# Titanic Dataset EDA Report

## 📌 Dataset Overview

The Titanic dataset contains information about passengers such as age, gender, passenger class, fare, and survival status.
The dataset includes 891 records and 12 variables, with Survived as the target variable.

## 📌 Data Quality Observations

- Missing values were found mainly in the Age and Cabin columns.
- The Cabin column contains a large number of missing values, making it less useful for direct analysis.
- Numerical columns such as Age and Fare contain outliers, as observed from boxplots.
- Categorical columns like Sex and Pclass are well defined and suitable for analysis.

## 📌 Survival Analysis

- A higher number of passengers did not survive compared to those who survived.
- Survival outcomes are not evenly distributed, indicating the influence of other factors.

## 📌 Gender Based Insights

- Female passengers had a significantly higher survival rate compared to male passengers.
- Gender is one of the most influential factors affecting survival.

## 📌 Passenger Class Insights

- Passengers traveling in First Class (Pclass = 1) had the highest survival rate.
- Third Class passengers had the lowest survival rate.
- This suggests that socio economic status played a major role in survival probability.

📌 Age Based Insights

- Most passengers were between 20 and 40 years of age.
- Younger passengers generally had a better chance of survival.
- Elderly passengers showed lower survival rates.

📌 Fare Analysis

- Passengers who paid higher fares had better survival outcomes.
- Fare shows a positive correlation with survival, indicating that wealthier passengers were prioritized.

📌 Correlation Analysis

- The correlation heatmap revealed:
- A positive correlation between Fare and Survival
- A negative correlation between Pclass and Survival
- These correlations support observations from visual analysis.

📌 Key Insights Summary

- Gender, passenger class, and fare strongly influence survival.
- Female and first class passengers had higher survival probabilities.
- Younger age groups were more likely to survive.
- The dataset contains missing values and outliers that should be handled before modeling.

## 📌 Conclusion

- Exploratory Data Analysis of the Titanic dataset revealed clear and meaningful patterns related to passenger survival.
- EDA helped in understanding the structure, quality, and relationships within the data.
- These findings provide a strong foundation for further steps such as data preprocessing, feature engineering, and predictive modeling.