

Redbus Data Scraping with Selenium & Dynamic Filtering using Streamlit

Project Overview

This project involves scraping bus route data from the Redbus website using Selenium, storing the data in a SQL database, and developing an interactive Streamlit application for data filtering. This project aims to demonstrate web scraping techniques, data storage, and creating a user-friendly interface for data analysis and filtering.

Technologies Used

- **Python:** For scripting and automation.
- **Selenium:** For web scraping and extracting data from the Redbus website.
- **MySQL:** For storing the scraped data.
- **Streamlit:** For building an interactive web application.
- **Libraries:** Pandas, NumPy, MySQL Connector.

Dataset

The dataset used in this project is scraped from the Redbus website. It includes information about various bus routes, bus types, departure and arrival times, prices, ratings, and seat availability.

Database Schema

The data is stored in a MySQL database with the following schema:

- **id:** Auto-incrementing primary key
- **route_name:** TEXT
- **route_link:** TEXT
- **busname:** TEXT
- **bustype:** TEXT
- **departing_time:** DATETIME
- **duration:** TEXT
- **reaching_time:** DATETIME
- **star_rating:** FLOAT
- **price:** DECIMAL
- **seats_available:** INT

Implementation Steps:

1. Web Scraping with Selenium

- **Objective:** Automate the extraction of bus travel data from the Redbus website.
- **Tools Used:** Python, Selenium WebDriver, ChromeDriver
- **Process:**
 1. **Setup:** Install and configure Selenium and ChromeDriver.
 2. **Automation:** Write Python scripts to navigate the Redbus website, locate and extract data fields such as route name, bus type, departure time, duration, arrival time, star rating, price, and seats available.
 3. **Data Extraction:** Use Selenium to dynamically interact with web elements and extract the required data.

2. Data Storage

- **Objective:** Store the extracted data in a structured format for efficient retrieval and analysis.
- **Tools Used:** MySQL, SQLAlchemy
- **Process:**
 1. **Database Setup:** Create a MySQL database named redbus.
 2. **Schema Design:** Define the database schema with appropriate tables and columns.
 3. **Data Insertion:** Write scripts to insert the scraped data into the database.

3. Data Cleaning and Preprocessing

- **Objective:** Ensure the data is clean and ready for analysis.
- **Tools Used:** Python, Pandas
- **Process:**
 1. **Data Cleaning:** Remove any inconsistencies, handle missing values, and correct data types.
 2. **Preprocessing:** Prepare the data for analysis and visualization by normalizing and formatting it.

4. Streamlit Application Development

- **Objective:** Develop an interactive web application for data filtering and visualization.
- **Tools Used:** Streamlit, Python, Pandas
- **Process:**
 1. **Setup:** Install and configure Streamlit.
 2. **Interface Design:** Design a user-friendly interface with options for filtering data based on various parameters (route, bus type, departure time, price).
 3. **Visualization:** Create dynamic visualizations to display filtered data using Matplotlib and Seaborn.

Features

- **Real-time Data Extraction:** Scrapes live data from the Redbus website, ensuring up-to-date information.
- **Interactive Interface:** User-friendly Streamlit application for easy navigation and data filtering.
- **Dynamic Filtering:** Allows users to filter data based on multiple criteria, providing flexibility in data analysis.
- **Data Storage:** Efficiently stores and retrieves data from a MySQL database.

Key Results

- Successfully automated the extraction of bus travel data from the Redbus website.
- Structured and stored the data in a MySQL database for efficient retrieval and analysis.
- Developed an interactive Streamlit application that allows users to filter and visualize bus travel data in real-time.
- Provided valuable insights into travel patterns, helping users make informed travel choices.

Future Work

- Expand the application to include more filtering options and additional data fields.
- Implement real-time data updates to keep the information current.
- Enhance the visualizations with more detailed analysis and insights.

Conclusion

This project demonstrates the effective use of web scraping, data storage, and web application development to create a tool that provides valuable insights into bus route data. The combination of Selenium, MySQL, and Streamlit offers a powerful solution for dynamic data analysis.

Sample Outputs:

Image 1

Filters

Select Route

Choose an option

Kozhikode to Bangalore Bus

Chittoor (Andhra Pradesh) to Bang...

Kadapa to Hyderabad Bus

Mysore to Kozhikode Bus

Motihari to Agra Bus

Delhi to Hamirpur (Himachal Prade...

