Red bus data scraping with selenium & Dynamic Filtering using streamlit

PROJECT DESCRIPTION

This project involves creating a web scraping tool using Selenium to extract bus travel data from the Red Bus website (or a similar bus booking website). The scraped data includes details such as bus names, travel times, ticket prices, available seats, and departure/arrival locations. This data is then made available through a user-friendly interface built with Streamlit, where users can dynamically filter and search for buses based on various criteria such as price range, travel duration, seat availability, etc.

Tech Stack:

1. **Selenium**: For automated web scraping of Red Bus data.
2. **Streamlit**: For building the dynamic filtering interface.
3. **Python**: Backend for handling scraping, data processing, and filtering.
4. **Pandas**: For data manipulation and filtering.
5. **Mysql**: For relational database management system.

Key Features:

1. **Web Scraping with Selenium**:
    - Use Selenium to navigate and scrape data from the Red Bus website dynamically.

- Extract information such as:
  - Bus Name
  - Bus Route
  - Departure Time
  - Arrival Time
  - Ticket Prices
  - Available Seats
  - Ratings
  - Duration of the journey
- Handle dynamic loading content (e.g., loading new results as the user scrolls).

2. **Data Storage**:
   - Store the scraped data in a structured format (Pandas Data Frame or CSV).
   - Ensure real-time scraping when required or use cached data to improve performance.
3. **Streamlit Interface**:
   - Build an interactive front-end using Streamlit.
   - Display the bus data in a clean and user-friendly format (like tables, cards, etc.).
4. **Dynamic Filtering**:
   - Enable users to filter the buses dynamically based on criteria such as:
     - **Price Range**: Filter buses based on the minimum and maximum price.

        - **Bus Operator**: Choose specific operators.
        - **Departure/Arrival Time**: Time-based filtering to select buses within a specific time range.
        - **Duration of Journey**: Short or long journeys.
        - **Seat Availability**: Only show buses with available seats.
    - Provide instant feedback on filters, dynamically updating the results.
5. **Sorting**:
    - Add sorting functionality to arrange buses by:
        - Price (ascending/descending)
        - Travel Time (shortest/longest)
        - Ratings (highest/lowest)
6. **Search**:
    - Add a search bar to allow users to search for specific routes or operators.

Detailed Steps:

<u>1. Web Scraping with Selenium:</u>

- **Setup**: Set up Selenium with appropriate drivers (Chrome Driver/Gecko Driver).
- **Navigation**: Use Selenium to navigate to Red Bus's search page and input travel details (e.g., departure city, destination, date).
- **Data Extraction**:
    - Locate and extract bus details from the search results page.

    - Handle dynamic content loading (pagination, scrolling, etc.).
    - Parse the HTML content to extract desired information.
- **Error Handling**: Implement retries and exception handling for failed page loads or missing elements.

2. Data Preprocessing:

- **Storage**: Save the scraped data into a Pandas Data Frame or CSV for easy access.
- **Clean Data**: Handle any missing or incomplete data, standardize formats (e.g., convert time to a consistent format).

3. Building Streamlit UI:

- **Design Interface**:
    - Use Streamlit widgets (sliders, dropdowns, multiselect, etc.) to create filters and sorting options.
    - Display the filtered bus data dynamically.
- **Real-Time Updates**: Ensure that filters instantly update the displayed data.
- **Styling**: Customize Streamlit components (e.g., cards, tables) to improve user experience.

4. Dynamic Filtering:

- Use Pandas to apply filters to the data in real-time based on the user's input.
- Implement sliders for price and time filtering.
- Allow sorting and searching functionality to enhance the user experience.

5. Deployment:

- Host the Streamlit app using a cloud platform (e.g., Heroku, Streamlit Sharing, or AWS).
- Automate the scraping process to update the bus data periodically (optional, using tools like cron jobs).

