Titanic Dataset - Exploratory Data Analysis (EDA)

Objective

Perform Exploratory Data Analysis (EDA) on the Titanic dataset to:

- Understand data distribution
- Identify patterns, trends, and anomalies
- Visualize relationships between features

Tools Used

- Python
- Pandas
- Matplotlib
- Seaborn
- Plotly (for interactive plots)
- VS Code

Files in This Repository

- titanic_data.py : Python script containing the EDA code.
- README.md : This file. Summary of what the project does.
- Titanic-Dataset.csv : Dataset file (if uploaded, or specify source).

Key Steps Performed

- Loaded Titanic dataset using Pandas
- Generated descriptive statistics (mean, median, std, etc.)
- Created histograms and boxplots to visualize numeric features

- Built correlation matrix and explored feature relationships
- Used Plotly to create interactive charts
- Identified:
  - Higher survival rate for women and 1st class passengers
  - Fare highly skewed (a few paid very high fares)
  - Outliers in age and fare
  - Negative correlation between Pclass and Fare

How to Run

1. Clone this repository
2. Make sure you have required packages installed:

pip install pandas matplotlib seaborn plotly

3. Run the Python script:

python titanic_data.py

Interactive charts will open in your browser.

Dataset Source

- Dataset: Titanic-Dataset.csv
- You can also load using: sns.load_dataset('titanic')

Author

Pranitha Bokketi