Badaun to Delhi Bus

Kadapa to Bangalore Bus

Seats Available

6

-

+

Deploy

Bus Route Information

	id	route_name	route_id	busname	bustype	departing_time	duration	reaching_time
434	435	Delhi to Shi	https://v	Bedi Travels	Volvo A/C Sei	1900-01-01 23: 09h 00m		1900-01-01 08:00:00
435	436	Delhi to Shi	https://v	Ram Dalal Holic	Mercedes Mu	1900-01-01 23: 08h 10m		1900-01-01 07:10:00
436	437	Delhi to Shi	https://v	Laxmi holidays	Scania Multi	1900-01-01 23: 08h 30m		1900-01-01 08:00:00
438	439	Delhi to Shi	https://v	Zingbus Plus	Volvo A/C Sei	1900-01-01 22: 08h 30m		1900-01-01 07:25:00
439	440	Delhi to Shi	https://v	Laxmi holidays	Volvo A/C Sei	1900-01-01 23: 08h 20m		1900-01-01 07:45:00
440	441	Delhi to Shi	https://v	Ram Dalal Holic	Volvo A/C Sei	1900-01-01 21: 09h 50m		1900-01-01 07:40:00
441	442	Delhi to Shi	https://v	Laxmi holidays	Bharat Benz	1900-01-01 22: 08h 50m		1900-01-01 07:40:00
446	447	Udaipur to	https://v	RSRTC - 163481	Express Non	1900-01-01 05: 19h 00m		1900-01-01 00:00:00
447	448	Udaipur to	https://v	RSRTC - 169076	Express Non	1900-01-01 12: 11h 50m		1900-01-01 00:00:00
448	449	Udaipur to	https://v	RSRTC - 193271	Express Non	1900-01-01 19: 06h 30m		1900-01-01 02:00:00

Image 2:

Filters

Select Route

Mysore to Kozhi...

Select Bus Type

Choose an option

Select Star Rating

3.50

1.00

4.00

Select Price Range

5000.00

186.00

5000.00

Seats Available

6

-

+

Deploy

Bus Route Information

	id	route_name	route_id	busname	bustype	departing_time	duration	reaching_time
128	129	Mysore to K	https://v	KSRTC (Kerala)	Swift Deluxe	1900-01-01 17: 06h 31m		1900-01-01 00:01:00
129	130	Mysore to K	https://v	KSRTC (Kerala)	Super Fast Ni	1900-01-01 18: 05h 30m		1900-01-01 00:00:00
130	131	Mysore to K	https://v	KSRTC (Kerala)	Swift Deluxe	1900-01-01 18: 05h 59m		1900-01-01 00:50:00
131	132	Mysore to K	https://v	KSRTC (Kerala)	Swift Deluxe	1900-01-01 19: 06h 10m		1900-01-01 01:20:00
132	133	Mysore to K	https://v	KSRTC (Kerala)	Super Deluxe	1900-01-01 19: 05h 35m		1900-01-01 01:25:00
133	134	Mysore to K	https://v	KSRTC (Kerala)	Swift Deluxe	1900-01-01 21: 06h 34m		1900-01-01 04:24:00
134	135	Mysore to K	https://v	KSRTC (Kerala)	Super Fast Ni	1900-01-01 22: 05h 55m		1900-01-01 04:25:00
135	136	Mysore to K	https://v	KSRTC (Kerala)	Super Fast Ni	1900-01-01 22: 07h 20m		1900-01-01 05:50:00
136	137	Mysore to K	https://v	KSRTC (Kerala)	Super Deluxe	1900-01-01 23: 06h 00m		1900-01-01 05:36:00
138	139	Mysore to K	https://v	NS Transports	Non A/C Seat	1900-01-01 00: 06h 00m		1900-01-01 06:30:00

Image 3:

Filters

Select Route

Chennai to Bang...

Select Bus Type

A/C Sleeper (2+1)

Select Star Rating

4.00

1.004.00

Select Price Range

2359.00

186.005000.00

Seats Available

7

Deploy

Bus Route Information

	+ id	route_nam	route_lir	busname	bustype	departing_tim	duration	reaching_time
	229	230	Chennai to	https://v	LIMOLINER	A/C Sleeper (1900-01-01 19: 11h 00m	1900-01-01 06:30:00
	232	233	Chennai to	https://v	Ramana Tours /	A/C Sleeper (1900-01-01 18: 12h 00m	1900-01-01 06:45:00

Image 4:

Filters

Select Route

Ongole to Hyder...

Select Bus Type

Super Luxury (N...

