**Task 5: Exploratory Data Analysis (EDA)**  
**Data Analyst Internship**  
**Dataset: Titanic Dataset**  
**Date:**28.4.2025  
**Name:**Harini Shanmugavel

**1. INTRODUCTION**

**Objective:**  
The goal of this project is to perform exploratory data analysis on the Titanic dataset to uncover hidden patterns, insights, and relationships among variables that influence survival.

**2. DATASET DESCRIPTION**

**Dataset Features:**

* PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, Embarked
* Target Variable: **Survived**

**3. EXPLORATORY DATA ANALYSIS (EDA)**

**3.1 Basic Information**

* **Rows**: 891
* **Columns**: 12
* Missing values in: **Age**, **Cabin**, and **Embarked**
* Data types: Mix of **numerical** (Age, Fare) and **categorical** (Sex, Embarked, Pclass)

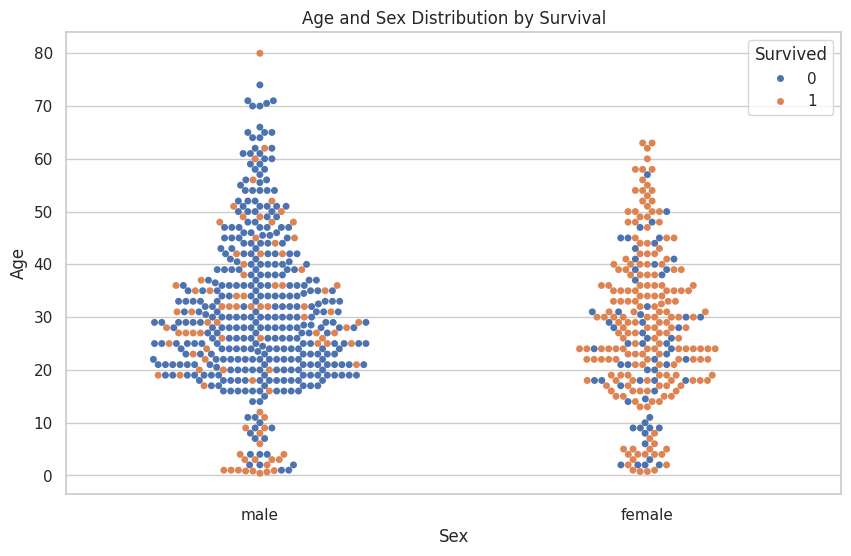
**3.2 Univariate & Bivariate Visualizations**

**a. Violin Plot: Age Distribution by Survival**

****

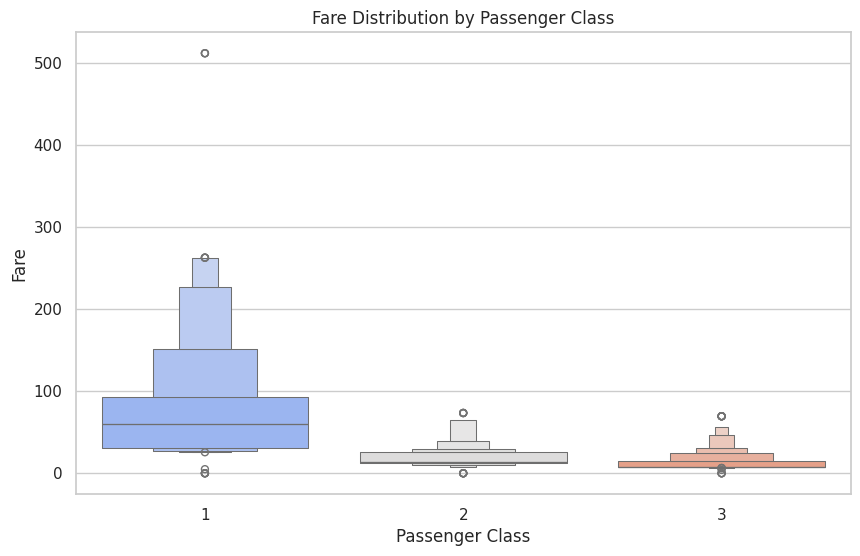
**Observation:**  
Survivors were generally younger. Non-survivors had a wider age range.

