Titanic EDA - Internship Task 5

This project is part of a data analyst internship task focused on **Exploratory Data Analysis (EDA)** using the Titanic dataset.

📌 Objective

Perform EDA to uncover insights, patterns, and trends within the Titanic dataset using Python libraries such as Pandas, Seaborn, and Matplotlib.

📁 Files Included

- `Titanic_EDA.ipynb` — Jupyter Notebook with complete code, visualizations, and insights
- `Titanic_EDA.pdf` — Exported PDF report from the notebook
- `train.csv` — Titanic dataset used

🛠️ Tools Used

- Python 3
- Pandas
- Matplotlib
- Seaborn
- Jupyter Notebook / Google Colab

🔍 Key Steps Performed

1. **Data Loading & Overview**
   - Loaded Titanic dataset
   - Used `.info()`, `.describe()`, `.isnull().sum()` for quick insights
2. **Univariate Analysis**
   - Visualized single-variable distributions (e.g., Age, Sex, Survived)
3. **Bivariate Analysis**
   - Compared survival with class, gender, and age
   - Used barplots and boxplots
4. **Multivariate Analysis**
   - Heatmap of numeric correlations
   - Pairplots for selected features
5. **Insights & Summary**
   - Females and first-class passengers had higher survival rates
   - Age and Fare had influence on survival probability
   - Cabin column had excessive missing data

By- Anandita Choudhary

📈 Sample Plots

- Count plots for gender and survival
- Histogram for age distribution
- Heatmap for feature correlation
- Pairplot showing variable relationships

📊 Summary of EDA Findings

- Female passengers had a much higher survival rate than males.
- Passengers in 1st class were more likely to survive than those in 2nd or 3rd class.
- Younger passengers had slightly better survival odds.
- Fare has a weak positive correlation with survival; higher fares often relate to 1st class.
- Many missing values were observed in the 'Cabin' column, which may affect model accuracy if used.
- No strong multicollinearity observed among major features.

✅ Submission

This task was completed and submitted as part of the internship requirements via GitHub and form submission.

By- Anandita Choudhary