

Análisis Climático en la Plataforma Petrolífera Troll A: Datos del Año 2024

Limpieza y Exploración de Datos Meteorológicos de Troll A

Fase: Limpieza y transformación de datos

Descripción general del proyecto

Este proyecto se centra en la limpieza, exploración y visualización de datos meteorológicos históricos de la estación TROLL A OIL PLATFORM NO para el año 2024, utilizando el conjunto de datos **Global Summary of the Day (GSOD)** proporcionado por el NOAA. El objetivo principal es explorar y comprender los patrones de temperatura, velocidad del viento y otras variables climáticas relevantes en esta ubicación específica, aplicando técnicas de limpieza, análisis exploratorio y análisis estadístico con Python, Pandas, Matplotlib y Seaborn.

Problema:

Se identificaron varios problemas en el conjunto de datos original, incluyendo:

- **Valores atípicos (outliers):** Se detectaron outliers en las columnas 'VISIB', 'WDSP', 'GUST', 'MAX' y 'MIN'. Los outliers en 'MAX' y 'MIN' se consideraron errores de medición, mientras que los de 'VISIB', 'WDSP' y 'GUST' requerirán un análisis más profundo.
- **Unidades de medida:** Las temperaturas estaban en grados Fahrenheit y las velocidades del viento en nudos, lo que requería transformación a Celsius y metros por segundo, respectivamente, para un análisis coherente.
- **Columnas de atributos:** Se detectó que existían errores en las columnas de atributos (tipo de dato incorrecto).

Solución:

Para abordar los problemas identificados, se realizaron los siguientes pasos:

- **Análisis de columnas de atributos:** Se realizó el análisis de las columnas de atributos para detectar errores.
- **Se transformaron las unidades:** De Fahrenheit a Celsius, Decimas de millas a kilómetros y nudos a kilómetros por hora.
- **Eliminación de outliers:** Los outliers de las columnas 'MAX' y 'MIN' fueron eliminados debido a su alta probabilidad de ser errores.
- **Creación de conjuntos de datos:** Se crearon dos conjuntos de datos uno con outliers (troll_a_2024_con_outliers.csv) y otro sin outliers (troll_a_2024_sin_outliers.csv).

- **Imputación de valores faltantes:** La imputación de valores faltantes se realizó utilizando la mediana de cada columna ('VISIB', 'WDSP', y 'GUST'). La mediana es una medida de tendencia central robusta, especialmente en presencia de outliers, ya que no se ve afectada por valores extremos. Esta técnica es adecuada para este conjunto de datos, donde los outliers son comunes y podrían sesgar la media (se guardaron en el conjunto troll_a_2024_sin_outliers.csv).

Detalles:

Se obtuvieron 2 conjuntos de datos listos para realizar un análisis comparativo y determinar si los outliers afectan significativamente los resultados.

Próximos pasos:

En la siguiente fase del proyecto se realizará un análisis exploratorio de datos (EDA). Se utilizarán técnicas de visualización y estadísticas descriptivas para descubrir patrones, tendencias y relaciones entre las variables climáticas. Adicionalmente, se realizará un análisis comparativo entre los dataframes con y sin outliers.