

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:

1. Web Scraping with Selenium:

- **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:

MySQL Workbench

PROJECT

File Edit View Query Database Server Tools Scripting Help

Navigator

SCHEMAS

Filter objects

- day_one
- joint
- red_bus
- red_bus_details
- redbus
- route
- Tables
- Views
- Stored Procedures
- Functions
- sys
- youtube

Administration Schemas Information

No object selected

Query 1 SQL File 3

```

1 create database route;
2 use route;
3 CREATE TABLE bus_routes (
4     id INT AUTO_INCREMENT PRIMARY KEY,
5     Bus_name VARCHAR(255),
6     Bus_type VARCHAR(255),
7     Start_time TIME,
8     End_time TIME,
9     Total_duration TIME,

```

Result Grid

	id	Bus_name	Bus_type	Start_time	End_time	Total_duration	Price	Seats_Available	Ratings	Route_link	Route_name
106	MMK Travels	A/C Seater / Sleeper (2+1)	22:30:00	07:30:00	09:00:00	899.00	13	4.3	https://www.redbus.in/bus-tickets/bangalore-to...	Bangalore to Kozhikode	
107	MMK Travels	A/C Seater / Sleeper (2+1)	21:30:00	06:00:00	08:30:00	750.00	13	4.4	https://www.redbus.in/bus-tickets/kozhikode-t...	Kozhikode to Bangalore	
108	A1 Travels	A/C Seater / Sleeper (2+1)	20:05:00	23:59:00	03:54:00	650.00	22	3.4	https://www.redbus.in/bus-tickets/kozhikode-t...	Kozhikode to Ernakulam	
109	Ubc travels	A/C Semi Sleeper (2+2)	20:40:00	01:00:00	04:20:00	900.00	1	4.4	https://www.redbus.in/bus-tickets/ernakulam-t...	Ernakulam to Kozhikode	
110	NS Transports	A/C Sleeper (2+1)	22:15:00	05:35:00	07:20:00	1100.00	13	4.5	https://www.redbus.in/bus-tickets/bangalore-t...	Bangalore to Kannur	
111	Kyros Connect	Bharat Benz A/C Semi Sleeper (2+2)	12:15:00	18:15:00	06:00:00	599.00	21	4.7	https://www.redbus.in/bus-tickets/kozhikode-t...	Kozhikode to Mysore	
112	NS Transports	A/C Sleeper (2+1)	21:15:00	05:30:00	08:15:00	999.00	2	4.5	https://www.redbus.in/bus-tickets/kannur-to-b...	Kannur to Bangalore	
113	BLM Transports	A/C Sleeper (2+1)	22:00:00	07:00:00	09:00:00	1450.00	14	3.5	https://www.redbus.in/bus-tickets/kozhikode-t...	Kozhikode to Thiruvananthapuram	
114	A1 Travels	A/C Seater / Sleeper (2+1)	23:45:00	06:30:00	06:45:00	800.00	20	4.2	https://www.redbus.in/bus-tickets/mysore-to-...	Mysore to Kozhikode	
115	KSRTC (Kerala) - 2841	Super Fast Non A/C Seater (2+3)	20:46:00	03:56:00	07:10:00	0.00	3	4.0	https://www.redbus.in/bus-tickets/bangalore-t...	Bangalore to Kalletta (Kerala)	
116	IntCity SmartBus	A/C Seater / Sleeper (2+1)	23:50:00	05:45:00	05:45:00	528.00	28	4.5	https://www.redbus.in/bus-tickets/hyderabad-t...	Hyderabad to Vijayawada	
117	IntCity SmartBus	A/C Seater / Sleeper (2+1)	23:45:00	05:45:00	06:00:00	499.00	31	4.7	https://www.redbus.in/bus-tickets/vijayawada-...	Vijayawada to Hyderabad	
118	Ch Vannanaptha Bus	A/C Seater / Sleeper (2+1)	21:40:00	06:15:00	08:35:00	573.00	18	4.5	https://www.redbus.in/bus-tickets/bangalore-t...	Bangalore to Chennai	

