Titanic Survival Analysis

This repository contains an analytical study of the survival factors aboard the RMS Titanic, leveraging Python and the Kaggle Titanic dataset. The analysis explores the influence of demographics, socioeconomic status, and family dynamics on survival outcomes during one of history's most significant maritime disasters.

Project Overview

The sinking of the Titanic in 1912 provides a unique opportunity to examine human behavior under crisis. This project analyzes key factors that influenced survival, including:

- Gender and ticket class.
- Age and family relationships.
- Economic factors, such as fare.
- Port of embarkation.

The project uses data visualization and statistical exploration to uncover patterns and derive meaningful insights from historical data.

Key Findings

1. **Gender**: Females had significantly higher survival rates (~74%) compared to males (~19%), reflecting prioritization during evacuation.

2. **Class**: First-class passengers had the highest survival rates (~63%), while third-class passengers experienced the lowest (~24%).
3. **Age**: Children (0-12) were prioritized, with survival rates of ~58%, whereas seniors and young adults faced the greatest challenges (~35%).
4. **Family Size**: Small family groups (2-4 members) had better survival outcomes (~70%), while solo travelers and large families struggled (~30%).
5. **Economic Factors**: Higher fares correlated with higher survival rates, emphasizing the role of socioeconomic privilege.
6. **Embarkation**: Passengers boarding at Cherbourg (C) fared better, likely due to a higher concentration of first-class passengers.

Technologies Used

- **Python**: Programming language for data analysis and visualization.
- **Pandas**: Data manipulation and preprocessing.
- **Matplotlib & Seaborn**: Data visualization libraries.
- **Google Colab**: Development platform for analysis and collaboration.

Project Files

- **notebooks/**: Contains the Jupyter Notebook with the full analysis and code.
- **data/**: Includes the Titanic dataset used for analysis.
- **visualizations/**: Contains visualizations generated during the analysis.
- **reports/**: The final written report summarizing insights and findings.

Acknowledgments

- Dataset sourced from Kaggle: Titanic Dataset.
- Tools used include Python, Pandas, Matplotlib, Seaborn, and Google Colab.