**Titanic Survival Prediction – Data Analysis Project**

This project explores the Titanic dataset to understand patterns in passenger survival. It involves data cleaning, exploration, visualization, and optional prediction modeling.

**📁 Files Included**

| **File Name** | **Description** |
| --- | --- |
| train.csv | Training dataset with labeled outcomes |
| test.csv | Test dataset without survival labels |
| gender\_submission.csv | Sample submission file for predictions |
| titanic\_analysis.ipynb | Main Jupyter notebook with analysis code |
| combined\_titanic.csv | (Optional) Merged file from all CSVs |
| titanic\_readme.md | This documentation file |

**🔧 Technologies Used**

* Python 3.x
* Jupyter Notebook
* Pandas
* NumPy
* Seaborn & Matplotlib (for visualization)
* Scikit-learn (for optional ML modeling)

**📊 Key Steps Performed**

1. **Data Import**  
   Merged train.csv, test.csv, and gender\_submission.csv from a ZIP file using Python.
2. **Data Cleaning**
   * Checked for missing values
   * Handled outliers using IQR
   * Created new columns like age/fare bins
3. **Exploratory Data Analysis (EDA)**
   * Visualized survival rates by gender, class, fare, etc.
   * Used sns.pairplot() and sns.heatmap() for insights
   * Correlation analysis of numerical features
4. **Feature Engineering** *(optional)*
   * Binned columns (e.g., Fare, Age)
   * Extracted titles from names
5. **(Optional)** Predictive Modeling
   * Logistic Regression, Random Forest, etc.
   * Trained on train.csv, tested on test.csv

**📈 Example Visualizations**

* Survival by gender (sns.barplot)
* Correlation matrix (sns.heatmap)
* Distribution of Age and Fare (sns.histplot, sns.boxplot)
* Pairwise relationships (sns.pairplot)