Detailed Summary of EDA2.ipynb

# 📚 Libraries Imported:

- pandas  
- matplotlib.pyplot  
- seaborn

# 📥 Dataset:

- Loaded titanic\_train.csv into a DataFrame.  
- Basic data exploration (head, tail, info, describe).  
- Checked value counts on the Sex column.

# 🛠 Data Cleaning:

- Identified missing values using isnull().sum().  
- Filled Age missing values with the mean.  
- Dropped the Cabin column (~77% missing values).  
- Dropped missing values from Embarked column.

# 📊 Exploratory Data Analysis (EDA) - Charts and Plots:

## Correlation Heatmap

- Heatmap using seaborn.  
- Survived positively related to Fare, negatively to Pclass.  
- SibSp and Parch moderately correlated.

## Pairplot

- sns.pairplot on Age, Fare, SibSp.  
- Observed relationships among numerical features.

## Scatter Plot

- sns.scatterplot between Age and Fare.  
- Weak or no strong correlation.

## Count Plot (Sex)

- sns.countplot for Sex.  
- More male passengers than females.

## Distribution Plot (Age)

- sns.histplot for Age distribution.  
- Most passengers were young adults (20s-30s).  
- Right-skewed distribution.

## Bar Plot (Survival Rate by Pclass)

- sns.barplot for Survival rate by Pclass.  
- 1st class had the highest survival rates.

## Count Plot (Embarked Port)

- sns.countplot by Embarked.  
- Most passengers boarded from 'S' port.

## FacetGrid/Category Plot (Survived vs. Gender)

- sns.catplot for gender survival.  
- Females had a higher survival rate.

# 📋 Key Observations:

- High survival rate in 1st class, low in 3rd class.  
- Fare positively correlated with survival.  
- Weak correlation between Age and Survival.  
- Female survival rate was significantly higher than male.  
- Most passengers were young adults.