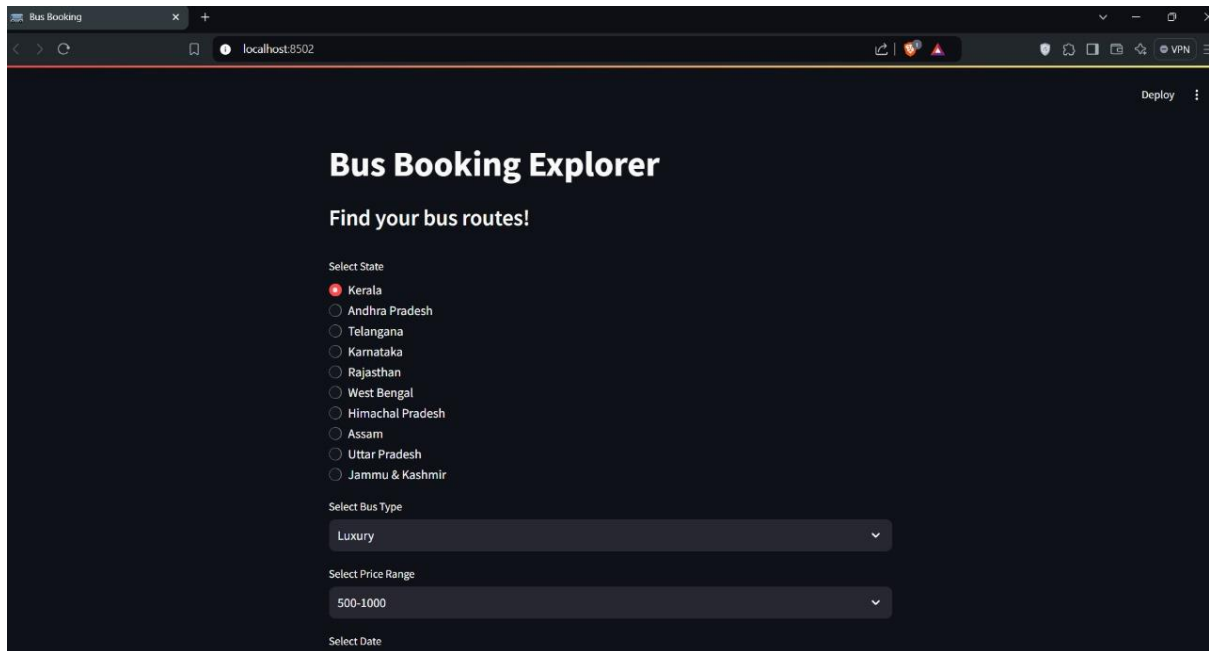
bus_routes 1 x

Output

Action Output

#	Time	Action	Message	Duration / Fetch
1	16:04:14	use route	0 row(s) affected	0.015 sec
2	16:04:19	select * from bus_routes LIMIT 0, 1000	99 row(s) returned	0.000 sec / 0.000 sec

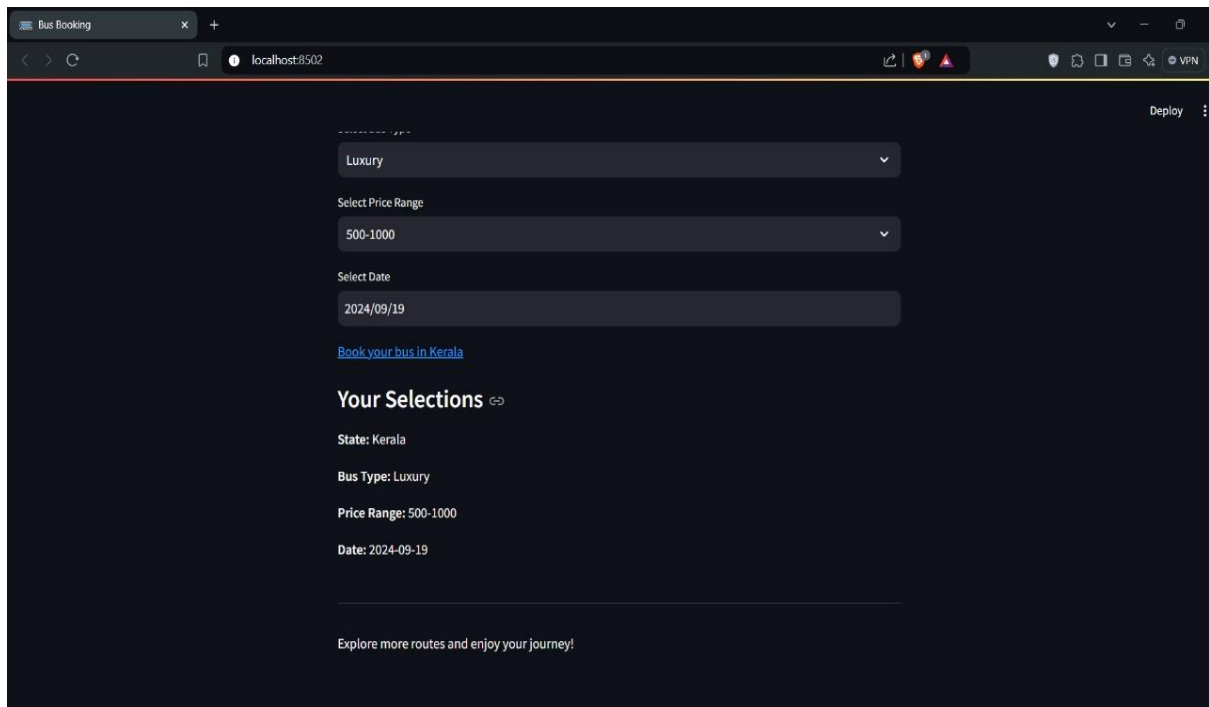
Step 4:



The screenshot shows a web browser window with the address bar displaying 'localhost:8502'. The page title is 'Bus Booking Explorer'. Below the title, the text 'Find your bus routes!' is displayed. The form contains the following elements:

- Select State:** A list of radio buttons with 'Kerala' selected. Other options include Andhra Pradesh, Telangana, Karnataka, Rajasthan, West Bengal, Himachal Pradesh, Assam, Uttar Pradesh, and Jammu & Kashmir.
- Select Bus Type:** A dropdown menu with 'Luxury' selected.
- Select Price Range:** A dropdown menu with '500-1000' selected.
- Select Date:** A label for the date selection field, which is currently empty.

Step 5:



The screenshot shows the same web browser window, but now the 'Your Selections' section is visible. The form elements are the same as in Step 4, but the 'Select Date' field now displays '2024/09/19'. Below the form, the 'Your Selections' section is displayed with the following information:

- State:** Kerala
- Bus Type:** Luxury
- Price Range:** 500-1000
- Date:** 2024-09-19

Below the selections, there is a link that says 'Book your bus in Kerala' and a button that says '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.