README - Task 2: Titanic Dataset EDA

Exploratory Data Analysis (EDA) on Titanic Dataset

Objective

Perform Exploratory Data Analysis (EDA) to understand the Titanic dataset using statistics and visualizations.

Dataset

- Source: Titanic Dataset (Kaggle)
- Rows: 891
- Columns: 12
- Key columns: Survived, Pclass, Sex, Age, Fare, SibSp, Parch, Embarked

Steps Performed

1. Data Overview
   - Checked shape (891 rows × 12 columns).
   - Identified missing values: Age (177 missing), Cabin (687 missing), Embarked (2 missing).

2. Summary Statistics
   - Average Age ~ 30 years
   - Average Fare ~ 32.2

3. Histograms
   - Age distribution skewed towards younger passengers.

- Fare distribution is right-skewed.

4. Boxplots

- Age has outliers.
- Fare has significant outliers.

5. Correlation Heatmap

- Pclass vs Fare (-0.55 correlation)
- SibSp and Parch are correlated (family size)
- Survived correlated slightly with Fare and negatively with Pclass.

Insights

- Higher-class passengers had higher survival chances.
- Females survived at a much higher rate than males.
- Higher fares linked to higher survival.
- Younger children had better survival chances.

Tools Used

- Pandas for data handling
- Matplotlib & Seaborn for visualization

How to Run

```bash
pip install pandas matplotlib seaborn
```

Run the notebook/script and explore visualizations.