

Propuesta y Organización CRISP-DM

1. Business Understanding (Entendimiento del Negocio)

Problema de Negocio:

Las empresas que gestionan grandes volúmenes de productos, como en el sector de ventas de bicicletas, enfrentan constantemente el reto de mantener un equilibrio óptimo entre la oferta y la demanda. Un exceso de inventario implica altos costos de almacenamiento, mientras que una falta de productos implica pérdida de ventas y clientes. El objetivo de este proyecto es desarrollar un sistema predictivo que permita anticipar las ventas futuras y ajustar los niveles de inventario de manera eficiente.

Pregunta de Investigación:

¿Cómo podemos predecir con precisión las ventas futuras para optimizar el inventario disponible y mejorar la planificación de la cadena de suministro?

Hipótesis Principal:

Si analizamos las tendencias históricas de ventas, considerando variables como el producto, fecha, región y canal de venta, entonces podemos construir un modelo que permita anticipar las ventas futuras con una precisión significativa.

Valor del Producto de Datos (MVP):

Un sistema de predicción de ventas integrado con el inventario permitirá a la empresa tomar decisiones informadas sobre compras, almacenamiento y distribución, lo cual puede traducirse en reducción de costos, mejor disponibilidad de productos y mayor satisfacción del cliente.

2. Data Understanding (Entendimiento de los Datos)

Origen de los Datos:

El dataset utilizado es el 'SAP Bikes Sales sample data', un conjunto de datos de muestra extraído del sistema contable SAP ERP y disponible públicamente en Kaggle. Representa información estructurada sobre ventas de bicicletas por fecha, categoría, región, canal de distribución, entre otros.

Frecuencia y Formato:

Los datos están organizados a nivel de transacción de ventas, con campos de fecha, producto, cantidad, ingresos, categoría, subcategoría, canal, etc. El rango temporal cubierto depende del dataset (por verificar tras la carga).

Variables Relevantes:

- Fecha de venta
- Producto y categoría
- Cantidad vendida

- Ingresos generados
- Canal de venta

Posibles Desafíos:

- Datos faltantes o inconsistentes en fechas o cantidades
- Valores atípicos en ingresos o unidades vendidas
- Necesidad de transformar datos categóricos para modelado
- Conversión y estandarización de formatos (por ejemplo, fechas)

3. Data Preparation (Preparación de los Datos)

Plan de Limpieza y Transformación:

- Valores faltantes: Evaluar y aplicar imputación (media/moda) o eliminación según el caso.
- Outliers: Identificación mediante visualización (boxplots) y tratamiento razonable (cap, transformación logarítmica o exclusión).
- Conversión de tipos: Asegurar que fechas estén en formato datetime, cantidades e ingresos como numéricos.
- Variables derivadas: Crear nuevas variables como “mes”, “año”, “día de la semana” a partir de la fecha.
- Integración de fuentes: Si se añaden más datasets (por ejemplo, inventario o costos), realizar integración con claves comunes.
- Documentación: Cada paso será documentado en notebooks y scripts para garantizar reproducibilidad.