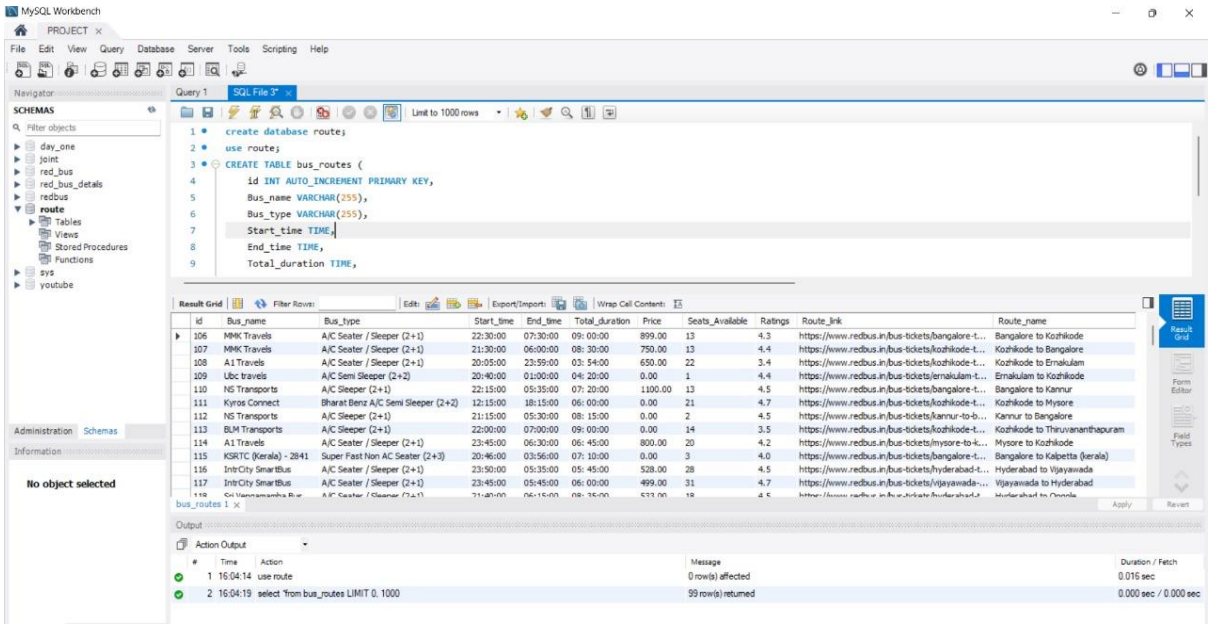
Step 1:

	Route_name	Route_link
0	Bangalore to Kozhikode	https://www.redbus.in/bus-tickets/bangalore-to...
1	Kozhikode to Bangalore	https://www.redbus.in/bus-tickets/kozhikode-to...
2	Kozhikode to Ernakulam	https://www.redbus.in/bus-tickets/kozhikode-to...
3	Ernakulam to Kozhikode	https://www.redbus.in/bus-tickets/ernakulam-to...
4	Bangalore to Kannur	https://www.redbus.in/bus-tickets/bangalore-to...
5	Kozhikode to Mysore	https://www.redbus.in/bus-tickets/kozhikode-to...
6	Kannur to Bangalore	https://www.redbus.in/bus-tickets/kannur-to-ba...
7	Kozhikode to Thiruvananthapuram	https://www.redbus.in/bus-tickets/kozhikode-to...
8	Mysore to Kozhikode	https://www.redbus.in/bus-tickets/mysore-to-ko...
9	Bangalore to Kalpetta (kerala)	https://www.redbus.in/bus-tickets/bangalore-to...

