******

Unamuno Acha, Isabel

2023 - 2024

Máster Universitario en

Análisis de Datos para la Inteligencia de Negocio / Business Analytics

**Análisis de inventario**

TRABAJO DE FIN DE MÁSTER

**ANÁLISIS DE INVENTARIO**

TRABAJO DE FIN DE MÁSTER PRESENTADO EN: Mondragon Unibertsitatea

PARA LA OBTENCIÓN DEL TÍTULO DE: Máster Universitario en Análisis de Datos para la Inteligencia de Negocio / Business Analytics

AUTOR/A: Isabel Unamuno Acha

DIRECTOR/A: XXX

TUTOR/A: Jon Perez Visaires

ORGANIZACIÓN EN LA QUE HA REALIZADO EL PROYECTO: Datua

FECHA DE DEFENSA: Bilbao, a DD de MM de AAAA

El autor/la autora del Trabajo de Fin de Máster, autoriza a la Facultad de Empresariales de Mondragon Unibertsitatea, con carácter gratuito y con fines exclusivamente de investigación y docencia, los derechos de reproducción y comunicación pública de este documento siempre que: se cite el autor/la autora original, y el uso que se haga de la obra no sea comercial.

****

**Reconocimiento – NoComercial – CompartirIgual (by-nc-sa):** No se permite un uso comercial de la obra original ni de las posibles obras derivadas, la distribución de las cuales se debe hacer con una licencia igual a la que regula la obra original.

DECLARACIÓN DE ORIGINALIDAD

Yo Isabel Unamuno Acha

Declaro que este Trabajo de Fin de Máster es original, fruto de mi trabajo personal, y que no ha sido previamente presentado para obtener otro título o calificación profesional.

Las ideas, formulaciones, imágenes, ilustraciones tomadas de fuentes ajenas han sido debidamente citadas y referenciadas.

**RESUMEN**

**LABURPENA**

**ABSTRACT**

**ÍNDICE**

**ÍNDICE DE FIGURAS**

**ÍNDICE DE TABLAS**

# INTRODUCCIÓN

## Presentación de la empresa

## Problematica

## Marco teórico

En el entorno empresarial actual, la aplicación de algoritmos de Machine Learning para la gestión de inventarios se ha convertido en una herramienta esencial. Estas técnicas avanzadas permiten a las empresas predecir con mayor precisión la demanda futura y ajustar sus niveles de stock de manera eficiente. Al optimizar los costos de almacenamiento y reducir las pérdidas asociadas con la escasez de productos, las empresas pueden no solo mejorar su rentabilidad, sino también elevar la satisfacción del cliente al asegurar la disponibilidad continua de los productos más demandados. La implementación de algoritmos de Machine Learning en la gestión de inventarios proporciona una ventaja competitiva significativa. Estas herramientas permiten a las empresas operar de forma más eficiente, responder mejor a las fluctuaciones del mercado y adaptarse rápidamente a las necesidades cambiantes de los consumidores. En última instancia, la capacidad de predecir la demanda y gestionar el stock de manera efectiva no solo mejora la eficiencia operativa, sino que también impulsa el crecimiento y el éxito en un mercado cada vez más competitivo.

Por ello, en el siguiente apartado, primero de todo se ha resumido el concepto de machine learning y a continuación se han analizado cuáles son los diferentes algoritmos de machine learning que se emplean actualmente para la predicción de la demanda y la mejora en la gestión de stock. <https://openaccess.uoc.edu/bitstream/10609/129826/8/jde_masTFM0121memoria.pdf>

Los algoritmos de Machine Learning permiten identificar patrones complejos en grandes volúmenes de datos, infiriendo sus propias reglas para reconocer patrones similares en nuevos conjuntos de datos. Esto da lugar a la creación de sistemas inteligentes que mejoran de forma autónoma mediante la observación de datos. Estos sistemas pueden aprender a predecir comportamientos, detectar similitudes o anomalías de manera automática, y tomar decisiones adecuadas basadas en los datos analizados. La Inteligencia Artificial se basa en el aprendizaje a partir de ejemplos. Los modelos de Machine Learning analizan datos reales que representan el proceso que se desea mejorar o automatizar. Este proceso, conocido como entrenamiento de algoritmos, permite que el sistema extraiga conclusiones relevantes y aprenda a realizar tareas sin necesidad de programación previa, simplemente observando y analizando los datos proporcionados. En los proyectos de Machine learning, hay tres pasos que se deben seguir, que son los siguientes. Se comienza con la revisión de datos, el cual implica examinar los datos disponibles para determinar cuáles son útiles, asegurando su correcto almacenamiento y formatos adecuados. Es crucial limpiar los datos eliminando registros antiguos, incompletos o erróneos para garantizar la calidad del dataset. Se sigue con la organización de datos, siendo esta fundamental para facilitar el aprendizaje automático. Se seleccionan datos que reflejen las decisiones que el sistema deberá tomar automáticamente, estructurándolos de manera que optimicen el proceso de entrenamiento del modelo. Y por último, el entrenamiento y la validación del modelo. Durante esta fase, el modelo analiza los datos para identificar patrones y relaciones, ajustando sus parámetros para poder detectar estas características automáticamente en futuros datos. Este proceso permite al modelo configurar su propio sistema de detección basado en la información observada. <https://www.iic.uam.es/inteligencia-artificial/machine-learning-deep-learning/>

**Redes neuronales y modelos Machine Learning en gestión de inventarios**

Las redes neuronales son ampliamente utilizadas para predecir el comportamiento del inventario. Los modelos de backpropagation (BPNN) destacan por su estructura simple y capacidad de aprendizaje. He (2013) mejoró la precisión de BPNN añadiendo una compensación de error y utilizando una arquitectura de tres capas con nueve parámetros de entrada.

Šustrová (2016) desarrolló varios modelos de redes neuronales para optimizar la cantidad de stock, seleccionando el modelo con menor error cuadrático medio (MSE) y un coeficiente de determinación (R2) cercano a 1. Su modelo final fue una red feed-forward con backpropagation, función de entrenamiento TRAINGDX y función de transferencia TANSIG.

Barbosa, Pestana y Goliatt (2017) abordaron el problema del desequilibrio de clases en datos mediante "Imbalanced learning" para mejorar el nivel de servicio al cliente, identificando materiales con alta probabilidad de rotura de stock.

