

## Titanic Survival Analysis

This project analyzes the Titanic dataset to predict passenger survival based on various factors such as socio-economic status, age, gender, and family connections. The goal is to explore the relationships between different attributes and their impact on survival outcomes.

### Project Overview

This analysis uses the Titanic dataset, which contains information about passengers on the ill-fated RMS Titanic. The dataset includes both numerical and categorical features, such as age, class, fare, and family connections, as well as whether the passenger survived or not.

The project aims to:

- Perform exploratory data analysis (EDA) to understand the distribution of key variables.
- Visualize the relationships between different features.
- Build predictive models to predict passenger survival based on these attributes.

### Dataset Description

The dataset contains the following attributes:

Attribute	Type	Description	Possible Values
PassengerId	int64	Unique identifier for each passenger.	Sequential integer values.
Survived	float64	Indicates whether the passenger survived or not.	1 = Survived, 0 = Did not survive
Pclass	int64	Class of the ticket the passenger purchased.	1 = First class, 2 = Second class, 3 = Third class
Name	object	Full name of the passenger.	String (e.g., "John Doe"). Can contain titles like Mr., Mrs., Miss.
Sex	object	Gender of the passenger.	male, female
Age	float64	Age of the passenger in years.	Numeric (e.g., 22.0, 35.0). Missing values might occur.

Attribute	Type	Description	Possible Values
<b>SibSp</b>	int64	Number of siblings or spouses the passenger was traveling with.	Non-negative integers (e.g., 0, 1, 2).
<b>Parch</b>	int64	Number of parents or children the passenger was traveling with.	Non-negative integers (e.g., 0, 1, 2).
<b>Ticket</b>	object	The ticket number.	String (e.g., "A/5 21171"). Can be used for grouping or exploration of ticket pricing.
<b>Fare</b>	float64	The amount of money paid for the ticket.	Numeric (e.g., 7.25, 71.2833).
<b>Cabin</b>	object	The cabin the passenger stayed in.	String (e.g., "C85"). Many missing values. Could be used to extract deck information or dropped.
<b>Embarked</b>	object	The port where the passenger boarded the Titanic.	C = Cherbourg, Q = Queenstown, S = Southampton
<b>dataset</b>	object	Categorization of the dataset (e.g., training or testing data).	String (e.g., "train", "test").

## Key Insights

- **Survival and Socio-Economic Status:** Higher class passengers (lower Pclass number) had a higher likelihood of survival.
- **Family Connections:** The number of family members (SibSp and Parch) had a weak correlation with survival chances.
- **Fare and Survival:** Passengers who paid higher fares had a slightly higher chance of survival.

## Installation

1. Clone the repository:

```
git clone https://github.com/your-username/titanic-survival-analysis.git
```

## Contributing

Contributions are welcome! Please fork the repository and submit a pull request with your changes.

## License

This project is licensed under the MIT License - see the LICENSE file for details.

### Notes:

- Replace `your-username` with your actual GitHub username in the clone URL.
- You can expand the usage section to include specific analysis steps or more detailed model training examples as you develop the project.
- The `requirements.txt` should include the libraries required to run your project (e.g., `pandas`, `seaborn`, `scikit-learn`, etc.).

This template provides a well-structured overview of the Titanic survival analysis project, making it easy for others to understand and contribute.