

Task 5 - Exploratory Data Analysis (EDA) Report

Objective

Perform Exploratory Data Analysis (EDA) to extract insights and visualize trends using the Titanic dataset. The aim is to understand the structure, relationships, and patterns in the data to aid further analysis or predictive modeling.

Tools Used

- Python
- Pandas
- Matplotlib
- Seaborn
- Jupyter Notebook

Key Steps Performed

1. Loaded Titanic dataset from Kaggle.
2. Inspected data types, structure, and missing values using `.info()`, `.describe()`, `.isnull().sum()`.
3. Conducted univariate analysis with histograms and boxplots.
4. Conducted bivariate analysis using bar plots and boxplots.
5. Conducted multivariate analysis using correlation matrix and heatmap.
6. Created visualizations to explore survival patterns by gender, class, and age.
7. Documented observations and summarized findings.

Visualizations Created

- Histogram: Age Distribution
- Boxplot: Fare Distribution
- Countplot: Survival by Gender
- Boxplot: Age by Pclass
- Heatmap: Feature Correlation
- Pairplot: Age, Fare, Pclass vs Survived

Task 5 - Exploratory Data Analysis (EDA) Report

Insights & Observations

- Females had a significantly higher survival rate than males.
- First-class passengers were more likely to survive compared to third-class passengers.
- Younger passengers had higher chances of survival.
- There is a moderate correlation between Fare and Pclass.
- Missing values were present in 'Age' and 'Cabin' columns.

Dataset Source

Titanic - Machine Learning from Disaster (<https://www.kaggle.com/c/titanic/data>)

Files Submitted

- Titanic_EDA.ipynb (Jupyter Notebook)
- Titanic_EDA.pdf (PDF Export of Notebook)
- README.md (Project Summary)

Conclusion

EDA helped uncover key patterns and relationships in the Titanic dataset. These insights can be used to inform further analysis or build predictive models, such as survival prediction.