Titanic Data Analysis Project

Objective:

The objective of this project is to analyse the Titanic dataset to identify patterns and relations between passenger characteristics and survival probability.

Raw Dataset:

The raw dataset contains passenger information such as Name, Age, Sex, Passenger Class (Pclass), Embarked port, and whether the passenger survived.

Steps Performed:

1. Imported raw Titanic dataset into Jupyter Notebook.
2. Checked for missing values and handled them by imputing with median or mode.
3. Performed initial data exploration and statistical summary.
4. Applied data visualization using seaborn and matplotlib to explore relationships between variables.

Cleaned Dataset:

The cleaned dataset is free from missing values.

Visualizations:

Count Plots: Survival distribution by gender, Pclass, and Sex.

Pie Charts: Relationship between Embarked and survival status.

Histograms: Age distribution of survivors vs. non-survivors.

Conclusion:

The analysis of the Titanic dataset revealed several key insights into passenger survival patterns. Our findings indicate that survival rates were strongly influenced by socio-economic status, as passengers from higher classes (Pclass 1) had a significantly greater chance of survival compared to those from lower classes. Gender also played a critical role, with females having a much higher survival rate than males, reflecting the "women and children first" evacuation policy. Age distribution analysis showed that younger passengers, particularly children, were more likely to survive, while survival rates for older passengers declined. Additionally, the point of embarkation (Embarked) influenced survival, with passengers boarding from certain ports showing better chances of survival, likely due to class distribution differences. The cleaning and preprocessing of the dataset ensured accuracy in our visualizations and statistical findings, making the conclusions reliable for further research or predictive modeling.