# Capstone

About: This project is all about scraping data from website, storing the extracted information and providing search functionality based on given criteria.

Description:

* Scraping data from <https://www.redbus.in/>.
* Storing of data in structured format in Mysql repository.
* Implementing search functionality using streamlit application.

Installation:

1. Install Selenium using pip: !pip install selenium.
2. Download chrome driver and add a path to environment variables.
3. Download mysql and install.
4. Install streamlit using pip: !pip install streamlit.

Scraping and saving the data:

* Selenium is used to interact with the chrome browser through chrome web driver.
* The Selenium script controls browser actions to navigate the Redbus.in website.
* On running the code, we enter Redbus.in website
* Then manually enter each state service webpage.
* Using code, automatically redirects into each route of that particular state service and gets all the information of every bus in the route.
* Load page until every bus service in that route got loaded.
* Using inspect elements, access elements on the website through classnames or ids or css styles.
* Elements include route link, route name, bus name, bus type, departure time, arriving time, price, seats available etc.
* Create a connection to mysql database.
* Convert all the data into a list and saving it in the MySQL database.

Creating table in mysql:

* Create a database name “mde92”.
* Create a table name “redbusguvi” with columns: primary id, route name,route link,bus name,bus type,departure time,duration,arrival time,price,seats available.
* The scraped data will be saved in the “redbusguvi” table.

Searching the criteria:

* Create a user interface using streamlit.
* User interface elements include:

1. Route name: User can select any route using streamlit dropdown. This drop down is populated from database.
2. Seat type: User can select seater or sleeper from streamlit dropdown which is static.
3. Bus type: User can select A/C or NON-A/C from static streamlit dropdown.
4. Departure time: User can select time range from static streamlit dropdown.
5. Rating: User can slide the range of bus ratings using a Streamlit slider.
6. Price: User can slide the range of prices using a Streamlit slider.
7. Search button: Streamlit button which is used to get data from database based on user selected criteria.

A screenshot of a computer

Description automatically generated

* By clicking the search button, creating the connection to database and getting all the information regarding all the buses which satisfies the criteria where user selected from the above elements.
* Retrieved information is shown in the dataframe using streamlit dataframe.
* For example, if the user selects the "Vijayawada to Rajahmundry" route, "A/C seater" bus type, departure time between 8 PM to 11 PM, ratings ranging from 1 to 5, and price ranging from 0 to 1000, and then clicks the search button, the output might look like the following:

In this scenario, the search criteria result in an output of 2 rows in the dataframe as shown in the image above.

A screenshot of a computer

Description automatically generated

Conclusion:

This project demonstrates a comprehensive approach to web scraping, data storage, and user-friendly data retrieval using Streamlit. By leveraging technologies such as Selenium for web scraping, MySQL for data storage, and Streamlit for creating an interactive user interface, this project provides a robust solution for extracting and managing travel information from the Redbus website. Users can easily search and filter bus services based on their preferences, enhancing the overall travel planning experience.