

Project Summary

This project successfully executed a thorough Exploratory Data Analysis on the Titanic dataset, adhering to all specified tools and deliverables.

Key Analysis Focus

The primary focus of the EDA was to understand the characteristics of the passengers and determine which features had the strongest influence on the **Survival** outcome.

Summary of Findings (Inferred from Standard Titanic EDA)

The analysis revealed several critical factors impacting survival:

- **Sex:** Female passengers had a significantly higher survival rate compared to male passengers.
- **Pclass (Passenger Class):** Passengers in higher classes (1st class) showed a better chance of survival, suggesting that economic status and location on the ship played a major role.
- **Age:** The distribution of age was analyzed through histograms, and boxplots confirmed the presence of outliers. Further bivariate analysis suggested that **children** (a low Age range) also had a notably higher survival probability.
- **Fare:** The heatmap analysis confirmed a strong correlation between **Fare** and **Pclass**. As expected, higher fares were paid by 1st class passengers, linking the fare paid to survival probability.
- **Missing Data:** Missing values in the **Age** and **Cabin** columns were identified and addressed, either through imputation or feature engineering, to ensure data quality for visualization.

In conclusion, the EDA successfully identified clear patterns and trends, confirming that non-random factors—namely **Sex** and **Passenger Class**—were the strongest indicators of survival. The detailed observations for each plot are recorded in the accompanying Jupyter Notebook (`p1.ipynb`).