

Titanic Data Analytics Project – Executive Summary

Overview

This project analyzes the Titanic dataset to identify the factors that influenced passenger survival. The work involved data cleaning, feature engineering, exploratory data analysis (EDA), and insights reporting. No predictive modeling was applied; the focus was on understanding the data and communicating findings clearly.

Data Preparation

- Reviewed the dataset structure, distributions, and missing values.
- Filled in missing values for *Age* and *Embarked*; dropped *Cabin* due to excessive nulls.
- Converted categorical variables (*Sex*, *Embarked*) into numerical form.
- Created additional features to enhance analysis:
 - **FamilySize** = SibSp + Parch + 1
 - **IsAlone** = indicator for solo travelers
 - **AgeGroup** = Child, Teen, Adult, Senior
 - **Title** = extracted from passenger names

Exploratory Data Analysis

Gender

- Female survival rate: 74%
 - Male survival rate: 19%
- Women were about four times more likely to survive than men.

Passenger Class

- First Class: 63% survived
 - Second Class: 47% survived
 - Third Class: 24% survived
- Higher social classes were associated with greater survival chances.

Age

- Average age: 29.7 years

- Children (<12): 58% survived
 - Adults (20–49): 36% survived
 - Seniors (50+): 23% survived
- Younger passengers were more likely to survive.

Family Size

- Solo travelers: 30% survived
 - With family: 51% survived
- Passengers traveling with family had better survival rates.

Fare

- Lower fares: 19% survived
 - Higher fares: 62% survived
- Wealthier passengers had significantly higher survival chances.

Port of Embarkation

- Cherbourg: 55% survived
- Queenstown: 39% survived
- Southampton: 34% survived

Correlation Analysis

- **Sex** (−0.54) and **Passenger Class** (−0.34) had the strongest correlations with survival.
- **Fare** (+0.26) showed a positive relationship with survival.
- **Age** (−0.08) had only a weak negative correlation.

Key Insights

- Gender and passenger class were the most important survival factors.
- Women, children, and first-class passengers had the best survival outcomes.
- Family presence and higher fares improved chances of survival.
- Economic and social status were decisive in shaping survival odds.

Skills Practiced

- Data cleaning and preprocessing with Pandas and NumPy
- Exploratory data analysis with Matplotlib and Seaborn
- Feature engineering and handling missing values
- Visualizing and reporting insights in a clear, accessible way

Conclusion

This project demonstrates the complete workflow of a data analyst: progressing from raw data to well-structured insights. Even without machine learning, the analysis reveals how social, economic, and demographic factors strongly influenced survival on the Titanic.