📑 Titanic Dataset – Exploratory Data Analysis (EDA)

Name: Shaik Vaheed
Tools Used: Python (Pandas, Matplotlib, Seaborn)

🔍 Objective

Perform exploratory data analysis on the Titanic dataset to extract patterns, detect anomalies, and uncover relationships using visual and statistical exploration.

📁 Dataset Overview

- **Source**: Titanic Dataset (tested.csv)
- **Rows**: 418
- **Columns**: 12

📊 Step 1: Data Preview

- Viewed top 5 rows
- Data types, missing values
- Basic statistics

📉 Step 2: Univariate Analysis

🟦 Gender Distribution

```
sns.countplot(x='sex', data=df)
plt.title('Gender Distribution')
plt.show()
```

✅ **Observation**: The dataset contains more male passengers than female passengers.

🟦 Age Distribution

```
df['age'].hist(bins=30)
plt.title('Age Distribution')
plt.xlabel('Age')
plt.show()
```

✅ **Observation**: Most passengers are between 20–40 years. The distribution is right-skewed.

🟦 Age Boxplot

```
sns.boxplot(x=df['age'])
plt.title('Age Boxplot')
plt.show()
```

✅ **Observation**: There are several outliers, including infants and elderly passengers.

📈 **Step 3: Bivariate Analysis**

🟪 **Survival by Gender**

```
sns.countplot(x='survived', hue='sex', data=df)
plt.title('Survival Count by Gender')
plt.show()
```

✅ **Observation**: Females had a significantly higher survival rate than males.

🟪 **Survival by Passenger Class**

```
sns.countplot(x='survived', hue='pclass', data=df)
plt.title('Survival by Class')
plt.show()
```

✅ **Observation**: 1st class passengers had a much better chance of survival than 3rd class passengers.

🔗 **Step 4: Correlation Analysis**

🔶 **Correlation Heatmap**

```
plt.figure(figsize=(10,6))
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
plt.show()
```

✅ **Observation**: Fare and Pclass show moderate correlation. Survival is positively correlated with Fare and negatively with Pclass.

🔗 **Step 5: Pairplot**

```
sns.pairplot(df[['age', 'fare', 'pclass', 'survived']], hue='survived')
plt.show()
```

✅ **Observation**: Higher fare and lower Pclass (1st class) passengers show better survival clustering.

📝 Summary of Insights

- 🚹 **Gender**: Women had much higher survival rates.
- 💼 **Class**: 1st class passengers were more likely to survive.
- 💰 **Fare**: Higher-paying passengers had better survival chances.
- 👶 **Age**: Younger passengers (especially children) were slightly more likely to survive.
- 🔍 **Outliers**: Found in Fare and Age.
- 🔄 **Missing Values**: Some nulls in Age, Embarked, etc.
- 📈 **Correlations**: Survival weakly correlated with fare and class.

✅ Deliverables

File Name	Description
Task_5_EDA.ipynb	Jupyter Notebook with code
titanic_eda_vaheed.pdf	Final EDA Report in PDF format