**Titanic Dataset EDA**

**📘 Project Overview**

This project performs **Exploratory Data Analysis (EDA)** on the famous Titanic dataset using Python. The goal is to extract insights from the data, visualize trends, and understand factors affecting passenger survival.

**🎯 Objectives**

* Analyze the dataset to identify trends and patterns.
* Handle missing values and perform data cleaning.
* Create new features to gain deeper insights (e.g., FamilySize, IsAlone).
* Visualize relationships between features and survival.
* Summarize findings for business or research insights.

**🛠 Tools & Libraries**

* **Python 3.x**
* **Pandas** – Data manipulation
* **NumPy** – Numerical operations
* **Matplotlib** – Data visualization
* **Seaborn** – Statistical plotting
* **Jupyter Notebook** – Interactive analysis

**📊 Steps Performed**

1. **Data Loading & Inspection**
   * Loaded train.csv
   * Checked data types, missing values, and summary statistics.
2. **Data Cleaning**
   * Filled missing Age values with median
   * Filled missing Embarked values with mode
   * Dropped Cabin column due to excessive missing data
3. **Feature Engineering**
   * FamilySize = SibSp + Parch + 1
   * IsAlone = 1 if FamilySize = 1, else 0
4. **Univariate Analysis**
   * Countplots for categorical features (Survived, Pclass, Sex, Embarked)
   * Histograms and boxplots for numerical features (Age, Fare)
5. **Bivariate Analysis**
   * Survival distribution by Pclass, Sex, Embarked
   * Boxplots of Age and Fare vs Survived
6. **Correlation Analysis**
   * Correlation heatmap of numeric features
   * Identified relationships like higher fare correlating with higher survival
7. **Pairplot Analysis**
   * Visualized multi-variable relationships with Age, Fare, SibSp, Parch, FamilySize, IsAlone, and Survived

**📈 Observations & Insights**

* Overall survival rate is ~38%.
* **Class Matters:** 1st & 2nd class passengers had higher survival rates than 3rd class.
* **Gender Impact:** Females had a higher survival rate than males.
* **Age Effect:** Younger passengers survived more often than older passengers.
* **Fare Influence:** Passengers who paid higher fares were more likely to survive.
* **Family Factor:** Passengers traveling alone (IsAlone) had lower survival compared to families.

**📂 Deliverables**

* Titanic\_EDA.ipynb – Jupyter Notebook with all analysis, plots, and observations
* EDA\_Report.pdf – PDF summary of findings and visualizations

**🔗 Dataset**

* Source: [Kaggle Titanic Competition](https://www.kaggle.com/c/titanic/data)
* File used: train.csv