

TITANIC SURVIVAL DATA ANALYSIS REPORT

Project Title: Demographic and Class-Based Survival Analysis of Titanic Passengers

Tools Used: Python, Pandas, Seaborn, Matplotlib

Dataset: 891 Titanic passenger records

OBJECTIVE:

To explore demographic and socioeconomic factors influencing survival on the Titanic.

DATA CLEANING:

- Dropped sparse/irrelevant columns: 'Cabin', 'Ticket', 'PassengerId'
- Filled missing 'Age' with median and 'Embarked' with mode

KEY FINDINGS:

- Women had significantly higher survival rates than men.
- First-class passengers were more likely to survive than those in third class.
- Children and younger passengers had higher survival rates.
- Passengers who paid higher fares tended to survive more.
- Port of embarkation influenced survival (highest from Cherbourg).

VISUALIZATIONS:

- Countplots for Sex, Survived, Pclass, Embarked
- Histograms for Age and Fare
- Boxplot for Age vs Survival
- Heatmap for feature correlation

CONCLUSION:

This project demonstrates data cleaning, EDA, and storytelling with visuals-essential skills for entry-level data science roles. It offers a strong foundation for further predictive modeling or dashboard projects.