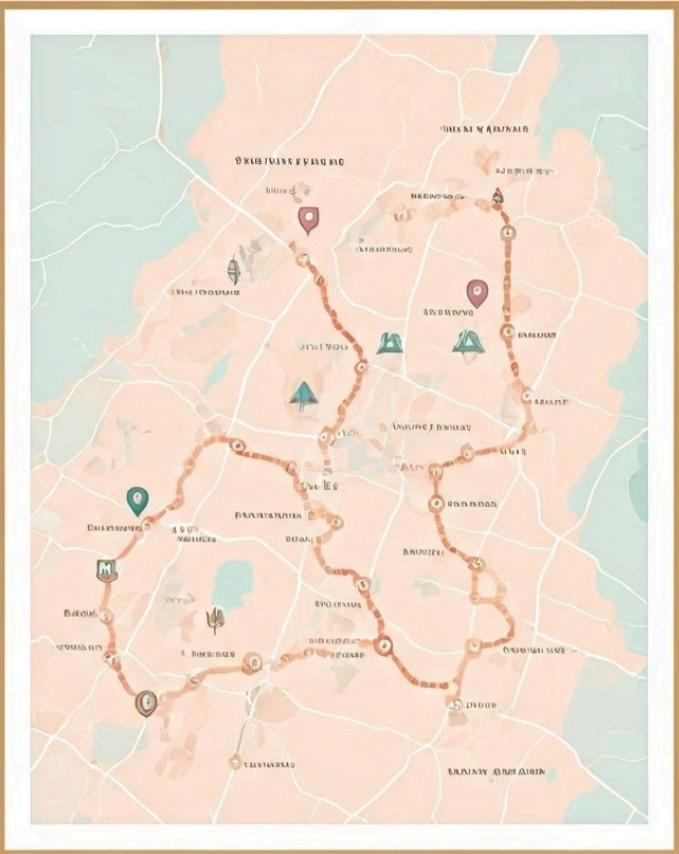


Redbus Data Scraping and Filtering with Streamlit Application

The 'Redbus Data Scraping and Filtering with Streamlit Application' is a powerful tool for collecting and analyzing bus travel data. This project utilizes Selenium for web scraping, automatically extracting detailed information from Redbus, including bus routes, schedules, prices, and seat availability.



by Jayasurya



Domain: Transportation

1 Bus Travel Optimization

Data analysis can help optimize bus routes, schedules, and pricing to improve efficiency and passenger satisfaction.

2 Competitive Analysis

By analyzing competitor data, companies can identify opportunities to improve their services and gain a competitive advantage.

3 Market Research

Data on bus travel patterns can provide valuable insights into market trends and customer preferences.

4 Operational Efficiency

Data analysis can identify areas for improvement in operations, such as reducing wait times and minimizing fuel consumption.



Skill Takeaways

Python Scripting

This project involves writing Python scripts to automate web scraping, data processing, and database interactions.