**b. Swarm Plot: Sex vs Age by Survival**



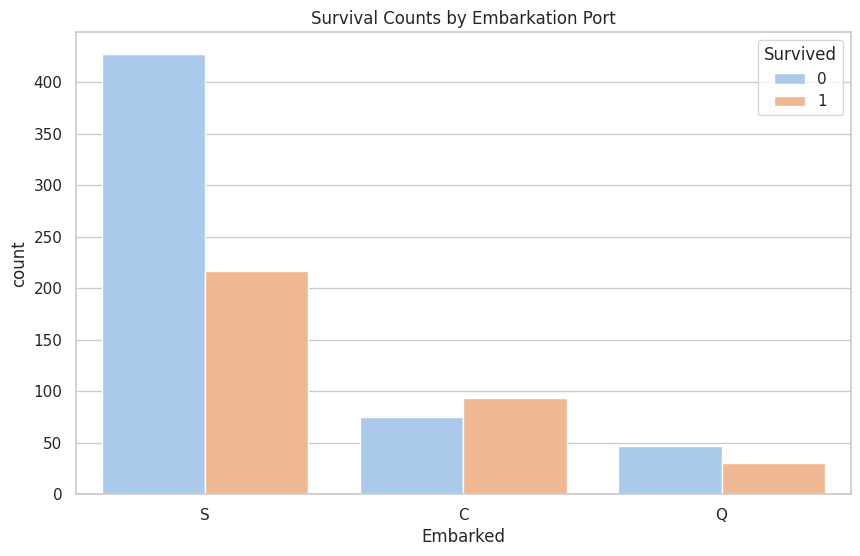
**Observation:**  
Female survivors were mainly younger to middle-aged. Male survivors were fewer and scattered.

**c. Boxen Plot: Fare Distribution by Passenger Class**



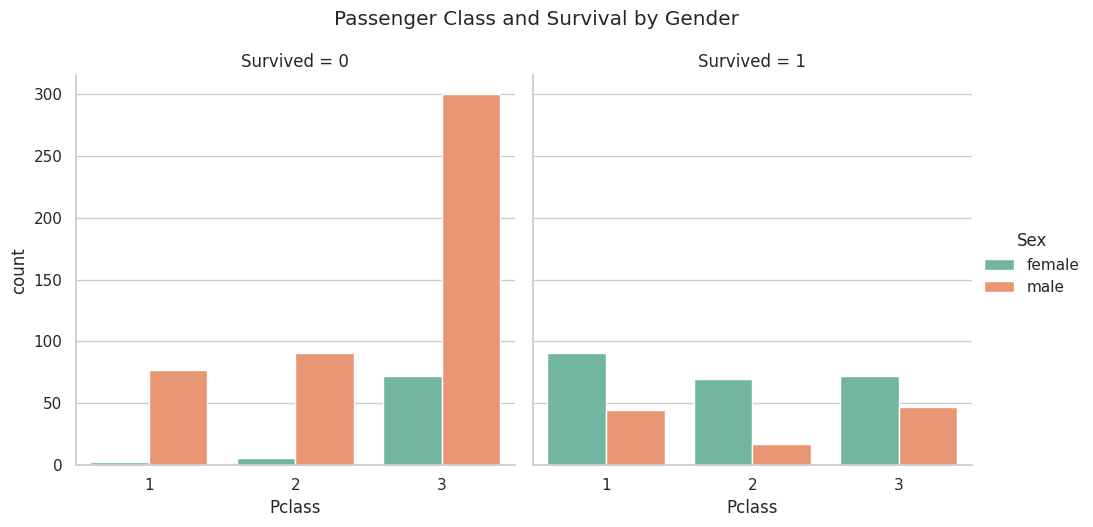
**Observation:**  
Fare clearly increases with class. 1st class passengers paid significantly more.

