# Redbus Data Scraping with Selenium DETAILED DOCUMENTATION

1. Selenium code for web scraping and MySQL database interaction are in the folder

-> “Final Redbus\Web Scrapping & upload to DB\Final\_full\_scraping\_to\_DB.ipynb"

Install packages

! pip install selenium

! pip install pandas

* Created the empty list to store the collected data during the scraping
* Select the states to scrap the data
* Use page navigation for next page
* Loop the states to scrap the data and store it as key values pair in list in bus\_routes = []
* Next loop through the bus\_routes and get link ,routes name, state name
* Loop the link to collect each routes data
* Collect the data in each bus\_routes through the link and store the data in bus\_details = []
* Use pandas to convert the collected data to dataframe to connect to DB

1. MySQL connection

* Install package

! mysql.connector

* Create the DB connection and
* CREATE TABLE to add data in DB
* INSERT THE DATA by looping through the dataframe
* Close all DB connections

1. StreamLit app located in

"Final Redbus\StreamLit\redbus.py"

* Import the packages

import pandas as pd

import streamlit as st

from streamlit\_option\_menu import option\_menu

import mysql.connector as db

* Connect to MySQL DB and convert the data to DataFrame using the pandas
* Used if conditions to filter the data and display using st.write function

1. Sample img dict

"Final Redbus\StreamLit\img"

* States

A screenshot of a computer

Description automatically generated

* Bus Selection

A screenshot of a bus route

Description automatically generated

* Filters

A screenshot of a bus route

Description automatically generatedA screenshot of a bus route

Description automatically generated