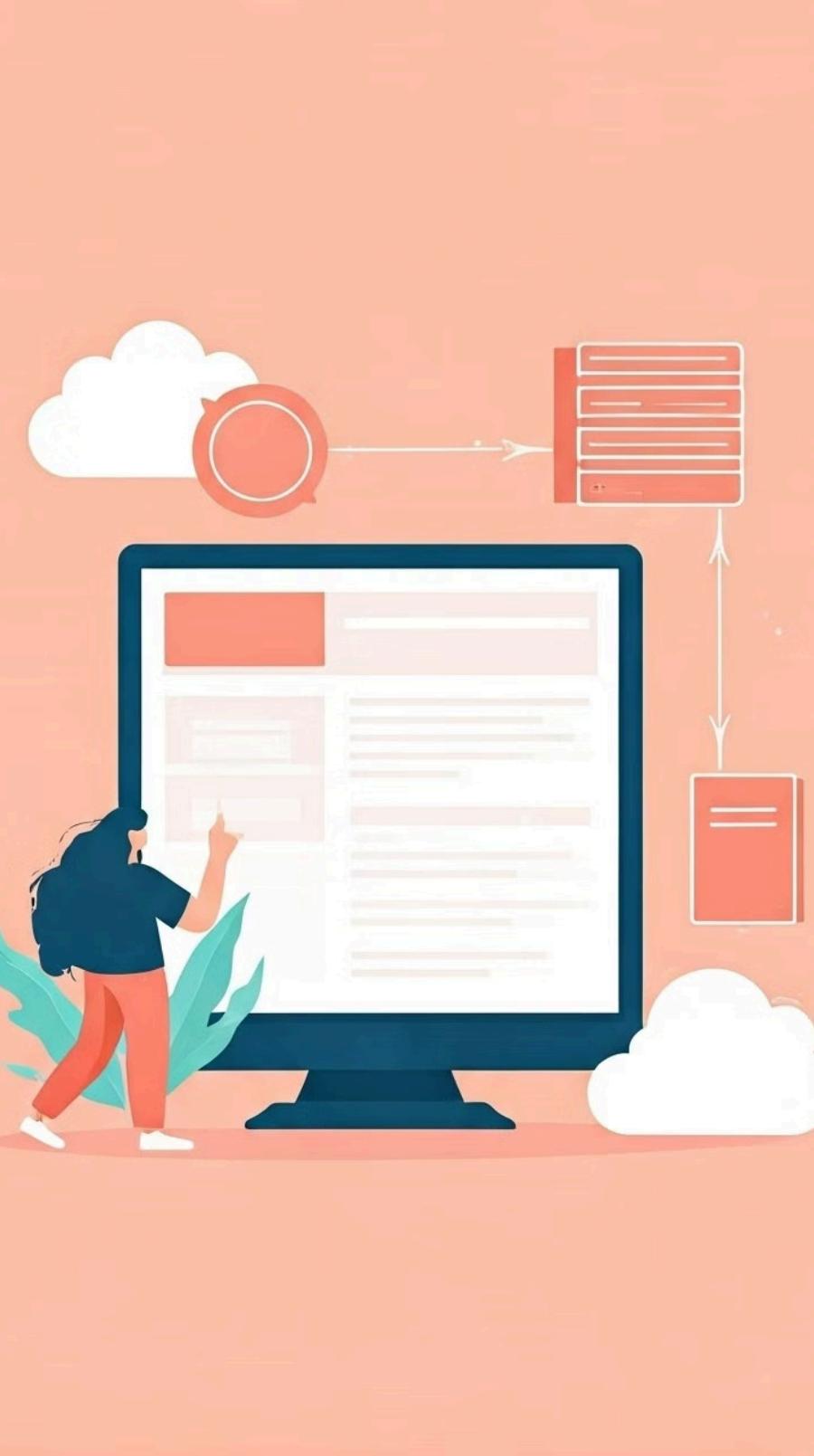
Selenium

Selenium is a powerful tool for web scraping and automating browser interactions, providing robust control over web elements.

Data Management using SQL

SQL is used to create and manage databases, storing and querying the collected bus data for analysis.

Technology Used



Python 3.9.I

The project utilizes Python 3.9.I as the primary programming language for scripting and data manipulation.

MySQL 8.0

MySQL 8.0 is used as the database management system to store and organize the collected bus data.

Streamlit

Streamlit is used to create a user-friendly web application, enabling interactive data visualization and user input.

Selenium

Selenium is the core library for web scraping, automating browser interactions to extract data from the Redbus website.

Retrieve the Bus Information

1

Web Scraping with Selenium

Selenium automates the process of searching for bus routes and extracting relevant details from the Redbus website.