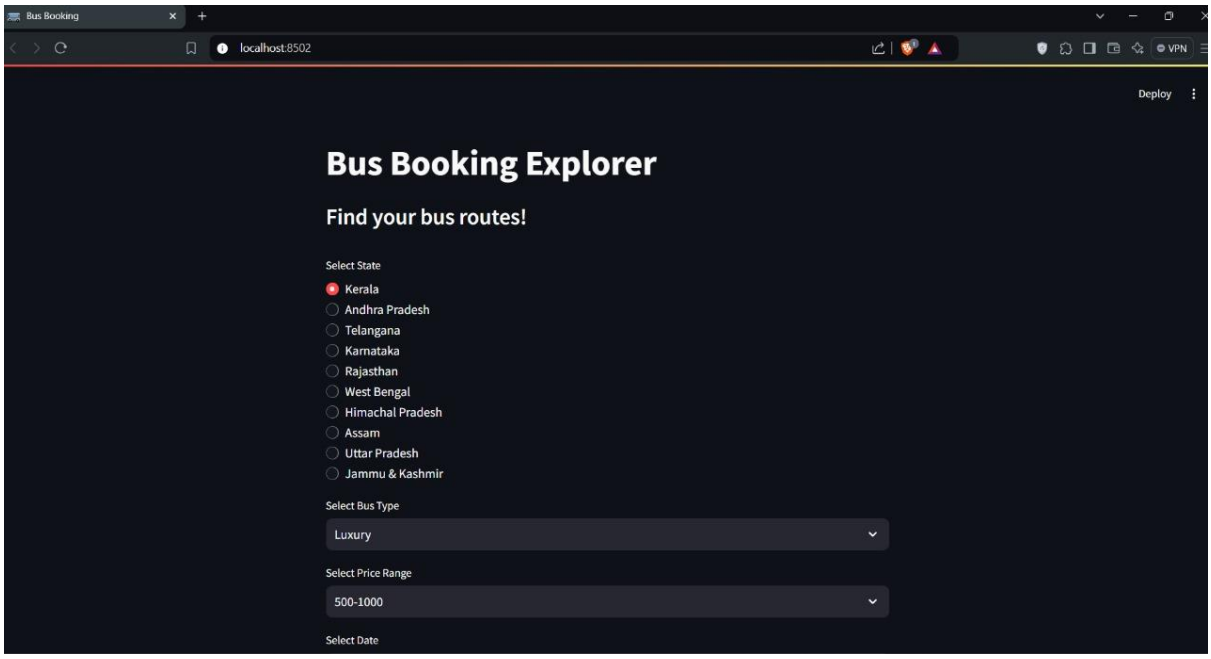
Step 2:

	Bus_name	Bus_type	Start_time	End_time	Total_duration	Price	Seats_Available	Ratings	Route_link	Route_name
0	MMK Travels	A/C Seater / Sleeper (2+1)	22:30	07:30	09h 00m	Starts from\nINR 899	13 Seats available\n3 Window	4.3	https://www.redbus.in/bus-tickets/bangalore-to...	Bangalore to Kozhikode
1	MMK Travels	A/C Seater / Sleeper (2+1)	21:30	06:00	08h 30m	Starts from\nINR 750	13 Seats available\n4 Single	4.4	https://www.redbus.in/bus-tickets/kozhikode-to...	Kozhikode to Bangalore
2	A1 Travels	A/C Seater / Sleeper (2+1)	20:05	23:59	03h 54m	Starts from\nINR 650	22 Seats available\n8 Single	3.4	https://www.redbus.in/bus-tickets/kozhikode-to...	Kozhikode to Ernakulam
3	Ubc travels	A/C Semi Sleeper (2+2)	20:40	01:00	04h 20m	INR 900	1 Seat available	4.4	https://www.redbus.in/bus-tickets/ernakulam-to...	Ernakulam to Kozhikode
4	NS Transports	A/C Sleeper (2+1)	22:15	05:35	07h 20m	Starts from\nINR 1100	13 Seats available\n1 Single	4.5	https://www.redbus.in/bus-tickets/bangalore-to...	Bangalore to Kannur
5	Kyros Connect	Bharat Benz A/C Semi Sleeper (2+2)	12:15	18:15	06h 00m	INR 599	21 Seats available\n6 Window	4.7	https://www.redbus.in/bus-tickets/kozhikode-to...	Kozhikode to Mysore
6	NS Transports	A/C Sleeper (2+1)	21:15	05:30	08h 15m	INR 999	2 Seats available	4.5	https://www.redbus.in/bus-tickets/kannur-to-ba...	Kannur to Bangalore
7	BLM Transports	A/C Sleeper (2+1)	22:00	07:00	09h 00m	Starts from\nINR 1450 1378\nredDeal applied	14 Seats available\n3 Single	3.5	https://www.redbus.in/bus-tickets/kozhikode-to...	Kozhikode to Thiruvananthapuram
8	A1 Travels	A/C Seater / Sleeper (2+1)	23:45	06:30	06h 45m	Starts from\nINR 800	20 Seats available\n9 Single	4.2	https://www.redbus.in/bus-tickets/mysore-to-ko...	Mysore to Kozhikode

