**Redbus Data Scraping with Selenium & Dynamic Filtering using Streamlit**

**1. Introduction**

The purpose of this project is to develop a web scraping solution to gather real-time bus route information from the Redbus platform. This project aims to analyze bus routes across multiple states in India, store the data in a MySQL database, and present it in an interactive Streamlit app.

**2. Objectives**

* **Data Collection**: Scrape bus route information from the Redbus website for 10 different states.
* **Data Storage**: Store the scraped data in a structured format in a MySQL database.
* **Interactive Application**: Build a Streamlit app to display and analyze the bus route data.
* **Data Analysis**: Analyze trends in bus routes, availability, and other relevant parameters.

**3. Project Scope**

* **Bus Routes**: The project covers 10 states, namely:
  + Kerala
  + Andhra Pradesh
  + Assam
  + Himachal Pradesh
  + Goa(Kadamba)
  + Rajasthan
  + South Bengal
  + Telangana
  + Uttar Pradesh
  + West Bengal
* **Web Scraping**: Selenium will be used to scrape the bus route information for these states.
* **Data Processing**: The data collected will be merged into a CSV file and then inserted into a MySQL database for easier access and manipulation.

**4. Tools and Technologies**

* **Web Scraping**: Selenium is used to automate the browser for collecting data from the Redbus website.
* **Database**: MySQL is used to store and manage the bus route information.
* **Frontend**: A Streamlit app will be developed to visualize the bus routes and provide an interactive user interface for querying the data.
* **Programming Language**: Python is used for scraping, database interaction, and developing the Streamlit app.

**5. Methodology**

1. **Data Scraping**:
   * Use Selenium to navigate the Redbus website and scrape bus route details.
   * The details include bus name, departure time, arrival time, price, route, and availability.
   * Save the data for 10 states into a CSV file ('merged\_buses.csv').
2. **Data Processing**:
   * Clean and format the data using Pandas.
   * Merge the bus route data from different states into a single file for further analysis.
3. **Data Storage**:
   * Create a MySQL database (REDBUS\_DETAILS) and a table (BUS\_ROUTES) to store the bus route information.
   * Insert the merged CSV data into the MySQL table for persistent storage.
4. **Streamlit App**:
   * Build a user-friendly interface that allows users to filter bus routes based on their preferences (e.g., state, price, availability).
   * Display analysis, such as busiest routes, common departure times, etc.

**6. Database Design**

* **Table Name**: BUS\_ROUTES
* **Fields**:
  + Bus\_Name: The name of the bus operator.
  + Start\_Time: The time the bus departs.
  + End\_Time: The time the bus arrives at its destination.
  + Total Duration : The total duration of the journey is
  + Route: The route the bus follows.
  + Price: The cost of the bus ticket.
  + Availability: The number of available seats.
  + Rating : The customer rating of the bus service

**7. Challenges**

* **Data Quality**: Ensuring the accuracy and completeness of the scraped data from the Redbus platform.
* **Rate Limiting**: Managing the web scraping operations within Redbus’s limitations to avoid bans or throttling.
* **Data Storage**: Efficiently storing and retrieving large amounts of bus route data from multiple states.
* **User Interaction**: Building an intuitive interface in Streamlit for easy data visualization and querying.

**8. Conclusion**

This project provides a comprehensive system to gather, store, and analyze bus route data from Redbus, helping users and stakeholder’s access and analyze transportation information. By integrating data scraping, storage, and visualization in a unified application, it simplifies the process of tracking bus routes across different regions in India.