Task-5

Exploratory Data Analysis (EDA)

Report

Motapothula Shivaram

shivaram7409@gmail.com

Company:ElevateLabs(Insternship)

# **Introduction:**

This project explores the Titanic dataset to uncover patterns related to passenger survival.  
Using Python libraries like Pandas, Matplotlib, and Seaborn, we clean and analyze the data.  
We investigate how factors like gender, class, age, and fare influenced survival rates.  
Visualizations help identify trends and correlations within the dataset.  
The goal is to gain insights and practice key exploratory data analysis (EDA) techniques.

# Libraries:

* **Python** for scripting
* **Pandas** for data handling
* **Matplotlib** and **Seaborn** for data visualization

# Load the Data:

Reads the Titanic CSV file into a Pandas DataFrame.

df is the main object we'll use for all analysis

data = pd.read\_csv("C:\\Elevate\\Titanic-Dataset.csv")

df = pd.DataFrame(data)

# Initial Exploration:

**.describe():** Shows statistics (mean, median, std, etc.) for numerical features.

.**info():** Displays column names, types, and null counts.

**.head(10):** Previews the first 10 rows.

# Handle Missing Values:

Age filled with median (central value).

Cabin dropped due to too many missing values.

Embarked filled with most frequent value (mode).

# Check for Duplicates:

Ensures there are no duplicate records in the dataset.

# Statistical Summary:

Includes statistics for all data types (numerical + categorical).

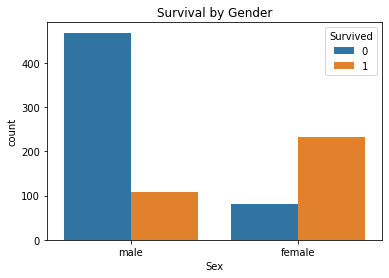
value\_counts() helps see distributions of categorical features.

# Data Visualization:

## Survival by Gender:

Compares male/female survival rates.

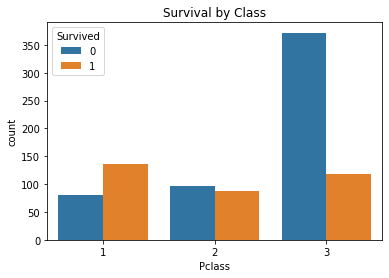
Females had a much higher survival rate



# 2. Survival by Class

sns.countplot(x='Pclass', hue='Survived', data=df)

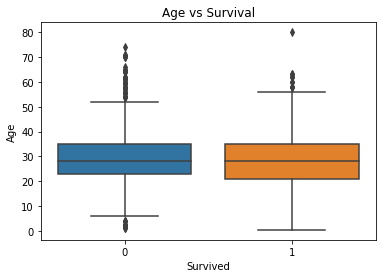
* First-class passengers had higher survival chances.



## 3. Age vs Survival:

sns.boxplot(x='Survived', y='Age', data=df)

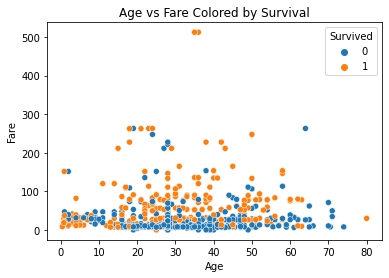
* Younger passengers had slightly higher chances of survival.



## 4. Age vs Fare (Colored by Survival)

sns.scatterplot(x='Age', y='Fare', hue='Survived', data=df)

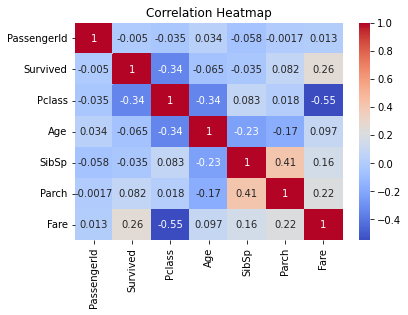
* Shows that passengers who paid more (higher class) were more likely to survive.



## 5. Correlation Heatmap:

sns.heatmap(df.corr(), annot=True, cmap='coolwarm')

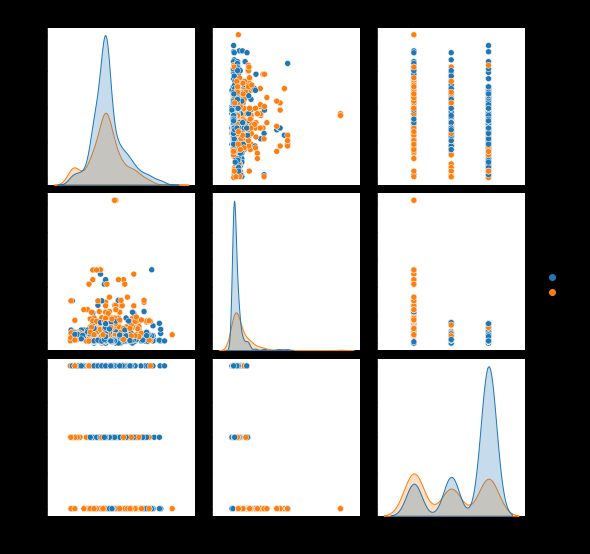
* Displays correlations between features like Fare, Age, Pclass, Survived.
* Fare and Pclass show moderate correlation with Survived.



## 6. Pairplot of Key Features:

sns.pairplot(df[['Survived', 'Age', 'Fare', 'Pclass']].dropna(), hue='Survived')

* Visualizes relationships between features.
* Helps detect patterns among survived vs non-survived passengers.



# Summary of Findings

* **Gender**: Women had much higher survival rates.
* **Class**: First-class passengers were more likely to survive.
* **Age**: Children had better chances, but age had a weak correlation overall.
* **Fare**: Higher fares linked with better survival (likely due to class).
* **Embarked**: Port of embarkation had minor effects.