2

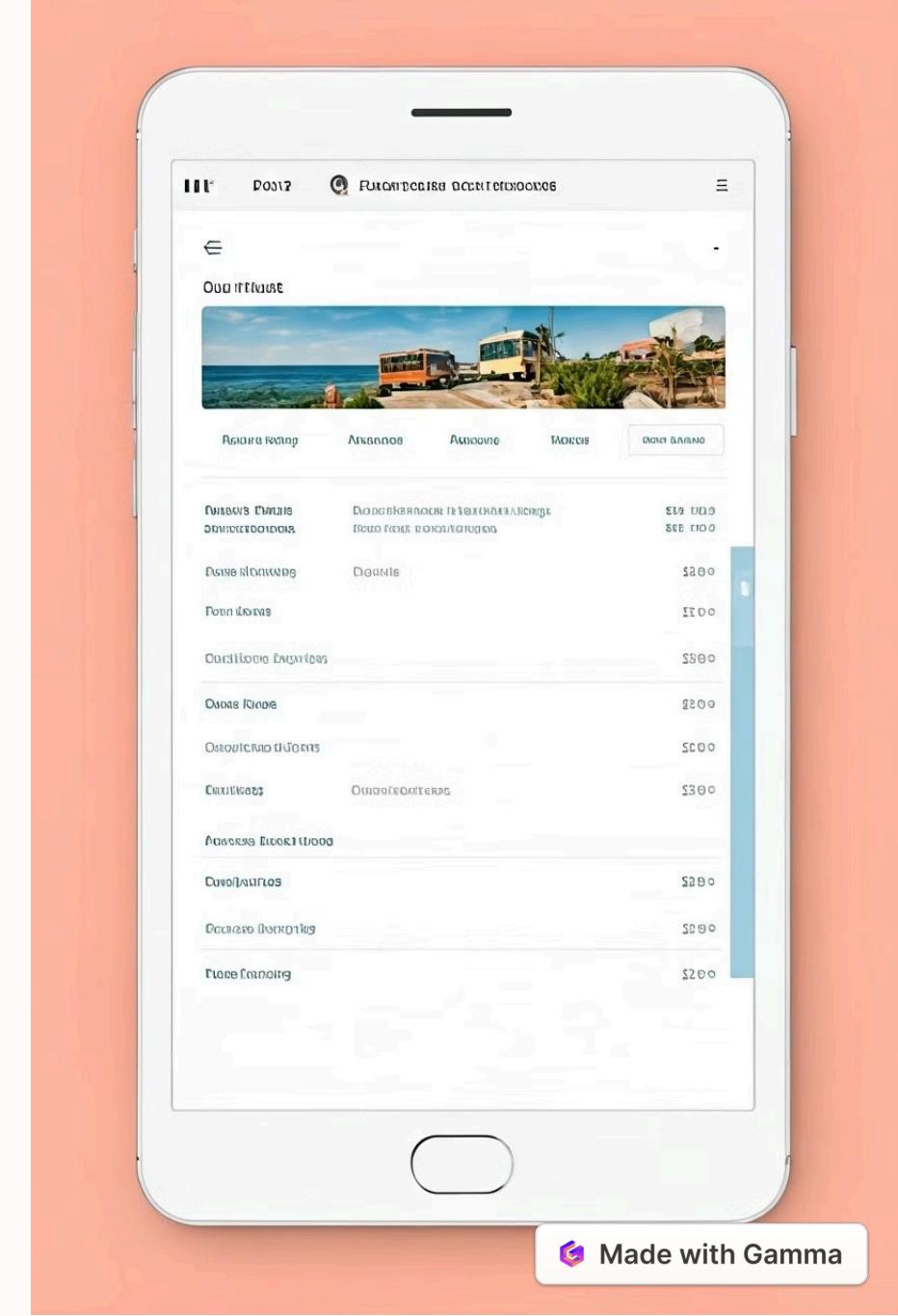
Targetting Web Elements

Selenium locates and interacts with specific web elements like input fields, buttons, and dropdowns.

3

Data Extraction

After searching for buses, the script extracts specific data points, such as route details, schedules, prices, and seat availability.





Store Data in Database

1

Data Transformation

The extracted bus data is transformed into a structured format, such as a Pandas DataFrame.

