Task 3: Exploratory Data Analysis (EDA)

➢ Objective

Perform Exploratory Data Analysis on a publicly available dataset to:

- Understand the dataset structure
- Identify patterns, trends, and relationships
- Detect missing values and outliers
- Visualize insights clearly
- Summarize findings in a concise report

➢ Dataset Used

Titanic Dataset (from Kaggle)

Why Titanic?

- Clean & well-known dataset
- Contains numerical + categorical features
- Ideal for demonstrating EDA skills clearly

Tools & Libraries

- Python
- Pandas
- NumPy
- Matplotlib
- Seaborn

➢ Step-by-Step EDA Workflow

1.Import Libraries & Load Dataset

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv("titanic.csv")
```

2. Basic Data Understanding

```
df.head()

df.info()

df.describe()
```

3.Missing Value Analysis

df.isnull().sum()

Observations

- Age has missing values

- Cabin has many missing values

- Embarked has few missing values

4.Filling Missing (Null) Values

df['Age'].fillna(df['Age'].mean(),inplace=True)

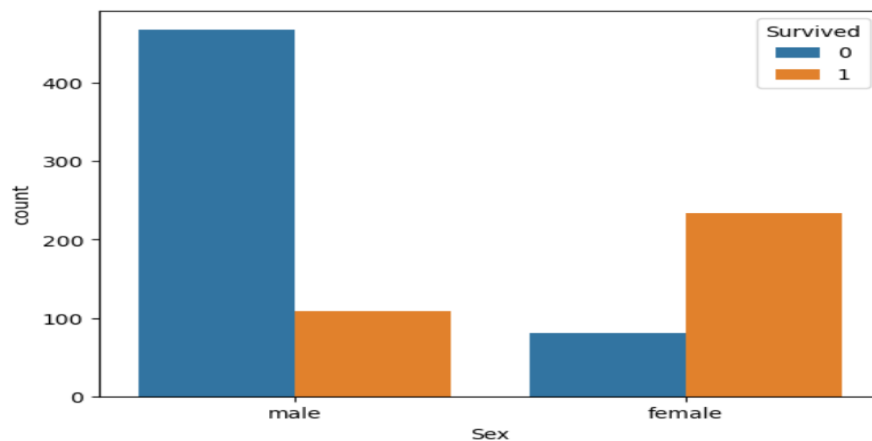df['Embarked'].fillna(df['Embarked'].mode()[0], inplace=True)

df.drop(columns=['Cabin'], inplace=True)

EDA Visualizations

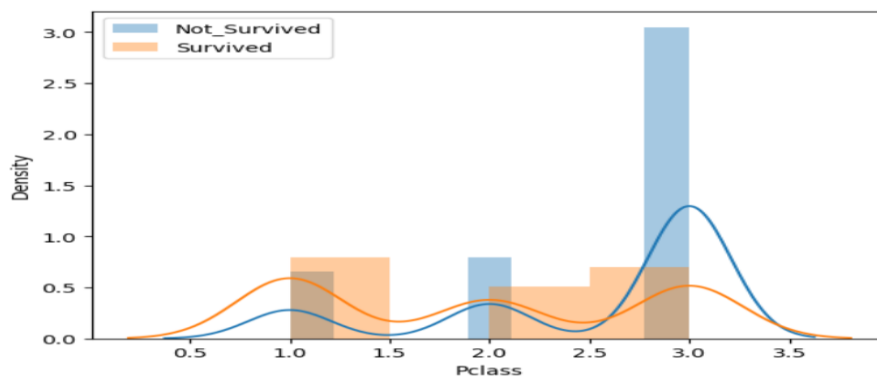- o Count Plot: Gender vs Survival
  sns.countplot(x=df.Sex,hue=df.Survived)

- o Dist Plot:Pclass vs Survival
  sns.distplot(df.Pclass[df.Survived==0],label='Not_Survived')
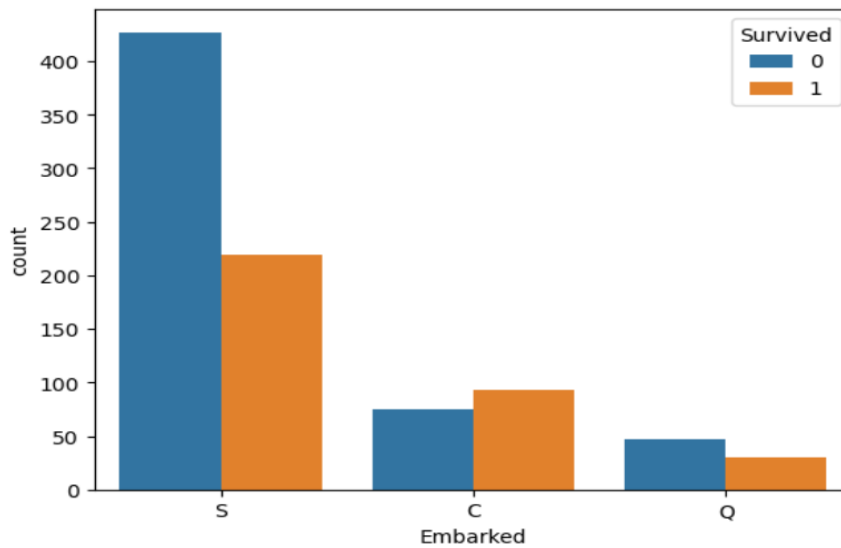 sns.distplot(df.Pclass[df.Survived==1],label='Survived')
  plt.legend()
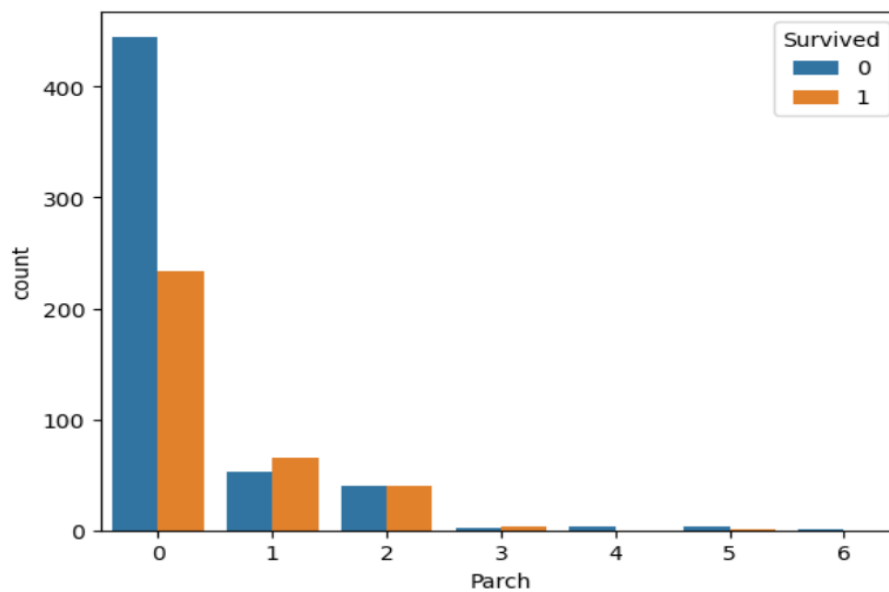 plt.show()

o Dist Plot:Embarked vs Survival
  sns.countplot(x=df.Embarked,hue=df.Survived)

o Dist Plot:Parch vs Survival
  sns.countplot(x=df.Parch,hue=df.Survived)

➢ Observations

- **More males did not survive** compared to females

- **Females had a significantly higher survival count**

- Indicates a **strong relationship between gender and survival**

➢ One-Line Explanation for EDA Report

- The count plot shows that female passengers had a much higher survival rate than male passengers, indicating gender was a significant factor in survival.