Titanic Dataset Visualization Using Shiny App

Created By: Dellah Salem

Table of Contents

---

1. Introduction

The goal of this project was to build a dynamic and interactive data visualization tool using the Titanic dataset from Kaggle. The Titanic dataset contains information about passengers aboard the RMS Titanic, including their survival status, demographic details, and class. The aim was to provide users with a tool to visualize various aspects of the Titanic dataset through a Shiny app, which allows for interactive exploration of the data. In this report, I will explain the dataset used, the features of the Shiny app, and provide a user manual for using the app.

2. Dataset Information

The dataset used for this project is the Titanic dataset from Kaggle. Specifically, the train.csv file contains the following key variables:

- **PassengerId**: A unique identifier for each passenger.
- **Pclass**: The passenger's class (1, 2, or 3).
- **Name**: The name of the passenger.
- **Sex**: The gender of the passenger.
- **Age**: The age of the passenger.
- **SibSp**: The number of siblings/spouses aboard the Titanic.
- **Parch**: The number of parents/children aboard the Titanic.
- **Fare**: The fare paid by the passenger.
- **Embarked**: The port of embarkation (C = Cherbourg, Q = Queenstown, S = Southampton).
- **Survived**: Whether the passenger survived (1 = Yes, 0 = No).

  For the Shiny app, I focused on visualizing the Pclass, Age, Sex, and Survived variables to understand survival trends by gender, age distribution, and passenger class.

3. Shiny App Features

The Shiny app is designed to provide an interactive way to explore the Titanic dataset. The main features of the app are:

1. **CSV File Upload**: Users can upload their own Titanic dataset in CSV format.
2. **Passenger Class Selection**: A dropdown menu allows users to filter the data by passenger class (1, 2, or 3).
3. **Age Range Slider**: A slider enables users to filter the data based on a selected age range (0-80 years).
4. **Survival Rate Plot**: A bar chart that shows the survival rate based on gender, with separate bars for survivors and non-survivors.
5. **Age Distribution Plot**: A histogram that displays the distribution of ages among passengers, filtered by the selected passenger class and age range.

These features give users the ability to interact with the dataset and explore trends related to survival, age, and class.

4. User Manual

To use the Titanic Dataset Visualization Shiny app:

1. **Upload a CSV file**:
    - Click on the “Choose CSV File” button to upload a Titanic dataset in CSV format (e.g., train.csv).
2. **Select Passenger Class**:
    - Use the dropdown menu labeled "Select Passenger Class" to filter the data by the passenger class (1, 2, or 3).
3. **Adjust Age Range**:
    - Use the Age Range slider to select an age range (e.g., 0 to 40 years). The app will automatically update the plots based on this selection.
4. **View the Plots**:
    - The app will display two plots:
        - **Survival Rate by Gender**: This bar chart shows the number of survivors and non-survivors, grouped by gender.
        - **Age Distribution of Passengers**: This histogram shows the age distribution of passengers within the selected age range and class.

The plots will update dynamically as you adjust the inputs.

5. References

Titanic Dataset: https://www.kaggle.com/competitions/titanic/data