Exploratory Data Analysis (EDA)

1.Overview of the Dataset

Data from Titanic passengers (train.csv) is the source.
891 rows
12 columns (11 after cleaning)
Survived is the target variable (0 = No, 1 = Yes).

2. Data Purification

used the median to fill in the missing age values.
filled in the gaps values that started with the mode.
77% of the data in the Cabin column is missing.
After cleaning, there are no missing values.

3. Analysis of Univariates

Feature Results

AgeThe majority of the passengers were in the 20–40 age range.
The cost Extremely right-skewed; few fares were extremely high, while the majority were low.
The majority of travellers had either one or no siblings or spouses.
The majority of Parch had no parents or kids on board.
The majority of passengers took third class.
Intercourse There are more men than women on board.
Set out The majority of travellers left Southampton (S).
62% did not survive, while 38% did.

4. Bivariate Analysis of Survival Rate by Sex: ➔ The survival rate was significantly higher for females than for males.

Class-by-Class Survival Rate: ➔ Passengers in first class had the highest survival rates.

Survival Rate by Embarked: ➔ Passengers from Cherbourg (C) have a slight advantage.

Survival vs. Age: ➔ Children were more likely to survive.

Survival vs. Fare: ➔ Travellers who paid more for their tickets typically lived longer.

Heatmap of Correlation:

Fare and Survived have a positive correlation.

P class (higher number = lower class) and Survived have a negative correlation.

5. Important Takeaways (Summary):

Gender was important: rescue efforts gave preference to women.

First-class travellers were given preference.

Money was important because a higher fare meant a better chance of surviving.

Younger passengers had better chances because age mattered.

The boarding point had a minor impact.

Conclusion:

Age, gender, class, and ticket price all had a significant impact on the Titanic survival rate. Survival outcomes were significantly influenced by social factors.