🚢 Titanic Dataset – Exploratory Data Analysis Report

Date: August 2025
Tools Used: Python, Pandas, Matplotlib, Seaborn, Jupyter Notebook
 Objective: To uncover survival patterns and trends using statistical and visual exploration of the Titanic dataset.

📁 *Dataset Overview*

- **Source**: [Kaggle Titanic Dataset](#)
- **Total Records**: 891 passengers
- **Key Features**:
- `Survived`: Binary indicator of survival (0 = No, 1 = Yes)
- `Pclass`: Passenger class (1st, 2nd, 3rd)
- `Sex`, `Age`, `Fare`, `Embarked`, `Cabin`, `SibSp`, `Parch`

🔍 *Methodology*

1. Data Inspection

- Used `.info()`, `.describe()`, and `.isnull().sum()` to assess structure and missing values.
- Found missing data in `Age`, `Cabin`, and `Embarked`.

2. Univariate Analysis

- **Age**: Most passengers were between 20–40 years old.
- **Fare**: Skewed distribution; majority paid lower fares.
- **Pclass**: Most passengers were in 3rd class.
- **Sex**: More males than females.

3. Bivariate Analysis

- **Survival by Sex**: Females had significantly higher survival rates.
- **Survival by Class**: 1st class passengers had better survival odds.
- **Age vs Survival**: Survivors tended to be younger.
- **Fare vs Survival**: Survivors paid higher fares on average.

4. Multivariate Analysis

- **Correlation Heatmap**:
  - Moderate negative correlation between `Pclass` and `Fare`.
  - Weak correlation between `Survived` and numeric features.
- **Pairplot**:
- Clear separation in survival based on fare and age clusters.

📈 *Summary of Findings*

- **Gender Bias**: Females had a survival rate of ~74%, while males had ~19%.
- **Class Divide**: 1st class passengers had a survival rate of ~63%, compared to ~25% in 3rd class.
- **Fare Impact**: Higher fares correlated with better survival odds.
- **Age Factor**: Younger passengers had slightly better survival chances.
- **Missing Data**: `Cabin` had extensive missing values; `Age` and `Embarked` had moderate gaps.
- **Feature Relationships**: `Pclass` and `Fare` showed moderate correlation; survival patterns were more categorical than numeric.

📊 *Business Implications*

- **Evacuation Prioritization**: Gender and class influenced survival, suggesting systemic biases or protocol-driven prioritization.
- **Socio-Economic Influence**: Fare and class are proxies for wealth, which impacted survival odds.
- **Data Quality**: Missing values in key features like `Cabin` and `Age` must be addressed before predictive modeling.

📄 *Deliverables*

- ✅ `Titanic_EDA.ipynb`: Full notebook with code, visuals, and inline observations
- ✅ `Titanic_EDA_Report.pdf`: This report, formatted for stakeholder review

.