Shoujing Zhang et al. (2020) combinaron BPNN con clustering mejorado y PCA para gestionar el inventario de piezas de repuesto en mantenimiento de camiones, evitando el impacto subjetivo de los gestores de inventario.

Por otro lado, Praveen K.B. et al. (2020) utilizaron el modelo de regresión XGBoost para predicciones de demanda en pequeñas y medianas empresas, logrando reducir el stock y el capital dedicado. Utilizaron early-stopping-runs para evitar el sobreajuste.

También se mencionan la importancia de considerar la estacionalidad y los patrones semanales de demanda en los modelos de predicción.

<https://openaccess.uoc.edu/bitstream/10609/129826/8/jde_masTFM0121memoria.pdf> Honek jarri dun bibliografiako linka rebai hartu

Para realizar las tareas de predicción, que es el objetivo del proyecto, se han tenido en cuenta algunos de los algoritmos recogidos en el análisis de la literatura y que mejores resultados han obtenido. Específicamente, Khalilzadeh (2008), He (2013) o Šustrová (2016) emplearon redes neuronales para realizar sus predicciones, Barbosa et al. (2017) usaron Random Forest y Gradient Tree Boosting para las suyas, Praveen et al. (2020) utilizó XGBOOST (Extreme Gradient Boosting) para hacer predicciones de la demanda y finalmente Shoujing Zhang et al. (2020) usó un algoritmo de clustering mejorado (Mahalanobis distance clustering algorithm) combinado con una BP NN, usando PCA (Principal Component Analysis) antes del algoritmo mejorado para realizar una reducción de dimensionalidad de datos y K-Means como método de clustering, con el fin de averiguar el grado de importancia de las piezas de repuesto. Se han realizado pruebas con varios de estos algoritmos que se detallarán explicando el resultado obtenido. <https://openaccess.uoc.edu/bitstream/10609/129826/8/jde_masTFM0121memoria.pdf>

1. Red Neuronal Artificial multicapa con Embeddings

Los Embeddings son una técnica para convertir variables categóricas en representaciones continuas, que se utilizan como entradas en un modelo de red neuronal para aprendizaje supervisado. Estas representaciones permiten que las categorías similares se ubiquen más cerca entre sí y se ajustan durante el entrenamiento para minimizar la pérdida del modelo. Al emplear Embeddings, se pueden obtener predicciones más precisas en comparación con el uso exclusivo de las unidades de stock. En este proyecto, además de los Embeddings y las unidades de stock, también se considerarán las unidades de venta para mejorar la precisión de las predicciones.

1. Gradient Boosting Regressor

Este estimador construye un modelo aditivo por etapas y permite optimizar funciones de pérdida diferenciables arbitrarias. En cada etapa se ajusta un árbol de regresión sobre el gradiente negativo de la función de pérdida dada. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html>

1. Random Forest Regressor

Un bosque aleatorio es un metaestimador que ajusta una serie de árboles de decisión regresores a varias submuestras del conjunto de datos y utiliza el promedio para mejorar la precisión predictiva y controlar el sobreajuste. Los árboles del bosque utilizan la mejor estrategia de división, lo que equivale a pasar splitter=«best» al DecisionTreeRegressor subyacente. El tamaño de la submuestra se controla con el parámetro max\_samples si bootstrap=True (por defecto); en caso contrario, se utiliza todo el conjunto de datos para construir cada árbol.

Traducción realizada con la versión gratuita del traductor DeepL.com

1. Decision Tree Regressor

Un [árbol de decisión](https://www.geeksforgeeks.org/decision-tree/) es un tipo de algoritmo de aprendizaje supervisado que se usa comúnmente en el aprendizaje automático para modelar y predecir resultados basados ​​en datos de entrada. Es una estructura en forma de árbol donde cada nodo interno prueba el atributo, cada rama corresponde al valor del atributo y cada nodo hoja representa la decisión o predicción final. El algoritmo del árbol de decisión entra en la categoría de [aprendizaje supervisado](https://www.geeksforgeeks.org/supervised-unsupervised-learning/) . Se pueden utilizar para resolver**problemas tanto de regresión** como de clasificación . <https://www.geeksforgeeks.org/decision-tree-introduction-example/>

1. Light Gradient Boosting Machine Regressor

Light GBM (Light Gradient Boosting Machine) es un marco de código abierto ampliamente utilizado para el refuerzo por gradiente. Destacado por su capacidad para manejar conjuntos de datos de gran tamaño, ofrece un rendimiento superior en comparación con otros marcos de refuerzo por gradiente como XGBoost y CatBoost.

Una de las características distintivas de Light GBM es su método de muestreo unilateral basado en el gradiente para dividir los árboles. Este enfoque reduce significativamente el uso de memoria y mejora la precisión del modelo. Además, Light GBM emplea un crecimiento por hojas en lugar de crecimiento por niveles, lo que lo hace considerablemente más rápido que los métodos tradicionales de crecimiento en profundidad.

Light GBM es versátil y puede manejar diversos tipos de datos, incluidos categóricos, numéricos y de texto. Además, incorpora funciones integradas para el preprocesamiento de datos, validación cruzada y ajuste de hiperparámetros, lo que facilita a los usuarios la optimización de sus modelos. <https://samanemami.medium.com/light-gradient-boosting-machine-b4f1b9e3f7d1>

1. Extreme Gradient Boosting Regressor

El Extreme Gradient Boosting, conocido como XGBoost, fue desarrollado por Tianqi Chen (Chen & Guestrin, 2016) y es un algoritmo de aprendizaje automático supervisado utilizado para problemas de clasificación y regresión. XGBoost es un tipo de método de refuerzo por gradiente que se diferencia de otros modelos de refuerzo de gradiente de las siguientes maneras:

**Regularización L1 y L2:** XGBoost incorpora tanto L1 como L2, lo que ayuda a mejorar la generalización del modelo y a reducir el sobreajuste.

**Cálculo del gradiente:** Mientras que los modelos tradicionales de refuerzo por gradiente utilizan la tasa de error para calcular el gradiente a través de la derivada parcial de la función de pérdida, XGBoost va un paso más allá al utilizar la segunda derivada parcial de la función de pérdida. Esto proporciona información adicional sobre la dirección del gradiente.

**Velocidad:** XGBoost es generalmente más rápido que otros métodos de gradient boosting debido a la paralelización en la construcción del árbol.

**Manejo de valores faltantes:** XGBoost puede gestionar valores faltantes en un conjunto de datos, lo que reduce el tiempo necesario para la preparación de los datos.

