

Redbus Data Scraping and Filtering with Streamlit Application

A Comprehensive Solution for Transportation Data Management

By
Harikrishnan G

Introduction

Efficient data management is crucial in the transportation industry. The "Redbus Data Scraping and Filtering with Streamlit Application" automates the collection, analysis, and visualization of bus travel data. By leveraging web scraping and interactive dashboards, this project provides real-time insights into bus schedules, prices, and seat availability, significantly enhancing decision-making and operational efficiency.

Problem Statement

Problem: Manual data collection is time-consuming and prone to errors, making it difficult to manage bus schedules, prices, and seat availability effectively.

Challenges: Discuss challenges like data inconsistency, lack of real-time information, and difficulty in visualizing complex datasets.

Goal: Automate data collection and provide tools for analysis and visualization.

Solution Overview

Data Collection: Using Selenium for web scraping.

Data Storage: Storing data in a SQL database.

Data Visualization: Using Streamlit to create an interactive user interface.

Outcome: Streamlined data management and enhanced decision-making capabilities.

Technology Stack

Web Scraping: Selenium for automated data extraction from Redbus.

Database: MySQL for structured data storage and querying.

Backend: Python for data processing and backend logic.

Frontend: Streamlit for creating an interactive dashboard.

Tools/libraries: Pandas, MySQL Connector,time,selenium & streamlit

Web Scrapping with Selenium

Explanation: Describe how Selenium is used to navigate Redbus, extract data, and handle dynamic content.

Features:

Automation of repetitive tasks.

Collection of complex data like bus schedules, prices, and seat availability.

Data Storage with MySQL

Server: 127.0.0.1 » Database: csv_db 10 » Table: bus_routes

[Browse](#) [Structure](#) [SQL](#) [Search](#) [Insert](#) [Export](#) [Import](#) [Privileges](#) [Oper](#)

[Table structure](#) [Relation view](#)

	#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
<input type="checkbox"/>	1	id	int(11)			No	None			Change Drop More
<input type="checkbox"/>	2	State	varchar(10)	utf8_general_ci		Yes	NULL			Change Drop More
<input type="checkbox"/>	3	route_name	text	utf8_general_ci		Yes	NULL			Change Drop More
<input type="checkbox"/>	4	route_link	text	utf8_general_ci		Yes	NULL			Change Drop More
<input type="checkbox"/>	5	busname	text	utf8_general_ci		Yes	NULL			Change Drop More
<input type="checkbox"/>	6	bustype	text	utf8_general_ci		Yes	NULL			Change Drop More
<input type="checkbox"/>	7	departing_time	time			Yes	NULL			Change Drop More
<input type="checkbox"/>	8	departure_location	text	utf8_general_ci		Yes	NULL			Change Drop More
<input type="checkbox"/>	9	duration	text	utf8_general_ci		Yes	NULL			Change Drop More
<input type="checkbox"/>	10	reaching_time	time			Yes	NULL			Change Drop More
<input type="checkbox"/>	11	arrival_location	text	utf8_general_ci		Yes	NULL			Change Drop More
<input type="checkbox"/>	12	star_rating	float			Yes	NULL			Change Drop More
<input type="checkbox"/>	13	price	decimal(8,0)			Yes	NULL			Change Drop More
<input type="checkbox"/>	14	seats_available	int(18)			Yes	NULL			Change Drop More

Data Visualization with Streamlit

→ ↺

localhost:8501

70% ☆

🔒 📧 🔔 🏠

Deploy

er Options

by state

RTC

by route_name

one

by busname (optional)

one

by bustype

Deluxe AC Seater 2+2

by departing_time

one

by departure_location

one

by reaching_time

one

by arrival_location

one

by star_rating

4.10

by price

645.00

Bus Routes Database Viewer

This app allows you to view and filter bus routes data from a SQL database.

Connected to database!

Filtered Bus Routes:

		route_link	busname	bustype	departing_time	departure_location	duration	reaching_time	arrival_location	star_rating	price	seats_available
4,257	jasthan	https://www.redbus.in/bus-tickets/sikar-to-jaipur	RSRTC - 193696	Deluxe AC Seater 2+2	18 hours	SIKAR RSRTC BUS STAND	02h 05m	20 hours	JAIPUR RSRTC BUS STAND	2.5	228	36
9,450		https://www.redbus.in/bus-tickets/jodhpur-to-ajmer	RSRTC - 193811	Deluxe AC Seater 2+2	13 hours	JODHPUR RSRTC BUS STAND	04h 25m	17 hours	AJMER RSRTC BUS STAND	3.9	404	37
9,584	y) to Ja	https://www.redbus.in/bus-tickets/beawar-to-jaipur	RSRTC - 193811	Deluxe AC Seater 2+2	16 hours	BEAWAR RSRTC BUS STAND	04h 25m	20 hours	JAIPUR RSRTC BUS STAND	3.6	371	37
9,595	y) to Ja	https://www.redbus.in/bus-tickets/beawar-to-jaipur	RSRTC - 194149	Deluxe AC Seater 2+2	18 hours	BEAWAR RSRTC BUS STAND	05h 15m	a day	JAIPUR RSRTC BUS STAND	3.6	371	37
9,736	jasthan	https://www.redbus.in/bus-tickets/sikar-to-jaipur	RSRTC - 194084	Deluxe AC Seater 2+2	11 hours	SIKAR RSRTC BUS STAND	02h 20m	13 hours	JAIPUR RSRTC BUS STAND	3.9	228	35
9,749	jasthan	https://www.redbus.in/bus-tickets/sikar-to-jaipur	RSRTC - 194044	Deluxe AC Seater 2+2	12 hours	SIKAR RSRTC BUS STAND	02h 00m	14 hours	SINDHI CAMP RSRTC BUS STAND	3.9	228	34
9,882	to Jod	https://www.redbus.in/bus-tickets/jaipur-to-jodhpur	RSRTC - 194132	Deluxe AC Seater 2+2	6 hours	JAIPUR RSRTC BUS STAND	07h 30m	13 hours	JODHPUR RSRTC BUS STAND	3.3	645	37
9,938	r (Raja	https://www.redbus.in/bus-tickets/jodhpur-to-beawar	RSRTC - 193811	Deluxe AC Seater 2+2	13 hours	JODHPUR RSRTC BUS STAND	03h 00m	16 hours	BEAWAR RSRTC BUS STAND	3.7	296	37
10,245		https://www.redbus.in/bus-tickets/sikar-to-bikaner	RSRTC - 194039	Deluxe AC Seater 2+2	7 hours	SIKAR RSRTC BUS STAND	03h 20m	11 hours	BIKANER RSRTC BUS STAND	4.1	431	35

Conclusion

This project can significantly improve operational efficiency, strategic planning, and overall user experience in the transportation sector.