**TASK-5 ( EDA )**

**1. Relationships and Trends Identified**

|  |  |
| --- | --- |
| **Relationship** | **Trend / Insight** |
| **Sex vs Survived** | Females had a much higher survival rate than males. Gender played a key role in survival decisions (“women and children first”). |
| **Pclass vs Survived** | Higher class passengers (1st class) survived more often. Lower class passengers (3rd class) had very low survival chances. |
| **Fare vs Survived** | Passengers who paid higher fares were more likely to survive — usually because they were in higher classes with better access to lifeboats. |
| **Age vs Survived** | Younger passengers had slightly higher survival rates; very old passengers had lower survival rates. |
| **Parch/SibSp vs Survived** | Those with small families (1–2 members) had better chances of survival compared to those traveling alone or with large families. |
| **Embarked vs Survived** | Passengers embarked from Cherbourg (C) had a higher survival rate compared to Southampton (S) or Queenstown (Q). |

**2. Observations for Each Visual**

**Pairplot**

* Shows relationships between numeric features like Age, Fare, and Pclass with survival.
* **Observation:** Survived passengers cluster around higher fares and lower Pclass (1st class).

**Heatmap**

* Displays correlation between numerical variables.
* **Observation:**
  + Fare and Pclass are negatively correlated (higher class → higher fare).
  + Survived positively correlates with Fare and negatively with Pclass.
  + Weak correlation of Age with survival — survival depended more on class and fare.

**Histogram (Age & Fare)**

* Shows data distribution.
* **Observation:**
  + Most passengers are between 20–40 years.
  + Fare distribution is right-skewed — few passengers paid very high prices.

**Boxplot (Fare)**

* Detects outliers in ticket prices.
* **Observation:**
  + Several high outliers (very expensive tickets) — likely 1st class passengers.
  + Median fare is quite low, suggesting most passengers traveled economically.

**Scatterplot (Age vs Fare, colored by Survived)**

* Shows whether survival depended on Age or Fare.
* **Observation:**
  + Fare seems more influential than Age.
  + Surviving passengers mostly had higher fares.

**Barplot (Sex vs Survived)**

* Displays average survival rate by gender.
* **Observation:**
  + About **75% of females** survived vs only **~20% of males**.
  + Gender was one of the strongest predictors of survival.

**Barplot (Pclass vs Survived)**

* Displays survival rate across passenger classes.
* **Observation:**
  + **1st Class:** High survival (~65%)
  + **2nd Class:** Moderate survival (~45%)
  + **3rd Class:** Very low survival (~25%)

**Summary of Findings**

1. The Titanic dataset contains **891 passengers** with features like Age, Sex, Fare, and Class.
2. **Missing data:** Over 77 % of data is missing in Cabin
3. **Gender and Class were the strongest predictors** of survival.
4. **Higher fare** and **lower Pclass (1st class)** strongly improved survival odds.
5. **Females and children** had better chances of survival.
6. **Fare and Pclass are strongly related** — indicating ticket price differences by class.
7. **Outliers** in Fare show that only a few passengers paid extremely high prices.