<https://www.sciencedirect.com/topics/computer-science/extreme-gradient-boosting>

Aunque

Volviendo al entorno empresarial, cada vez se habla más de estos términos. Por ejemplo, hablando del sector logístico, la tecnología, puede garantizar que cumpla con las siete condiciones esenciales: ofrecer el producto o servicio adecuado, en el lugar adecuado, en la cantidad y calidad adecuadas, de la manera adecuada, en el momento adecuado, para el cliente adecuado y al coste adecuado. Para ello, en este ámbito, se utilizan predominantemente técnicas de Machine Learning para predecir compras y ventas, así como métodos de optimización numérica para planificar y tomar decisiones eficientes. Siguiendo el ejemplo del sector logístico, la calidad de los datos utilizados para crear un modelo de Inteligencia Artificial es crucial. Aunque los algoritmos pueden ser similares, su configuración varía en función del objetivo, el área de aplicación y los datos disponibles. Para el ejemplo de la gestión de inventarios, la selección de datos de calidad es fundamental, y aquí es donde entra en juego el conocimiento del negocio. Un buen proyecto debe comprender el problema a resolver, identificar los factores más relevantes y comunicar esta información al modelo. Por tanto, la colaboración entre expertos en IA y profesionales con conocimientos del negocio es esencial para obtener resultados óptimos. Toda la estructura del modelo se basa en esta información.

El Machine Learning y la optimización son especialmente útiles cuando se dispone de una gran cantidad de datos que pueden representar situaciones recurrentes. Por ejemplo, el historial de ventas proporciona información valiosa para predecir ventas futuras, siendo útil tanto para modelos predictivos como prescriptivos.

Hablando sobre la aplicación de técnicas predictivas y prescriptivas es fundamental para alcanzar resultados óptimos. La elección tanto de los datos como de las herramientas de Inteligencia Artificial adecuadas depende directamente de los objetivos específicos y los resultados deseados.

La analítica predictiva va más allá de simplemente mostrar lo que ha ocurrido, como lo hace la analítica descriptiva. En cambio, se enfoca en prever lo que sucederá en el futuro. Utilizando técnicas de Machine Learning, esta modalidad analítica examina datos pasados para estimar eventos futuros, como ventas futuras, comportamiento del cliente, entre otros. Estas predicciones son esenciales para la gestión eficiente del inventario, permitiendo una planificación más precisa.

Por otro lado, la analítica prescriptiva utiliza estas predicciones como base para determinar la mejor estrategia de acción. Se centra en la optimización, simulando diversos escenarios para evaluar qué acciones y decisiones generarán los mejores resultados.En última instancia, el objetivo principal es optimizar la gestión del inventario para maximizar los beneficios o minimizar los costos. Para lograr esto, es crucial considerar una serie de restricciones, como los niveles máximos de stock, costos de transporte, tiempos de entrega, entre otros. Estas restricciones se formalizan y se utilizan para encontrar automáticamente la solución óptima o la mejor solución posible.

Centrandose en la aplicación de la inteligencia artificial en la gestión de stock, se puede concluir que esta tiene múltiples aplicaciones, abarcando diversas áreas logísticas. Los casos de uso que se mencionan a continuación demuestran cómo puede impactar directamente en la gestión de inventarios y otras etapas del proceso logístico:

Gestión de la Demanda: Predicción de ventas utilizando datos externos como ubicación de tiendas y estacionalidad.

Segmentación de mercado y asociación de nuevos clientes a perfiles existentes.

Gestión del Abastecimiento y del Inventario: Optimización de calendarios de pedidos a proveedores considerando restricciones como costo de almacenamiento y tiempo de entrega.

Manufactura y Capacidad Operativa: Selección personalizada de recursos humanos y predicción de rotación o absentismo. Además de la optimización de turnos y medición del sentimiento de los empleados.

Gestión del Cumplimiento: Mejora de operaciones en almacenes, incluyendo recepción, distribución y robotización en empaquetado y carga. Y desarrollo de modelos para optimización de pedidos, planificación de rutas y gestión de devoluciones.

Estos casos pueden materializarse en proyectos reales que combinan análisis predictivo y prescriptivo. <https://www.iic.uam.es/noticias/machine-learning-y-optimizacion-para-gestion-stock/>

## <https://www.eaemadrid.com/es/blog/machine-learning> Descubre la Optimización de Inventario con Machine Learning ( Hemendik info ona alda atea) Geo beitukoet ze gehitu ta ze kendu

Vivimos un momento de rápidos cambios en el área tecnológica, donde todas las empresas deben poder reaccionar rápidamente. Cualquier organización necesita adaptarse cada vez más rápido a un mundo en constante cambio. Además, cada día aumenta la complejidad de la cadena de suministro. Por lo que no hay otra manera de que las empresas sobrevivan y se adapten rápidamente a esos cambios que automatizando procesos en su gestión de suministros. Y convertir procesos manuales en procesos en los que las máquinas son supervisadas por humanos. Si bien no se pueden resolver todos los desafíos simplemente añadiendo tecnología, una estrategia de cadena de suministro moderna requiere una pila de tecnología actualizada.

Las cadenas de suministro, tradicionalmente lineales y predecibles, enfrentan una importante fragmentación de la demanda debido al crecimiento del comercio digital y nuevos modelos de cumplimiento. Esta evolución está haciendo que las cadenas actuales se vuelvan obsoletas. Aunque nadie puede prever con certeza el futuro de la cadena de suministro, es claro que la inteligencia artificial jugará un papel fundamental para impulsar su resiliencia. Se espera una mayor visibilidad de extremo a extremo con sistemas más interconectados y un mayor uso de la IA y el machine learning para la previsión de la demanda y prácticas más sostenibles, respondiendo a las demandas de los consumidores. En un futuro cercano, las cadenas de suministro podrían ser altamente autónomas, con sistemas impulsados por IA que gestionen la mayoría de los procesos, desde la adquisición hasta la entrega. Además, es probable que el concepto de economía circular se adopte más ampliamente, reduciendo el desperdicio y optimizando el uso de los recursos.

El machine learning ha demostrado tener un gran potencial en varios ámbitos, incluida la configuración de optimización del control de inventario. La optimización de inventario es fundamental para la gestión eficiente de los minoristas, independientemente de su tamaño. Implica manejar una gran cantidad de productos de manera regular con el objetivo de reducir costos operativos y aumentar las ventas. Una parte esencial de este proceso es el control de inventario, que implica decidir cuándo y cuánto pedir de un artículo en particular para mantener un equilibrio óptimo entre oferta y demanda.

