# Exploratory Data Analysis (EDA) Report

### *Dataset: Titanic Dataset*

Objective: Understand the dataset using statistical summaries and visualizations to identify patterns, trends, and anomalies before model building.

## 1. Summary Statistics

- Calculated mean, median, and standard deviation for numerical features.
- Observed age distribution and fare spread.
- Identified passenger class distribution.
- Noted right-skewness in Fare column.

## 2. Distribution Analysis (Histograms & Boxplots)

- Histograms revealed Age is slightly right-skewed.
- Fare showed heavy right-skew with extreme outliers.
- Boxplots confirmed presence of high-value fare anomalies.
- Majority passengers belonged to 3rd class.

## 3. Survival Analysis

- Overall survival rate was less than 50%.
- Female passengers had significantly higher survival rates than males.
- Higher-class passengers had better survival probability.
- Higher fare correlated positively with survival.

## 4. Correlation & Feature Relationships

- Correlation matrix showed negative correlation between Passenger Class and Survival.
- Fare had positive correlation with Survival.
- Age showed weak correlation with Survival.
- Pairplots indicated clearer separation in Fare vs Survival.

# Conclusion

Exploratory Data Analysis provided critical insights into feature distributions, outliers, and survival patterns. Key predictive features identified include Sex, Passenger Class, and Fare. These findings establish a strong foundation for subsequent machine learning model development.