**d. Countplot with Hue: Embarked vs Survival**



**Observation:**  
Passengers who boarded from Cherbourg (C) had the highest survival rate.

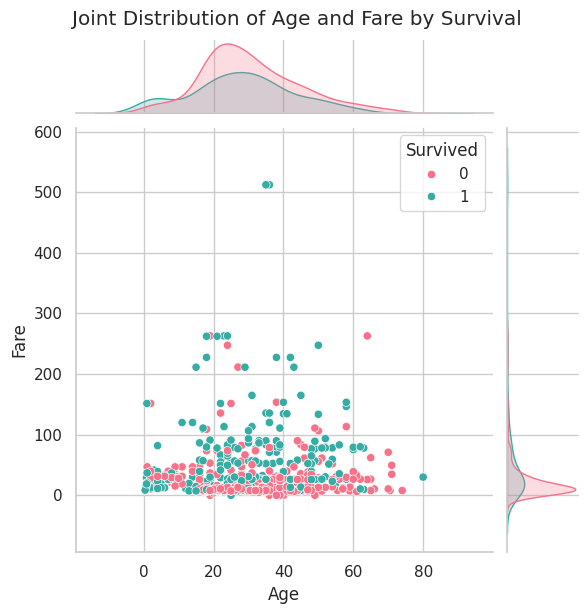
**e. Catplot: Pclass vs Survived by Gender**



**Observation:**  
Females in 1st class had the highest survival. Males in 3rd class had the lowest.

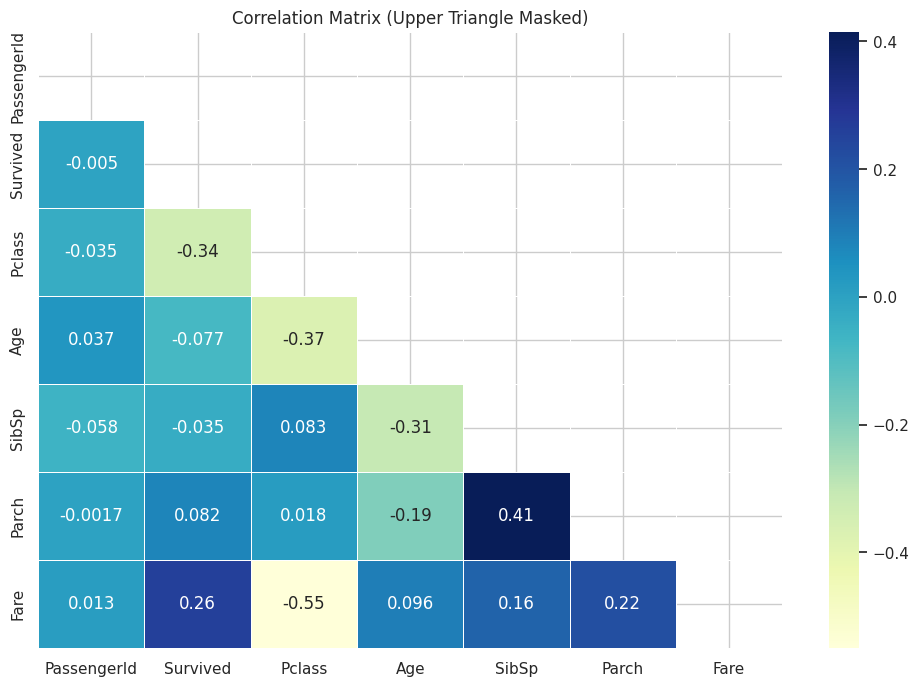
**3.3 Multivariate Analysis**

**f. Jointplot: Age vs Fare by Survival**



**Observation:**  
Survivors cluster at younger ages and higher fares.

**g. Correlation Heatmap with Mask**



**Observation:**

* Survival is positively correlated with Fare
* Survival is negatively correlated with Pclass

**4. FINAL SUMMARY OF FINDINGS**

* Young, 1st class passengers had the highest survival rate
* Females, especially in 1st class, had better odds
* Fare and Pclass are strong indicators of survival
* Port of embarkation also played a role (Cherbourg passengers survived more)