**Name: Cibiya D**

**Date of submission: 31-Jul-2024**

**Redbus Data Scraping and Filtering with Streamlit Application**

**Overview**

The Redbus Data Scraping and Filtering with Streamlit Application"\*\* project provides a robust solution for collecting, analyzing, and visualizing bus travel data. By leveraging Selenium for web scraping and Streamlit for data visualization, this project aims to enhance the efficiency of the transportation industry through comprehensive data collection and insightful analysis.

**Objectives**

Automate Data Collection: I have used Selenium to scrape detailed information from Redbus, including bus routes, schedules, prices, and seat availability. I have stored the extracted data in a SQL database for structured storage and easy retrieval. Developed a user-friendly Streamlit application to display, filter, and interact with the data, enabling better decision-making and strategic planning.

Approach

Data Scraping:

Tool: Selenium

Objective: Automate the extraction of data from the Redbus website.

Data Collected:

Bus Routes: 10 Government transports bus routes were extracted.

Schedules: Departure and arrival times for buses.

Prices: Fare details for various bus services.

Seat Availability: Information on available seats.

Steps:

* Navigate to Redbus: Open the Redbus website and access the relevant sections for bus routes.
* Extract Data: Use Selenium to identify and extract data points such as routes, schedules, and pricing.
* Handle Dynamic Content: Implement scrolling and interaction to ensure all data is collected.

2. Data Storage

Tool: SQL Database

Objective: Stored scraped data in a structured format for easy querying and analysis.

Database Schema: Define tables for routes, schedules, prices, and availability to organize the data efficiently.

Steps:

* Design Schema: Create a database schema that supports the types of data collected.
* Insert Data: Use SQL queries to insert scraped data into the database.
* Ensure Integrity: Implement checks to maintain data consistency and accuracy.

3. Streamlit Application

Tool: Streamlit

Objective: Develop an interactive web application to display and filter the scraped data.

Features:

Data Display: Show routes, schedules, and pricing information.

Filters: Allow users to filter data by bus type, route, price range, star rating, and availability.

Steps:

* Develop Interface: A user-friendly interface for displaying data created.
* Implement Filters: Filter options are added to refine search results based on user preferences.
* Connect to Database: Retrieved data from the SQL database and update the Streamlit application dynamically.

4. Data Analysis/Filtering

Tool: SQL Queries and Streamlit Filters

Objective: Enable users to analyze and filter data based on specific criteria.

Methods:

SQL Queries: Retrieve and filter data from the database based on user inputs.

Streamlit Interactions: Use Streamlit components to provide interactive data filtering.

Steps:

* Query Optimization: Write efficient SQL queries to handle large datasets and provide quick responses.
* Interactive Filters: Implement interactive elements in Streamlit to allow users to customize their data view.
* Data Visualization: Display filtered results in an easy-to-understand format.

Conclusion

The "Redbus Data Scraping and Filtering with Streamlit Application" project offers a comprehensive solution for collecting and analyzing bus travel data. By combining Selenium for scraping, SQL for data storage, and Streamlit for visualization, this project aims to enhance decision-making and operational efficiency in the transportation industry. The application supports various business scenarios, from improving customer service to conducting market and competitor analysis, ultimately driving strategic planning and business growth.