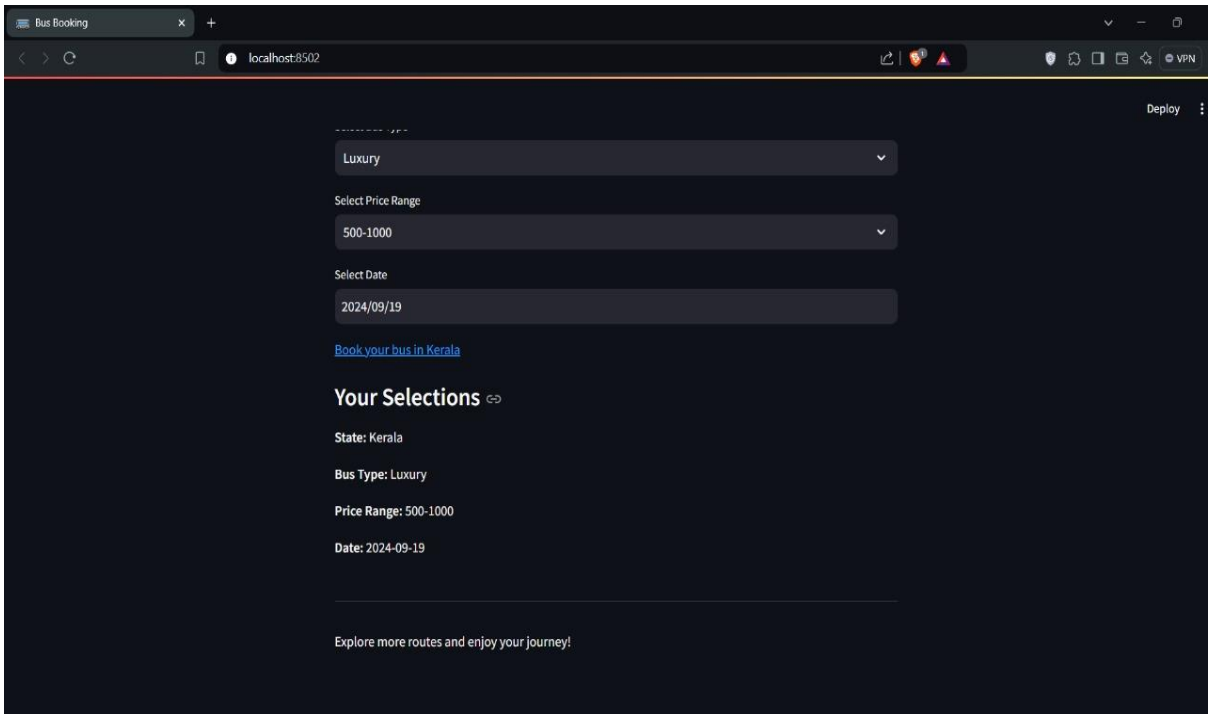
Step 3:



Step 4:


Bus Booking
localhost:8502
VPN
Deploy
Bus Booking Explorer
Find your bus routes!
Select State
Kerala
Andhra Pradesh
Telangana
Karnataka
Rajasthan
West Bengal
Himachal Pradesh
Assam
Uttar Pradesh
Jammu & Kashmir
Select Bus Type
Luxury
Select Price Range
500-1000
Select Date

Step 5:


Bus Booking
localhost:8502
VPN
Deploy
Luxury
Select Price Range
500-1000
Select Date
2024/09/19
Book your bus in Kerala
Your Selections
State: Kerala
Bus Type: Luxury
Price Range: 500-1000
Date: 2024-09-19
Explore more routes and enjoy your journey!

Challenges & Solutions:

- **Dynamic Loading**: Handling lazy-loaded bus data on the website with Selenium by simulating user scrolling or loading more results.
- **Captcha or Anti-Scraping Measures**: Employ appropriate techniques like browser simulation, user-agents, or waiting times to avoid detection.
- **Data Volume**: Efficiently manage large data volumes by paginating results on the front-end and compressing data where possible.

Future Enhancements:

1. **Bus Detail Pages**: Scrape individual bus detail pages to get more information like bus amenities, boarding points, etc.
2. **Historical Price Trends**: Analyze the price changes over time and suggest the best booking time.
3. **Additional Filters**: Add filters for amenities like Wi-Fi, AC, or sleeper coaches.
4. **User Authentication**: Allow users to save their preferences and routes.
5. **Notification Feature**: Add a feature to notify users of price drops or seat availability.

Conclusion:

This project provides an effective way to extract and visualize bus data from Red Bus using Selenium for scraping and Streamlit for the front-end. The dynamic filtering capability ensures that users can find the most suitable bus quickly and easily.