Exploratory Data Analysis on the Titanic Dataset

The Titanic dataset contains passenger details such as age, gender, class, and survival status. The objective of this analysis is to explore the dataset statistically and visually to identify key trends, patterns, and anomalies.

Dataset Overview:
- Rows: 891
- Columns: 12
- Target Variable: Survived (0 = No, 1 = Yes)
- Contains missing values in Age, Cabin, and Embarked.

Univariate Analysis:
- Survival Count: Majority did not survive (~62%).
- Gender Distribution: Male passengers dominated (~65%), females (~35%).
- Age Distribution: Most passengers were between 20–40 years. Some outliers above 70.
- Fare Distribution: Positively skewed; most paid low fares, with few very high values.

Bivariate Analysis:
- Survival by Gender: Females (~74%) survived more than males (~19%).
- Survival by Class: 1st Class (~63%) survived most; 3rd Class (~24%) survived least.
- Age vs. Survival: Younger children survived more; elderly survival was rare.

Multivariate Analysis:
- Survival by Class & Gender: Female 1st Class had highest survival, Male 3rd Class had lowest.
- Correlation: Pclass negatively correlated with survival; Fare positively correlated.
- Pairplot Insights: Higher fares linked with survival; overlap in ages.

Key Findings:
1. Gender is the strongest survival factor – women survived at higher rates.
2. Passenger Class influenced survival chances – wealthier passengers had better chances.
3. Fare also played a role – higher fare associated with survival.
4. Children had slightly better chances, elderly survival was rare.
5. Missing values (especially in Cabin) require preprocessing.

Conclusion:
This analysis highlights the importance of demographic and socio-economic factors in survival on the Titanic. These findings can serve as a foundation for predictive modeling, where gender, class, and fare would likely be the most important features.