

Informe de Progreso

El presente informe detalla el progreso alcanzado en el proyecto, que tiene como objetivo desarrollar un modelo eficaz para predecir las ventas mensuales de un producto.

Competición [en Kaggle](#)

Análisis Inicial de la Relación de las Columnas en el data set

En una fase inicial, se realizó un análisis de correlación entre las columnas del conjunto de datos utilizando la función `.corr()` y se generó un mapa de calor con `sns.heatmap` para visualizar las relaciones entre las variables.

Se aplicó un filtro a los valores de correlación, considerando solo aquellos mayores de 0.7, lo que nos permitió identificar las relaciones más fuertes. Esto proporcionó un punto de partida sólido para el análisis individual de cada columna.

Preprocesamiento de Datos

Se abordaron diversos aspectos del preprocesamiento de datos para garantizar la calidad y consistencia de los mismos. Se tomaron las siguientes medidas:

- **Limpieza de Valores Nulos:** Se identificaron y manejaron los valores nulos en las columnas "Quan_X" al rellenarlos con la media de la respectiva columna, evitando así alteraciones significativas en las métricas.
- **Normalización de Nombres de Columnas:** Se corrigieron errores en la nomenclatura de las columnas, como la confusión entre "Quant" y "Quan," asegurando la coherencia en la nomenclatura.
- **Corrección de Datos Erróneos en Columnas "Cat_X":** Se identificaron datos incongruentes en las columnas "Cat_X" y se igualaron a cero aquellos valores mayores o iguales a dos, ya que no se ajustaban a los parámetros iniciales.
- **Eliminación de Columnas Irrelevantes:** Las columnas Cat_X que contenían valores constantes en todas sus entradas se consideraron irrelevantes y se eliminaron del conjunto de datos.
- **Manejo de Fechas:** Se realizó una corrección en las fechas "Date_2," asegurando que los valores nulos sean iguales al menor de "Date_2," lo cual resulta coherente con la descripción de los campos.

- **Cálculo de Medias en Ventas Mensuales:** Para los "Outcomes_MX" (ventas mensuales), se implementó el cálculo de la media de las ventas para no afectar las métricas de estos datos.

Conclusiones Preliminares

El análisis inicial de correlación y el preprocesamiento de datos han sentado las bases para el desarrollo de modelos predictivos. Se ha logrado una comprensión más profunda de la relación entre las columnas y se han abordado problemas comunes, como la presencia de valores nulos y datos erróneos.

El proyecto avanza de manera sólida, y estos primeros pasos sientan las bases para el desarrollo de modelos predictivos eficaces. En las próximas etapas, se procederá a la construcción y evaluación de modelos utilizando el conjunto de datos preprocesado.

Próximos Pasos:

- Construcción de modelos predictivos utilizando el conjunto de datos preprocesado.
- Evaluación de modelos y ajuste de hiperparámetros.
- Validación cruzada y selección del modelo final.
- Interpretación de resultados y presentación de hallazgos.