La gestión efectiva del inventario es crucial para las empresas, ya que impacta tanto en la disponibilidad del producto como en los costos operativos. Aunque reducir los tiempos de revisión y aumentar los pedidos puede mejorar la disponibilidad y reducir la pérdida de ventas, también puede aumentar los costos de inventario, incluidos los de mantenimiento y pedidos. Por ello, es esencial optimizar meticulosamente las políticas de inventario, ajustándolas según parámetros como el periodo de revisión, el tiempo de entrega y el nivel de servicio objetivo para cada artículo.

El Machine Learning, al aprovechar algoritmos avanzados para analizar datos, identificar patrones y hacer predicciones precisas, revoluciona la gestión de inventario. Puede manejar relaciones no lineales, incorporar factores externos como el clima o tendencias de redes sociales, y ajustar los pronósticos en tiempo real según las condiciones del mercado. Esto mejora la precisión de la previsión de la demanda y optimiza los niveles de inventario en consecuencia.

Con su capacidad para aprender y adaptarse a partir de datos, el Machine Learning es más dinámico y flexible que el software tradicional de gestión de inventario, especialmente en la previsión y optimización de la demanda. Puede manejar patrones de demanda complejos, identificar tendencias y ajustar pronósticos según múltiples variables y restricciones. Esto permite a las empresas determinar puntos de reabastecimiento óptimos, niveles de existencias de seguridad y estrategias de asignación de inventario, lo que resulta en decisiones de gestión de inventario más precisas, una mejor gestión de la cadena de suministro, y una reducción del riesgo de desabastecimiento o exceso de inventario.

Implementar Machine Learning en la gestión de inventario ofrece una serie de beneficios significativos: Por un lado, la Mejora de la Precisión de Pronósticos. Los algoritmos de Machine Learning permiten determinar niveles óptimos de inventario considerando factores como tiempo de entrega, estacionalidad y limitaciones de costos. Identifican el equilibrio adecuado entre costos de mantenimiento y desabastecimientos, optimizando los niveles de inventario a través del análisis de datos históricos, ciclos de producción y pronósticos de ventas. Por otro lado la Reducción de Costos y Pérdidas, es decir, analizan datos históricos de ventas, tendencias del mercado y factores externos para pronosticar con precisión la demanda de los clientes. Permite generar pronósticos de demanda más precisos, optimizando los niveles de inventario, reduciendo desabastecimientos y evitando excesos de inventario. Por último, también optimiza el ciclo de vida del producto. Los algoritmos de Machine Learning pueden generar probabilidades asociadas con diferentes niveles de demanda, permitiendo una gestión más eficiente de los productos.

Por ejemplo, en una tienda de comestibles, pueden analizar fechas de vencimiento, patrones de demanda y datos históricos de ventas para optimizar los niveles de stock y minimizar el desperdicio debido a la caducidad del producto.

Estas son varias consideraciones y pasos clave para implementar Machine Learning en la gestión de inventario. La evaluación de necesidades, es decir, reflexionar sobre objetivos claros y específicos, como evitar el exceso de existencias, prevenir desabastecimientos o mejorar la previsión de la demanda. La selección de Modelos y Algoritmos, seleccionado modelos como regresión lineal, árboles de decisión o redes neuronales según la complejidad del problema y los datos disponibles. Entrenar los modelos con datos históricos, evaluando su rendimiento con métricas adaptadas a los objetivos de gestión de inventario, y ajustar y optimizar el modelo para alcanzar un rendimiento óptimo. Y por último, la integración con Sistemas Existentes, Colaborando estrechamente entre analistas de datos, profesionales de IT y expertos en gestión de suministros para integrar el modelo en el sistema de gestión de inventario. Alinear los resultados del modelo con los procesos existentes y garantizar una integración fluida en la toma de decisiones. Establecer un proceso de monitoreo y mantenimiento constante, dado que los modelos de Machine Learning son dinámicos y requieren actualizaciones periódicas.

Empresas líderes como Amazon, Walmart y Nike han implementado con éxito el Machine Learning en la gestión de inventario, obteniendo resultados significativos. En el caso de Amazon por ejemplo, utiliza el Machine Learning para predecir la demanda de productos considerando diversas variables como tendencias de búsqueda, datos históricos de ventas y condiciones climáticas. Esto le permite mantener niveles óptimos de inventario, evitando excesos o faltantes de stock. Para el caso de Walmart, emplea un sistema de gestión de inventario impulsado por IA para proporcionar a los clientes lo que necesitan, cuando lo necesitan y al costo esperado. Combina datos históricos con análisis predictivos para colocar estratégicamente los artículos en centros de distribución y tiendas, optimizando la experiencia de compra. Y por último, Nike, está construyendo una cadena de suministro digital a nivel mundial para atender directamente a los consumidores a escala. <https://www.eaemadrid.com/es/blog/machine-learning>

Aquí ya mencionar un ejemplo de un modelo o dos y hacer un ejemplo simple con uno de ellos.

<https://rua.ua.es/dspace/bitstream/10045/120108/1/TFG_Carlos-Garrido-Marin.pdf> Puede ser interesante

**Aplicación de Modelos de Machine Learning para la Predicción de Demanda**:

Un Caso de Estudio

Después de revisar varios modelos de machine learning, se ha procedido a aplicar un caso real siguiendo un enfoque práctico y detallado. Para ilustrar este proceso, se ha seleccionado un modelo de los previamente analizados y se ha implementado para predecir la demanda utilizando datos históricos disponibles.

En este estudio, se ha optado por utilizar el modelo de árboles de decisión, conocido por su interpretabilidad y capacidad para manejar relaciones no lineales en los datos. El procedimiento seguido se detalla a continuación:

**Selección del Modelo**: Los árboles de decisión fueron elegidos por su simplicidad y facilidad de interpretación, además de su capacidad para manejar tanto variables categóricas como continuas sin necesidad de preprocesamiento extensivo.

**Preparación de los Datos**: Se recopilaron y organizaron los datos históricos de demanda, incluyendo variables relevantes como fechas, volúmenes de ventas, promociones y otros factores contextuales. Los datos fueron limpiados y preparados para ser introducidos en el modelo, asegurando su calidad y consistencia.

