Vishnu Kumar D S
Data Analyst

Task 2 Summary: Data Cleaning, Analysis & Insights

As part of Task 2, I worked through a detailed process of understanding, cleaning, and analyzing a dataset to uncover key insights and improve its overall usability. I began by exploring the dataset column by column — identifying data types, counting unique values, and understanding how each feature contributes to the bigger picture. This helped me recognize which columns were most valuable from a stakeholder perspective.

Next, I focused on cleaning the data. I addressed missing values using appropriate methods like imputation and handled text inconsistencies, such as typos and inconsistent capitalization. Numerical data was checked for accuracy, outliers were treated, and date formats were standardized to ensure consistency across the dataset.

From this cleaned data, I identified five critical columns that offered the most value — including failure reasons, component types, timestamps, location, and cost of failure. These were used to generate clear visualizations like bar charts, pie charts, and trend lines to help stakeholders quickly grasp key patterns and problem areas.

One of the more insightful parts of the task involved working with free-text fields. By processing this text using basic NLP techniques, I extracted meaningful tags such as "overheating" or "sensor failure," which helped summarize complex issues in a clear, structured format.

Finally, I summarized my findings into actionable insights — like focusing maintenance on specific components or prioritizing high-cost failures — and noted discrepancies I resolved during the cleaning phase. All work was supported with Python scripts and shared alongside the cleaned dataset and visuals.

This task provided a complete view from raw data to meaningful insight, all aimed at helping stakeholders make informed decisions.

GITHUB LINK : LINK

Email: vishnukumardsvishnu@gmail.com

Plots/ Graph