# RedBus Data Scraper and Streamlit Application

## Project Overview

This project involves scraping data from the redbus website using Selenium, storing the data in a MySQL database, and building a Streamlit application to display and filter the bus route data. The data includes details such as route names, bus names, bus types, departing and reaching times, star ratings, prices, and seat availability.

## Requirements

- Python 3.x
- Selenium
- MySQL Connector
- Streamlit
- Pandas
- WebDriver (e.g., ChromeDriver for Chrome)

## Installation

1. **Clone the repository:**

   ```
   git clone https://github.com/Shabanabacker/redbus-data-scraper.git
   cd redbus-data-scraper
   ```

2. **Install the required Python packages:**

   ```
   pip install <python libraries>
   ```

3. **Download and set up WebDriver:**
   - Download the appropriate WebDriver for your browser from ChromeDriver or the corresponding site for other browsers.
   - Place the WebDriver executable in a directory that's in your system's PATH.

# Project Structure

```
redbus-data-scraper/
│
├── redbus_scraping.ipynb         # Script to scrape data from RedBus and
insert it into the MySQL database
├── redbusdetails.py     # Streamlit application to display and filter the
bus route data
├── README.md
└── Redbus_Data_Scraper_and_Streamlit_Application.pdf    # Documentation
```

# Scraping Script (`redbus_scraping.ipynb`)

This script uses Selenium to navigate through various state transport pages on the redbus website, scrape bus route data, and store it in a MySQL database.

## Configuration

- **MySQL Database:**
  - Host: `localhost`
  - User: `root`
  - Password: `<your password>`
  - Database: `REDBUS`

## Usage

1. **Ensure your MySQL database and table are set up:**

   ```
   CREATE TABLE bus_route (
       id INT AUTO_INCREMENT PRIMARY KEY,
       routename VARCHAR(255),
       routelink VARCHAR(255),
       busname VARCHAR(255),
       bustype VARCHAR(255),
       departingtime TIME,
       duration VARCHAR(255),
       reachingtime TIME,
       starrating FLOAT,
       price DECIMAL(10,2),
       seatavailable INT(200)
   );
   ```

2. **Run the script:**

   ```
   python redbus_scraping.ipynb
   ```

# Streamlit Application (`redbusdetails.py`)

This script creates a web application using Streamlit to display and filter bus route data.

### Usage

1. **Run the Streamlit application:**

   ```
   streamlit run redbusdetails.py
   ```

2. **Access the application in your web browser at:**

   ```
   http://localhost:8501
   ```

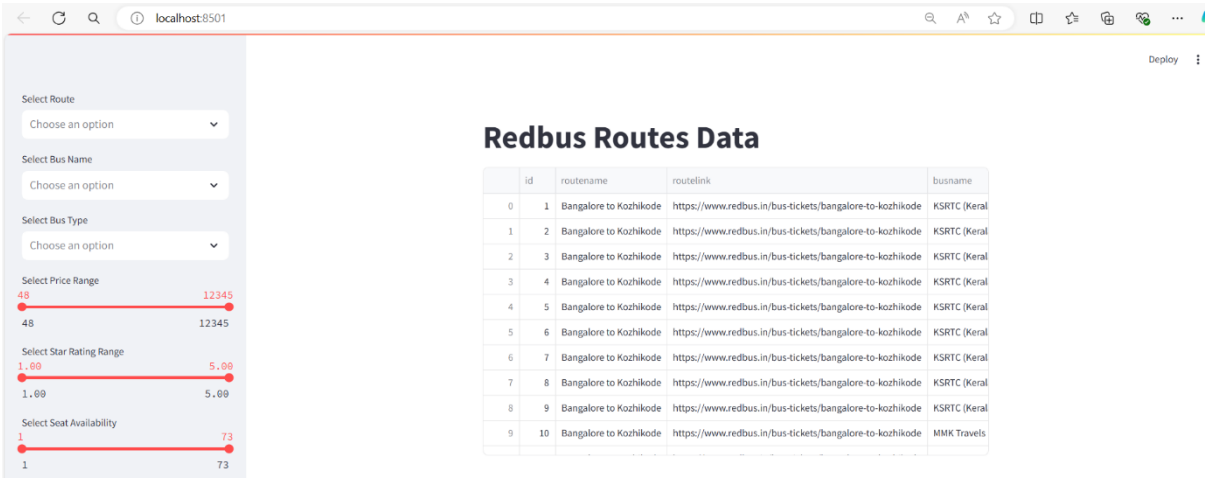
### Features

- **Filters:**
  - Select Route
  - Select Bus Name
  - Select Bus Type
  - Select Price Range
  - Select Star Rating Range
  - Select Seat Availability

## Configuration

- Modify the database connection settings in `redbusdetails.py` if needed.

## Application Screenshot




## Acknowledgements

- Selenium for web scraping
- Streamlit for creating the web application
- MySQL for data storage

## Contact

For any questions or feedback, please reach out to shabana.backer@gmail.com.