**Entrenamiento del Modelo**: El modelo de árboles de decisión fue entrenado utilizando los datos históricos. Durante esta fase, se aplicaron técnicas de validación cruzada y ajuste de hiperparámetros para optimizar el rendimiento del modelo y evitar el sobreajuste.

**Evaluación del Modelo**: Una vez entrenado, el modelo fue evaluado utilizando un conjunto de datos de prueba. Se calcularon métricas de rendimiento, como el error cuadrático medio (MSE) y el coeficiente de determinación (R²), para asegurar que el modelo proporcionaba predicciones precisas y confiables.

**Predicción de la Demanda**: Con el modelo validado, se procedió a predecir la demanda futura. Las predicciones se realizaron para un periodo específico, utilizando los datos disponibles para anticipar los niveles de demanda.

**Análisis de Resultados**: Los resultados de la predicción fueron comparados con los datos reales para evaluar la precisión del modelo. Se observó una alta correlación entre los valores predichos y los valores reales, indicando que el modelo de árboles de decisión es adecuado para la tarea de predicción de demanda en este contexto.

En conclusión, la aplicación práctica de un modelo de machine learning basado en árboles de decisión para la predicción de demanda ha demostrado ser efectiva y precisa. Este estudio no solo valida la capacidad del modelo seleccionado, sino que también proporciona una guía metodológica para la implementación de técnicas de machine learning en problemas reales de predicción de demanda.

## Objetivos

El objetivo de este proyecto es mejorar las prácticas de gestión de inventarios de Bibitor, LLC, una empresa minorista de vinos y licores que opera en múltiples ubicaciones. WBibitor, LLC gestiona ventas y costes de bienes vendidos que alcanzan los cientos de millones, lo que hace que las hojas de cálculo tradicionales resulten insuficientes. El vasto volumen de datos, que incluye millones de registros de ventas, compras e inventario, requiere un enfoque sofisticado para un análisis eficaz. El objetivo de este proyecto es aprovechar el análisis exhaustivo de datos para optimizar el control del inventario y extraer información valiosa del funcionamiento de la empresa, en particular de las ventas y las compras. Las tareas del proyecto se dividen en dos objetivos principales:

Analizar el proceso de gestión de inventarios y ofrecer recomendaciones para su optimización.

Extraer información empresarial significativa de los datos y ofrecer recomendaciones prácticas.

## Planificación

xxxxxxxx

# DESARROLLO

Xxxxxxxx

<https://www.kaggle.com/datasets/bhanupratapbiswas/inventory-analysis-case-study>

## Análisis de los datos

Teoría bezela:

El análisis de datos constituye un proceso fundamental en la interpretación y utilización efectiva de conjuntos de datos para obtener insights valiosos y respaldar la toma de decisiones informadas. Este proceso implica una serie de pasos interrelacionados que van desde la entrada de datos hasta la optimización de los resultados. A continuación, se han seguido los siguientes pasos para llevar a cabo el análisis.

* Entrada de Datos: Este paso implica la determinación de los requisitos de datos y la recopilación de información relevante. Se realiza mediante la interacción con stakeholders y la obtención de acceso a los datos necesarios.
* Preparación de Datos: La preparación de datos implica la limpieza y consolidación de datos brutos para transformarlos en una forma adecuada para el análisis. Se verifica continuamente para garantizar la calidad y relevancia de los datos.
* Exploración de Datos: En esta fase, se estudia el conjunto de datos a través de técnicas como el muestreo, análisis estadístico y visualizaciones para identificar patrones y tendencias que puedan conducir a insights valiosos.
* Enriquecimiento de Datos: Los datos se enriquecen con información adicional para proporcionar una perspectiva más completa y revelar insights que de otra manera podrían pasar desapercibidos.
* Ciencia de Datos: Se aplican métodos avanzados de análisis de datos, como algoritmos, aprendizaje automático e inteligencia artificial, para obtener insights más profundos y complejos que pueden ser difíciles de extraer con enfoques convencionales.
* Inteligencia Empresarial: La combinación de datos, software y procesos comerciales junto con la intuición humana permite obtener insights accionables que respaldan la toma de decisiones empresariales.
* Generación de Informes: Los resultados del análisis se organizan y presentan de manera efectiva para compartir los conocimientos adquiridos de manera comprensible y útil.
* Optimización: Se ajustan y mejoran continuamente los modelos y procesos de análisis para mantener su relevancia y eficacia en un entorno en constante cambio.

ESTA PODRIA SER LA RUTA A SEGUIR PARA EL APARTADO DE DESARROLLO.

<https://www.alteryx.com/es/glossary/data-analytics>

Para llevar a cabo la lectura de datos y el procesamiento, se ha optado por emplear el entorno de desarrollo integrado Visual Studio Code, aprovechando su funcionalidad y facilidad de uso en la programación en Python, un lenguaje ampliamente utilizado en el ámbito del análisis de datos y la ciencia de datos.

En la fase inicial del proceso, se ha procedido a importar las bibliotecas y paquetes necesarios para la manipulación, análisis y visualización de datos. Entre estas herramientas esenciales, se incluyen Pandas, una biblioteca de manipulación y análisis de datos que proporciona estructuras de datos flexibles y eficientes, así como funciones para la lectura y escritura de datos en diversos formatos; NumPy, una biblioteca fundamental para la computación numérica en Python, que ofrece soporte para matrices y funciones matemáticas de alto nivel; Matplotlib, una biblioteca ampliamente utilizada para la visualización de datos en dos dimensiones, que permite la creación de gráficos estáticos, interactivos y animados; y Seaborn, una biblioteca basada en Matplotlib que proporciona una interfaz de alto nivel para la creación de gráficos estadísticos atractivos y informativos.

La elección de estas herramientas se sustenta en su capacidad para facilitar las tareas de lectura, procesamiento y análisis de datos, así como en su versatilidad y potencia para generar visualizaciones claras y efectivas que permitan comprender y comunicar adecuadamente los resultados obtenidos durante el análisis de datos.