2

Database Setup

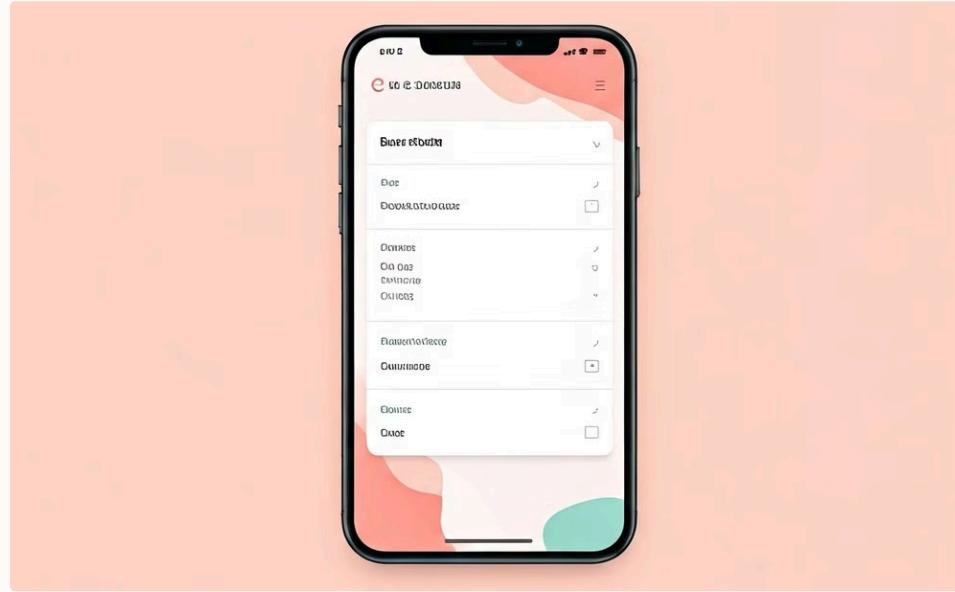
A MySQL database is created and configured to store the bus data effectively.

3

Data Insertion

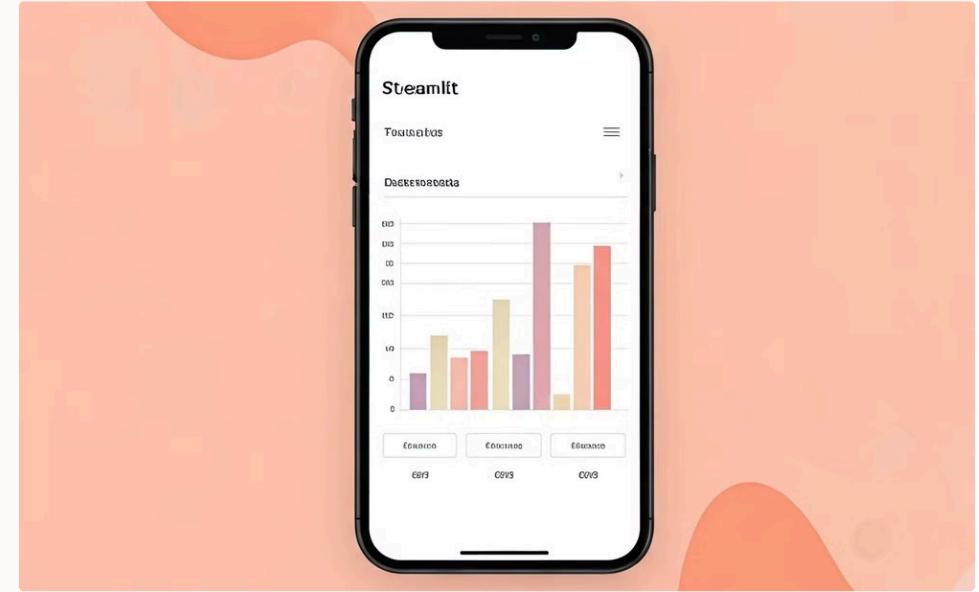
The structured data is inserted into tables within the MySQL database, organizing the information for efficient retrieval.

