Titanic Dataset - Exploratory Data Analysis (EDA)

This report presents an Exploratory Data Analysis (EDA) of the Titanic dataset. We investigate demographics, socio-economic factors, and their relationship with survival outcomes.

Dataset Overview

- Rows: 891

- Columns: 12

- Target Variable: Survived (0 = No, 1 = Yes)

- Missing Values: Age (177), Cabin (687), Embarked (2)

Key Findings

1. Survival Distribution: About 38% of passengers survived.

2. Gender Effect: Women had a much higher survival rate than men.

3. Passenger Class: 1st class passengers had the best survival chances, 3rd class the worst.

4. Age: Most passengers were between 20-40 years.

5. Fare: Skewed distribution. Higher fares linked to better survival rates.

6. Correlation: Negative correlation between Pclass and Survival. Fare positively correlated with Survival.

Visual Summary

- Countplots show survival variation by gender and passenger class.

- Histogram shows age distribution centered around 29 years.

- Fare distribution is right-skewed with a few very high values.

- Heatmap highlights correlations: class and fare are significant factors for survival.

Conclusion

The EDA reveals strong survival differences across gender, class, and fare. Women and first-class passengers had much higher chances of survival. The dataset also highlights missing data issues (Cabin, Age) that should be handled in predictive modeling.