1. Exploración de datos

A continuación, se han cargado los 6archivos Excel con los que se dispone para llevar adelante el trabajo. Estos corresponden a operaciones comerciales del ejercicio cerrado en 2016. El conjunto de datos incluye las siguientes bases de datos: Inventario inicial del 2016, inventario final del 2016, facturas de compra correspondientes a 2016, precios de compra, datos de compras y datos de ventas. Una vez cargados los datos se han impreso las primeras 5lineas de todos los datos para obtener una visión general de los datos. Una vez leídos las bases de datos, se ha comenzado con el análisis exploratorio de ellos, obteniendo una tabla donde se han recogido detalles clave sobre los datos. Entre ellos, se han recogido los nombres de las columnas, el número de filas, el tipo de los datos, el número de valores únicos de cada columna, los valores no informados y el porcentaje de valores no informados frente al total. Así con la foto clave de cada tabla se han sacado varias conclusiones para después proceder a la limpieza y preprocesamiento de los datos.

Se ha visto que la columna *InventoryId* tiene los mismos registros únicos que los totales, lo que sugiere que podrían servir como claves primaria y foránea para unir las tablas. También se ha visto que varias tablas contienen valores no informados que es necesario identificar y tratar adecuadamente. Además, las columnas de tamaño muestran incoherencias en todas las tablas debido a la variación de unidades de medida. Por ello, es necesario convertirlas a una unidad única y coherente. Además de la columna de volumen, en alguna de las tablas también existe la columna tamaño, la cual contiene el mismo dato, por lo que se puede eliminar una de las dos columnas. Hablando del formato de las fechas se ha observado que utilizan diferentes formatos en cada una de las tablas por lo que es conveniente unificar y ponerlos de la misma manera. Y por último, en la columna *VendorName* se pueden ver espacios o signos tras la palabra que se pueden eliminar.

1. Limpieza y procesamiento

En el apartado anterior se han mencionado cuales son los puntos que se deben abordar para llevar a cabo la limpieza y el procesamiento de los datos, por lo que a continuación se han producido dichos cambios.

* 1. Tratamiento de Missings

La integridad y la calidad de los datos son aspectos fundamentales en cualquier análisis de datos o proyecto de investigación. Los valores no informados, comúnmente conocidos como "missing values" en inglés, representan una preocupación significativa en el contexto del análisis de datos, ya que pueden distorsionar los resultados y conducir a interpretaciones erróneas si no se manejan adecuadamente. Por lo tanto, es crucial detectar y abordar de manera efectiva estos valores faltantes.

Los valores no informados pueden surgir debido a una variedad de razones, que van desde errores en la entrada de datos hasta fallos en la recopilación o transferencia de información. Independientemente de su origen, la presencia de valores no informados puede comprometer la validez y la fiabilidad de cualquier análisis realizado sobre los datos afectados.

Durante el análisis del conjunto de datos que comprende varias tablas, se han detectado valores no informados en cuatro de las tablas analizadas. A continuación, se describen las acciones tomadas para abordar estos valores faltantes:

* Tabla “end\_inventory” : En la columna “City se ha identificado la ausencia de valores. Tras un análisis detallado, se ha observado que todos los nombres de ciudades están presentes excepto para una ubicación de tienda específica. Dado este hallazgo, se ha decidido asignar el nombre "TYWARDREATH" a esta ciudad, basándose en el número de tienda correspondiente.
* Tabla “purchases”: Se han encontrado solo tres valores faltantes en la columna "size", lo que representa un porcentaje mínimo del total de registros. Por lo tanto, se ha optado por por eliminar estas filas para preservar la integridad de los datos restantes.
* Tabla “sales”: En la columna "Approval" se muestra un alto porcentaje de valores faltantes, lo que representa el 93% del total de registros. Dada la magnitud de esta ausencia de información, se ha tomado la decisión de eliminar completamente esta columna para evitar cualquier sesgo o distorsión en el análisis posterior.
* Tabla purchase\_prices”: Se han identificado valores faltantes en las columnas "Description", "Size" y "Volume". Dado que estos valores no informados representan registros individuales, se ha decidido eliminar las filas correspondientes para mantener la coherencia y la integridad de los datos restantes.
  1. Tratamiento de duplicados

Mediante el método duplicated en Python, se ha realizado una verificación para determinar la presencia de duplicados dentro del conjunto de datos representado. Se ha guardado la información en una variable llamada duplicados, la cual captura el resultado booleano, proporcionando una indicación clara sobre la presencia o ausencia de duplicados en los datos. Este enfoque es esencial para garantizar la integridad y la calidad de los datos, ya que la presencia de duplicados puede distorsionar los resultados del análisis y conducir a conclusiones erróneas si no se manejan adecuadamente.

* 1. Irregularidades en los datos

Se han observado irregularidades significativas en las entradas de datos, especialmente en la columna "Size", que exhibe variaciones en todas las tablas de datos, con la excepción del conjunto de datos "purchase\_price". Estas discrepancias se refieren a la representación de información de volumen, expresada en diferentes unidades y formatos, como litros, mililitros, onzas, paquetes (pk) y combinaciones de estas unidades. Este panorama heterogéneo implica la necesidad de estandarizar estas medidas en una unidad de volumen única y coherente.

Para abordar esta cuestión, se ha llevado a cabo un análisis exhaustivo de los recuentos únicos en la columna "Size". Aquellos elementos que exhiben un solo recuento y no presentan un patrón discernible han sido asignados a una tasa equivalente estándar. Por otro lado, aquellos elementos con múltiples recuentos y un patrón discernible han sido transformados de acuerdo con estos patrones identificados.

Este enfoque metodológico busca garantizar la consistencia y la comparabilidad de los datos de volumen a lo largo de las diferentes tablas, lo que es esencial para facilitar un análisis coherente y significativo en el contexto de la investigación o análisis de datos en cuestión. La estandarización de las unidades de volumen contribuye a mitigar posibles sesgos o distorsiones en los resultados del análisis, permitiendo así una interpretación precisa y fiable de los datos.

Además, para garantizar la coherencia y la fiabilidad en el manejo de datos temporales, se ha llevado a cabo una normalización del formato de fecha y hora en un DataFrame específico. Es importante destacar que las columnas de fecha en los conjuntos de datos de "Inventario inicial" e "Inventario final" ya se encuentran en un formato adecuado, por lo que no requieren ninguna manipulación adicional. En consecuencia, nos enfocaremos en formatear las columnas de fecha en los demás conjuntos de datos.

Este proceso de normalización del formato de fecha y hora tiene como objetivo principal garantizar la cohesión y la consistencia en el tratamiento de datos temporales, lo que facilita su posterior análisis y visualización. Al estandarizar el formato de fecha y hora en todas las columnas pertinentes del DataFrame, se promueve la precisión y la interpretabilidad de los datos, lo que es esencial para obtener conclusiones sólidas y confiables en cualquier análisis posterior.