Web App - Streamlit



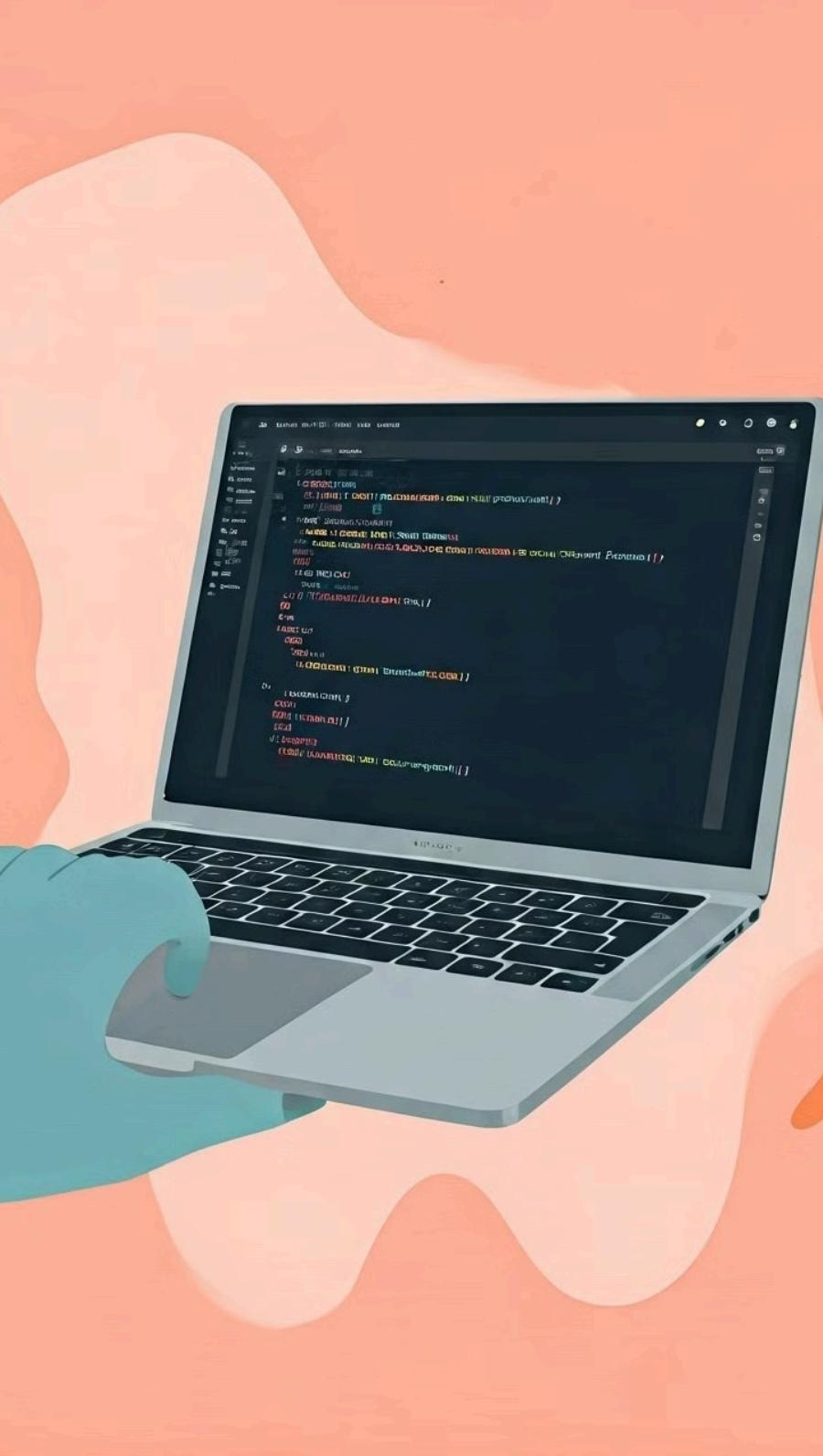
User Interface Design

Streamlit helps build a user-friendly interface that allows users to search for bus routes, view available buses, and get details like departure times and prices.



Data Visualization

The Streamlit application allows for interactive data visualization, enabling users to analyze trends and patterns in the collected bus data.



Packages and Libraries

pandas as pd

For data manipulation and analysis.

mysql.connector

For interacting with MySQL database.

time

For controlling the execution flow and pausing operations.

streamlit asslt

For building the interactive web application.

datetime

For working with dates and times in the data processing.

streamlit_option_menu

For creating a multi-page menu in the Streamlit application.

selenium import webdriver

For controlling the web browser and automating scraping tasks.