Select Star Rating

4.00

1.004.00

Select Price Range

2359.00

186.005000.00

Seats Available

7

Deploy

Bus Route Information

	+ id	route_nam	route_lir	busname	bustype	departing_tim	duration	reaching_time
	296	297	Ongole to H	https://v	APSRTC - 5866	Super Luxury	1900-01-01 19: 10h 01m	1900-01-01 05:46:00
	298	299	Ongole to H	https://v	APSRTC - 45901	Super Luxury	1900-01-01 21: 07h 20m	1900-01-01 04:20:00
	302	303	Ongole to H	https://v	APSRTC - 5927	Super Luxury	1900-01-01 21: 06h 30m	1900-01-01 04:15:00
	308	309	Ongole to H	https://v	APSRTC - 5869	Super Luxury	1900-01-01 22: 06h 30m	1900-01-01 04:45:00
	313	314	Ongole to H	https://v	APSRTC - 5925	Super Luxury	1900-01-01 22: 05h 30m	1900-01-01 04:15:00
	314	315	Ongole to H	https://v	APSRTC - 5921	Super Luxury	1900-01-01 22: 07h 36m	1900-01-01 06:21:00
	322	323	Ongole to H	https://v	APSRTC - 48822	Super Luxury	1900-01-01 23: 06h 30m	1900-01-01 06:20:00

Image 5:

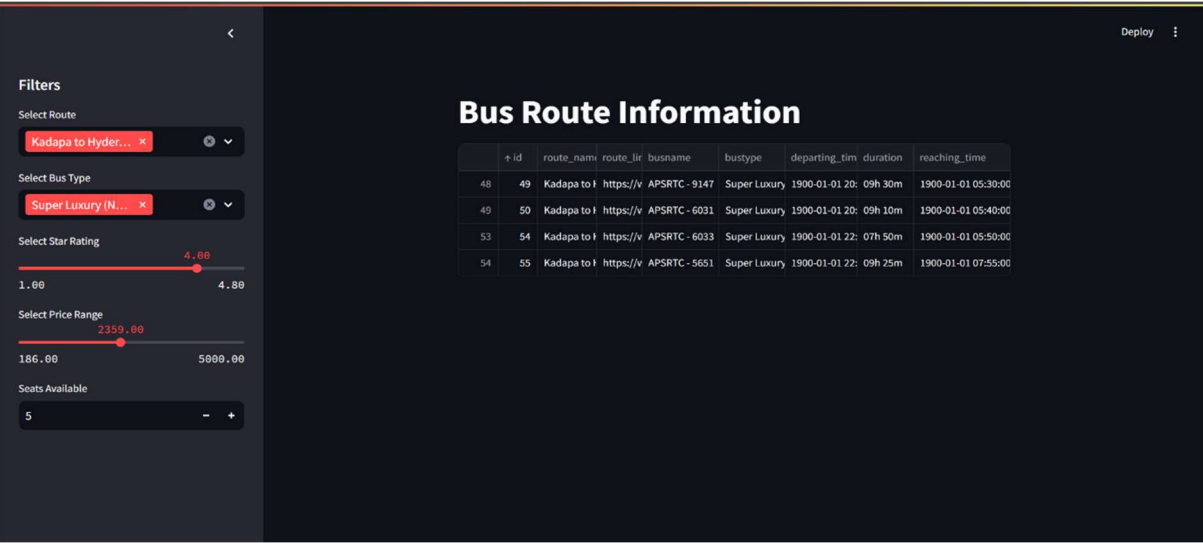


Image 6:



Image 7:

Filters

Select Route

Kadapa to Bang...

Select Bus Type

Super Luxury (N...

Select Star Rating

4.00

1.004.00

Select Price Range

2359.00

186.005000.00

Seats Available

6

Deploy

Bus Route Information

	id	route_name	route_id	busname	bustype	departing_time	duration	reaching_time	s
169	170	Kadapa to f	https://v	APSRTC - 6207	Super Luxury	1900-01-01 22: 05h 56m		1900-01-01 04:26:00	
171	172	Kadapa to f	https://v	APSRTC - 5227	Super Luxury	1900-01-01 22: 06h 45m		1900-01-01 05:30:00	
172	173	Kadapa to f	https://v	APSRTC - 6065	Super Luxury	1900-01-01 22: 05h 50m		1900-01-01 04:35:00	
176	177	Kadapa to f	https://v	APSRTC - 48483	Super Luxury	1900-01-01 23: 06h 45m		1900-01-01 06:00:00	
178	179	Kadapa to f	https://v	APSRTC - 9254	Super Luxury	1900-01-01 23: 06h 00m		1900-01-01 05:30:00	
179	180	Kadapa to f	https://v	APSRTC - 48516	Super Luxury	1900-01-01 23: 07h 00m		1900-01-01 06:30:00	
182	183	Kadapa to f	https://v	APSRTC - 5977	Super Luxury	1900-01-01 23: 06h 40m		1900-01-01 06:30:00	