Por último, se ha identificado una discrepancia en los nombres de las ciudades presentes en las tablas de "Inventario inicial" y "Inventario final", donde los nombres no corresponden a ubicaciones geográficas reales. Para abordar esta irregularidad, se ha realizado un proceso de corrección mediante el reemplazo de los nombres de las ciudades no reales por aquellos que sí lo son.

Para llevar a cabo esta corrección, se han extraído todos los nombres de ciudades presentes en las tablas y mediante la función replace se han remplazado con una lista de ciudades geográficamente válidas.

Además, se ha observado que el identificador de inventario (InventoryId) estaba influenciado por el nombre de la ciudad. Por lo tanto, se ha procedido a recrear este identificador combinando el número de tienda, el nombre de la ciudad y el número de marca, separados por un guion. Este proceso de recreación del identificador se ha aplicado en las tablas pertinentes donde se encontraba esta relación entre la ciudad y el InventoryId.

Este enfoque de corrección y recreación de registros asegura la coherencia y la precisión de los datos, lo que facilita su interpretación y análisis posteriores. Además, garantiza que los identificadores de inventario reflejen de manera precisa la relación entre la tienda, la ciudad y la bodega, proporcionando así una representación más fiel de la realidad en el conjunto de datos.

Para terminar, se han impreso otra vez todas las tablas, y se ha asegurado que se ha completado la limpieza y procesamiento de los datos, para así empezar con la siguiente fase.

## Análisis y optimización del inventario

Realizar un Análisis de Control de Inventario basado en parámetros como Tiempo de Entrega (LT), Stock de Seguridad, Punto de Reorden y Análisis ABC.

Extraiga información empresarial significativa relacionada con el inventario respondiendo a las siguientes preguntas:

* ¿Qué productos tienen recuentos de inventario elevados al final del período de inventario y pertenecen a la clase A en el análisis de inventario ABC?
* ¿Qué productos tienen un recuento de existencias bajo al final del periodo de inventario, registraron ventas el año anterior y pertenecen a la clase C en el análisis de inventario ABC?
* Identifique los productos con grandes inventarios pero bajas ventas anuales y, a la inversa, utilice la relación entre las ventas anuales y el inventario final para este análisis.
* ¿Los productos con mayores existencias de seguridad registran mayores volúmenes de ventas, y a qué categoría del análisis ABC pertenecen?
* Enumere los 10 principales productos que requieren un reaprovisionamiento inmediato junto con sus cantidades de reaprovisionamiento.
* Determine la contribución de cada categoría a los ingresos totales basándose en el análisis de inventario ABC.
* Identifique los 10 productos principales de cada categoría que más contribuyen a los ingresos totales anuales.
* Identificar los 10 productos de cada categoría que menos contribuyen a los ingresos totales anuales.

Parte III: extracción de información útil sobre compras y ventas

Extraiga información empresarial significativa relacionada con las compras y las ventas respondiendo a las siguientes preguntas:

* ¿Quiénes son los vendedores o proveedores importantes? Evalúelos en función de la cantidad de artículos suministrados, el valor de los mismos y su contribución a los ingresos totales de la empresa. vendedores más importantes en términos de cantidad y valor de los suministros.
* ¿Quiénes son los vendedores o proveedores menos importantes? Evalúelos en función de la cantidad de artículos suministrados, el valor de los artículos y su contribución a los ingresos totales de la empresa. proveedores más importantes en términos de cantidad y valor de los suministros
* ¿Cuál es el plazo medio de ejecución de los pagos de cada proveedor y cómo se correlaciona con la cantidad o el valor de la compra?
* ¿Qué productos se venden más en términos de cantidad e importe de ventas en dólares y, a la inversa, cuáles se venden menos?
* ¿Cómo se comparan los resultados de ventas de las distintas tiendas y cuáles son las 10 tiendas con mayor rendimiento de ventas?
* ¿Qué tamaño de producto se vende más?
* ¿Podemos proyectar e identificar las tendencias de venta a lo largo del año?

POWERBI:

<https://learn.microsoft.com/es-es/power-bi/create-reports/service-dashboards-design-tips>

**Modelo semántico en powerbi:**

La creación de un modelo semántico excelente es una de las tareas más importantes ya que al hacer este trabajo de forma correcta, ayudará a los usuarios a comprender mejor los datos, lo que facilitará la creación de informes de Power BI útiles para ellos y para el creador.

Un buen modelo semántico ofrece las siguientes ventajas: una exploración de datos más rápida, agregaciones más fáciles de crear, informes más precisos, dedicar menos tiempo al crear el informe y mejor seguimiento a futuro.

<https://learn.microsoft.com/es-es/training/modules/design-model-power-bi/1-introduction>

A la hora de crear el modelo, se ha optado por el modelo estrella.

Esquemas de estrella

Puede diseñar un esquema de estrella para simplificar los datos. No es la única manera de simplificarlos, pero es un método popular; por tanto, todos los analista de datos de Power BI deben comprenderlo. En un esquema de estrella, cada tabla del modelo semántico se define como una tabla de hechos o de dimensiones, como se muestra en el siguiente objeto visual.

