**REDBUS DATA SCRAPING USING SELENIUM**

**Project Overview**: This project is to scrape bus route details from redbus website and store them in a database. The data consists of information such as bus name, type, departure/arrival times, price, ratings, number of available seats, and others. This data is then applied in a Streamlit application where users can interface with it in order to compare bus information among different routes.

**Data Collection:** Data is collected using Selenium WebDriver an automation tool for browsers. The key information scraped includes Bus name, type, start time, end time, duration, price, seats available, ratings, route name, and route link.

The process includes:

* **Setup Initial:** A Selenium WebDriver object is initialized to interact with the bus booking website. XPath expressions are utilized to extract bus route links and traverse through multiple pages to gather all route details.
* **Data Scraping:** Bus route links are scraped and opened separately in order to gather in-depth bus data. Pagination is automatically taken care of in order to scrape information from all accessible pages.
* **Error Handling:** The application covers missing information and slow-loading websites with exception handling (e.g., NoSuchElementException, TimeoutException).
* **Data Organization:** The extracted information is placed in lists (e.g., Busname\_1, Bustype\_1) and transferred into a Pandas DataFrame. Final information is cleaned and stored to a CSV file for data storage and reusability.
* **Data Storage & Integration:** The data scraped is stored as a CSV and loaded into a MySQL database for organized storage. This enables simple querying and integration with the Streamlit app.

**Streamlit Application:** The Streamlit app offers an interactive interface for showing bus route data. Users are able to:

* Search for buses by route.
* See bus information such as timings, price, availability, and ratings.
* Filter results through dropdowns and search bars.
* Look at the bus route map (if available).

**Technologies Used:**

*Selenium:* Browser automation and web scraping.

*Pandas*: Data structuring and manipulation.

*MySQL:* Relational database for storing data.

*Streamlit*: Data visualization and web app development.