

# Titanic EDA Observations

## 1. **Missing Data:**

- 'Age', 'Cabin', and 'Embarked' have missing values.
- Cabin has too many missing entries to be useful directly.

## 2. **Univariate Insights:**

- Most passengers were in the age group of 20-40.
- Fare distribution is right-skewed with most fares under 50.
- Class distribution: Most passengers were in 3rd class.

## 3. **Survival Rate:**

- Overall survival rate is around 38%.
- Females had a significantly higher survival rate than males.

## 4. **Multivariate Observations:**

- Passengers in 1st class had better survival chances.
- Higher fares were generally paid by survivors.
- Age didn't have a strong linear correlation with survival.

## 5. **Correlations:**

- Strongest negative correlation with survival: Pclass (-0.34)
- Fare and survival have a positive correlation.

## 6. **Family Size (SibSp + Parch):**

- Small families (1-3 total members) had better chances than solo travelers or very large families.

## 7. **Embarked Port:**

- Most passengers boarded from 'S' (Southampton), but survival varied more by class/gender.

## 8. **Other Notes:**

- Data is imbalanced in terms of class and gender, which affects model performance if used for ML.