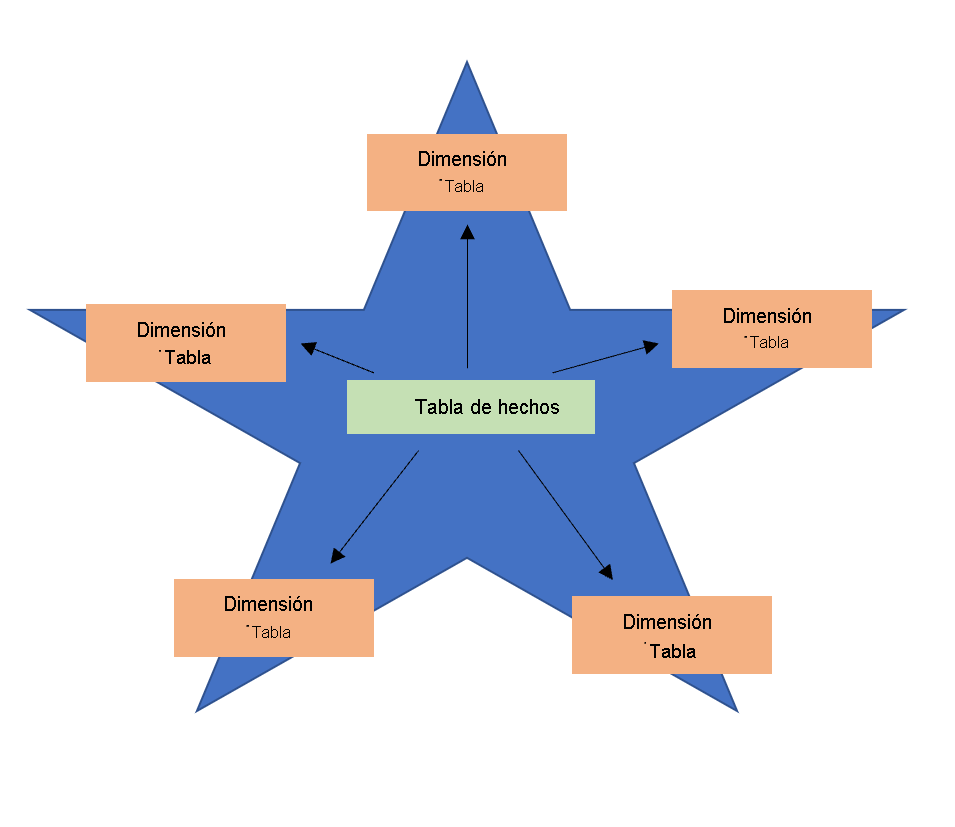


Ilustración del esquema de estrella con una tabla de hechos en el centro y tablas de dimensiones en cada una de las cinco puntas.

Las tablas de hechos contienen valores de datos de eventos o de observación: pedidos de ventas, recuentos de productos, precios, fechas y horas de transacciones, y cantidades. Las tablas de hechos pueden contener varios valores repetidos. Por ejemplo, un producto puede aparecer varias veces en varias filas para diferentes clientes en fechas distintas. Estos valores se pueden sumar para crear objetos visuales. Por ejemplo, un objeto visual del total de pedidos de ventas es una suma de todos los pedidos de ventas en la tabla de hechos. Con las tablas de hechos, es habitual ver columnas rellenadas con números y fechas. Los números pueden ser unidades de medida, como el importe de venta, o pueden ser claves, como un identificador de cliente. Las fechas representan el tiempo que se registra, como la fecha del pedido o la del envío.

Las tablas de dimensiones contienen los detalles sobre los datos de las tablas de hechos: productos, ubicaciones, empleados y tipos de pedido. Estas tablas están conectadas a la tabla de hechos a través de columnas de clave. Las tablas de dimensiones se usan para filtrar y agrupar los datos de las tablas de hechos. Las tablas de hechos, por otro lado, contienen los datos medibles, como ventas e ingresos, y cada fila representa una combinación única de valores de las tablas de dimensiones.

Las tablas de hechos son mucho más grandes que las de dimensiones, porque en ellas se producen numerosos eventos, como ventas individuales. Las tablas de dimensiones suelen ser más pequeñas porque está limitado al número de elementos que puede filtrar y agrupar. Por ejemplo, un año contiene solamente un número fijo de meses y Estados Unidos se compone de un número concreto de estados.

Por ello es muy importante la creación correcta de las tablas, escogiendo únicamente la información necesaria y mostrando de una manera ordenada. Así, como se ha mencionado anterior mente se han creado las siguientes tablas:

M\_CLIENTES

M\_PROVEEDORES

M\_PRODUCTO

M\_TIENDAS

M\_CALENDARIO (<https://learn.microsoft.com/es-es/training/modules/design-model-power-bi/3-date-table>) \* Explicación beyan

H\_VENTAS

H\_COMPRAS

H\_INVENTARIO\_INICIAL

H\_INVENTARIO FINAL

(Explicar la información que se recoge en cada tabla)

Creación de una tabla de fechas común

Se puede crear una tabla de fechas común de estas formas:Datos de origen,DAX,Power Query.

En este caso se ha decidido crear la tabla de calendario mediante DAX.

Puede usar las funciones de expresión de análisis de datos (DAX) CALENDARAUTO() o CALENDAR() para crear la tabla de fechas común. CALENDAR() devuelve un intervalo de fechas contiguo en función de fechas de inicio y de finalización que se especifican como argumentos en la función. Como alternativa, la función CALENDARAUTO() devuelve un intervalo de fechas completo y contiguo que se determina de forma automática a partir del modelo semántico. La fecha de inicio se elige como la más temprana del modelo semántico y la fecha de finalización es la última, además de los datos que se han rellenado para el mes fiscal, que puede optar por incluir como argumento en la función CALENDARAUTO(). Para los fines de este ejemplo, se usa la función CALENDAR() porque solo quiere ver los datos desde el 31 de mayo de 2011 (el primer día en el que las ventas empezaron a realizar el seguimiento de estos datos) en adelante, durante los próximos 10 años. En Power BI Desktop, seleccione **Nueva tabla** y escriba la siguiente fórmula DAX:

Ahora tiene una columna de fechas que puede usar. Pero esta columna es ligeramente dispersa. También quiere ver columnas solo del año, el número del mes, la semana del año y el día de la semana. Puede realizar esta tarea si selecciona **Nueva columna** en la cinta y escribe la ecuación DAX siguiente, que recuperará el año de la tabla de fechas.

Para el DATA EXPLORATORY AND CLEANING QUE SE HA HECHO HASTA AHORA SE HA UTILIZADO: https://www.kaggle.com/code/abdulmelikhmeda/inventory-purchase-sales-analysis-and-optimization

Para compras y ventas (gráficos)

**https://www.kaggle.com/code/bahadir23/inventory-optimization-and-sustainability-analysis**

## Configuración del modelo semántico y creación de relaciones entre tablas

Aunque se puede realizar el procesos de limpieza de datos en PowerQuery, en este proyecto se ha decido hacerlo mediante visual studio. Por ello, las tablas que se han ingestado en PowerBI cuentan con datos limpios. Así, en el apartado modelo se ha comenzado a crear las relaciones necesarias. (EXPLICACION DE LA CREACION DE RELACIONES)

# RESULTADOS

Xxxx

## Informe en PowerBI

# CONCLUSIONES Y DISCUSIÓN

xxxxxxxx

**BIBLIOGRAFÍA**

|  |
| --- |
| Anexos / Eranskinak |

