SkillCraft Technology Internship

Task 2 Report: Data Cleaning & EDA

Intern: Suzain Bargir

Position: Data Science Intern

Objective

Theobjective of this task was to perform data cleaning and exploratory data analysis (EDA) on the Titanic dataset to find patterns and insights about passenger survival.

Steps Performed

1. DataLoadingandPreparation:

- Uploaded Titanic dataset (train.csv.zip) in Google Colab.
- Extracted and read the train.csv file using Pandas.

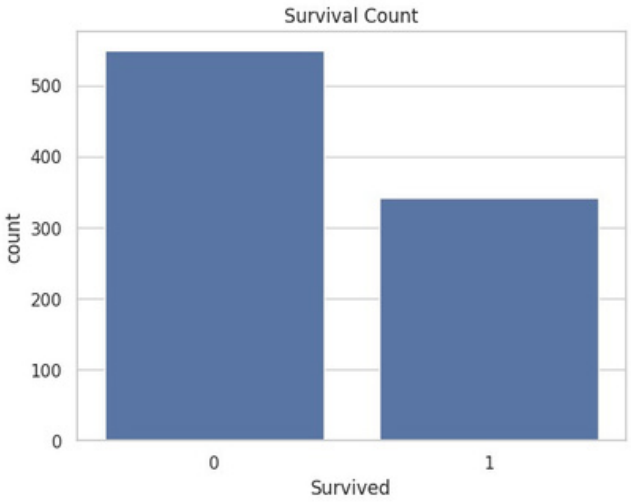
2. Data Cleaning:

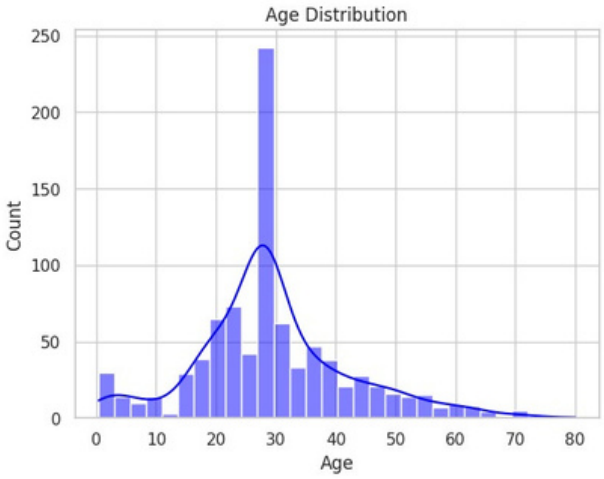
- Filled missing Age values with median.
- Filled missing Embarked values with most frequent category.
- Dropped Cabin column due to many missing values.
- Converted categorical variables (Sex, Embarked) to numeric.

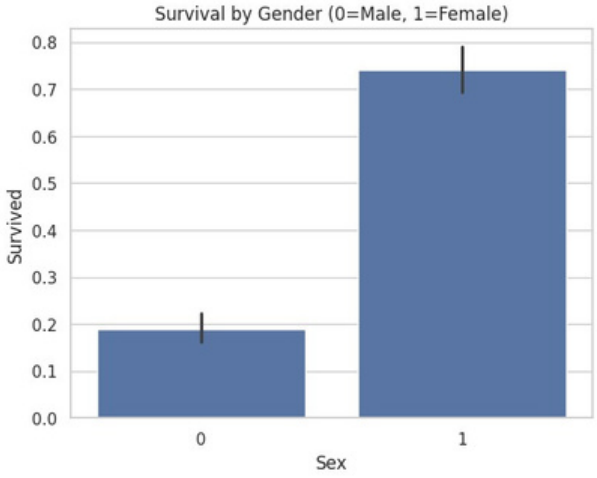
3. Exploratory Data Analysis:

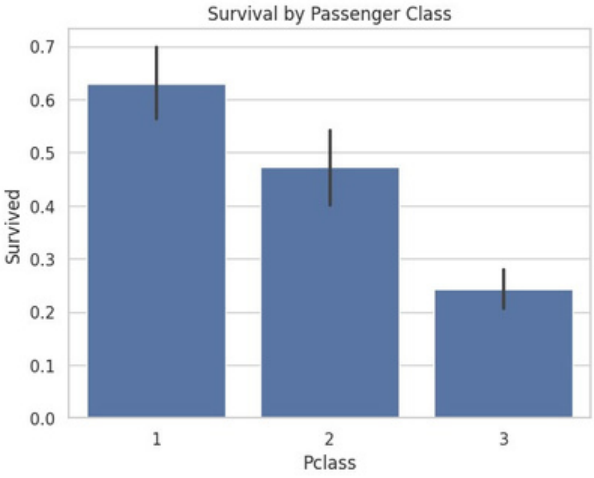
- Plotted survival count (overall survival ratio).
- Visualized age distribution of passengers.
- Analyzed survival rate by gender and passenger class.
- Generated correlation heatmap to find feature relationships.

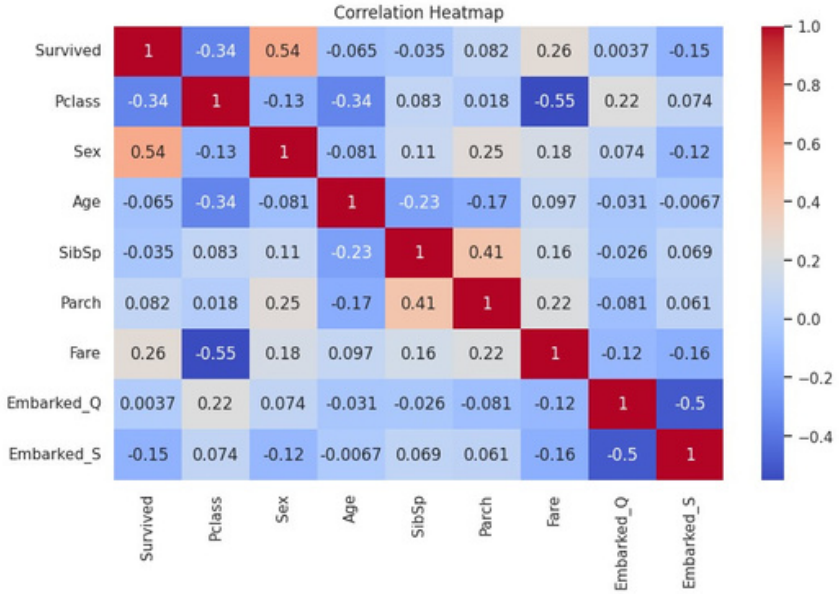
Graphs (Placeholders)











Insights and Conclusion

- Femalepassengershad a higher survival rate than males.
- Passengers in 1st class had better chances of survival compared to 2nd and 3rd class. - Most passengers were aged between 20 to 40 years. - Higher ticket fare showed a positive correlation with survival.

Conclusion: Gender, passenger class, and fare played a significant role in survival chances.