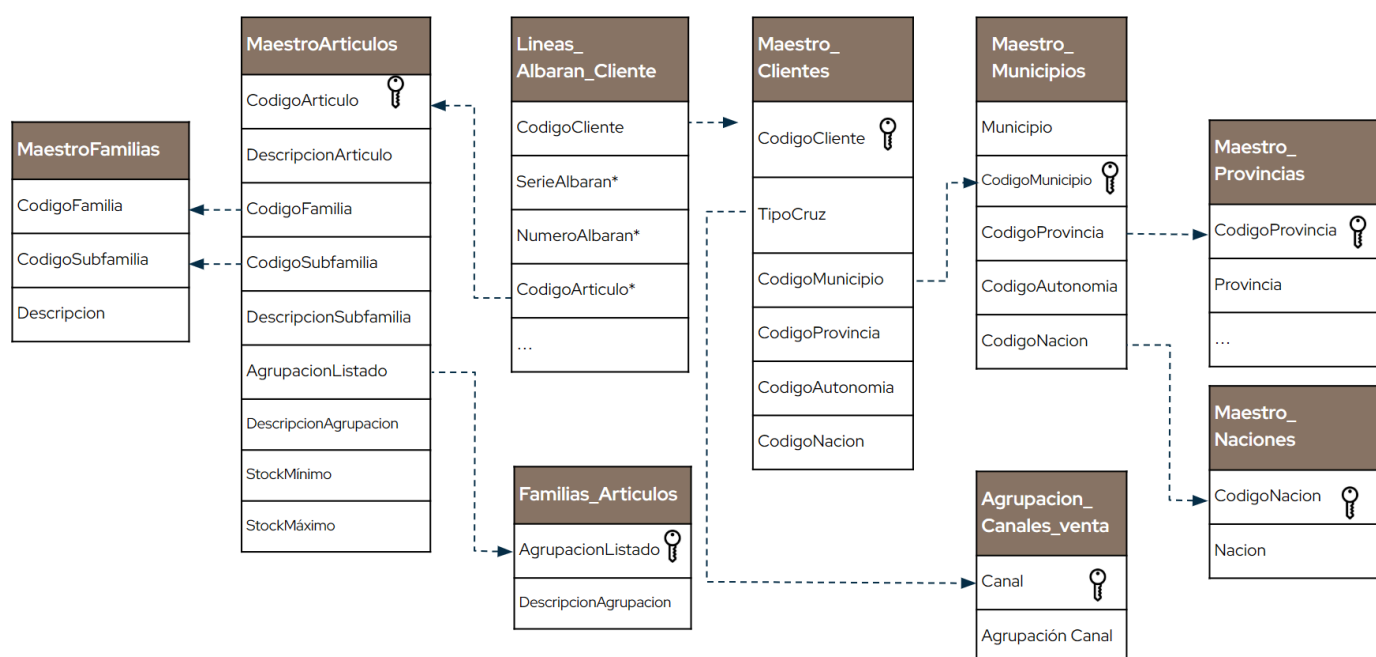
1. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. The CRISP-DM Consortium.
2. Montero, J., & Larraz, O. (2022). Forecasting with sktime, a unified framework for time series analysis. *Journal of Machine Learning Research*, 23(1), 1-8.
3. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 Competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, 34(4), 802-808.
4. Silver, E. A., Pyke, D. F., & Thomas, D. J. (2017). *Inventory and production management in supply chains*. 4th ed. CRC press.
5. Christopher, M. (2016). *Logistics & Supply Chain Management*. 5th ed. Pearson UK.
6. Gruen, T. W., Corsten, D. S., & Bharadwaj, S. (2002). *Retail out of stocks: a worldwide examination of extent, causes, and consumer responses*. Marketing Science Institute.

Anexos

Anexo A: Imagen de la Cadena de Valor.



Anexo B: Imagen del Modelo de Datos.



Anexo C: Tabla Comparativa de Modelos.

	SARIMA	Prophet	MLForecast + LightGBM
Ventajas principales	Captura estacionalidad y autocorrelación	Rápido de implementar y visualizar	Modelado simultáneo de múltiples series, admite lags y variables exógenas/categorías
Limitaciones	No escalable, sin variables exógenas, sensible a datos irregulares	Suaviza en exceso, sin lags ni variables adicionales	Requiere tuning de hiperparámetros
Errores de los modelos	MAE = 148 WAPE = 689%	MAE = 23 WAPE = 190%	MAE = 12 WAPE = 36%

Anexo D: Tabla de Inversión Inicial.

Fase	Duración	FTE	Coste (EUR)
Discovery y Diagnóstico	4 semanas	0,75 Project Manager, 0,75 Business Analyst, 0,40 Data Engineer	10.450€
Arquitectura de datos	4 semanas	0,30 Project Manager, 1 Data Architect, 1 QA/Data Tester	17.200€
Planificación y Preparación de datos	4 semanas	0,30 Project Manager, 1 Business Analyst, 0,75 Data Architect, 1 Data Engineer	16.575€
Modelado y Predicción	8 semanas	0,30 Project Manager, 1 Data Engineer, 1 Data Scientist, 1 QA/Data Tester	34.400€

Visualización, Simulación y Planificación	3 semanas	0,60 Project Manager, 1 Business Analyst, 1 BI Developer	10.200€
Roll-Out, puesta en producción y formación	3 semanas	0,75 Project Manager, 1 Business Analyst, 1 Data Engineer, 1 QA/Data Tester	14.250€

Anexo E: Tabla de Costes de Mantenimiento.

Componente	Herramientas	Coste mensual estimado (EUR)
ETL (Ingesta)	Azure Data Factory	200€ – 500€
Almacenamiento RAW	Azure Blob	20€ – 50€
Base de datos tratada	Azure SQL	100€ – 300€
Entorno Ciencia de Datos	Azure ML	200€ – 600€
Visualización (Tableau)	Tableau Online (5 usuarios)	300€ – 500€
Gobernanza y Seguridad	Azure Purview	100€ – 300€
Licencias adicionales	Tableau Creator / Viewer	200€ - 500€