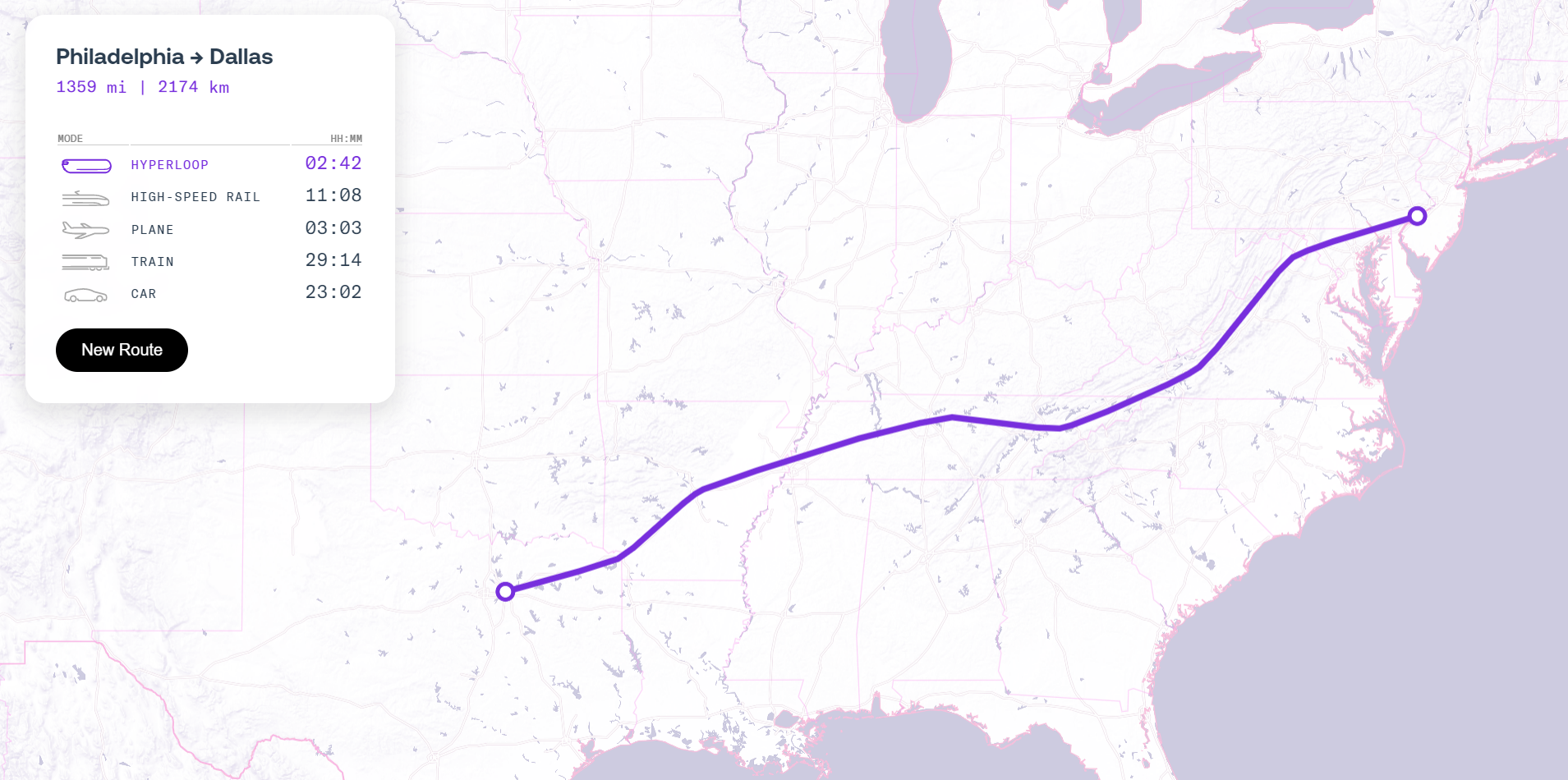
# Hyperloop Data Scraping Tool User Guide

## Robert Zhang, updated in 2022.08.21

Virgin Hyperloop is an American transportation technology company working to commercialize a high-speed travel concept called Hyperloop, designed to transport passengers or cargo at airspeed at a fraction of the cost of air travel. To compare the speed of Hyperloop with other transportation modes, they develop a website (<https://route.virginhyperloop.com/>) to show the travel time between any city pairs in the world (Figure 1).

Figure 1 Hyperloop Website



AECOM developed a data scraping tool to automatically extract the travel time and distance from that website with any input city pairs. This tool has three components. The first part is importing packages and setting up a web driver. The second part is the input data validator to check if the origin and destination cities can be identified correctly by the program. The third part is the main function, which is responsible for scraping the data and storing the outputs in a CSV file. The requirement of input datasets, relevant package and software, and instructions for running the program are listed below:

1. Choose the city pair and create the input file:

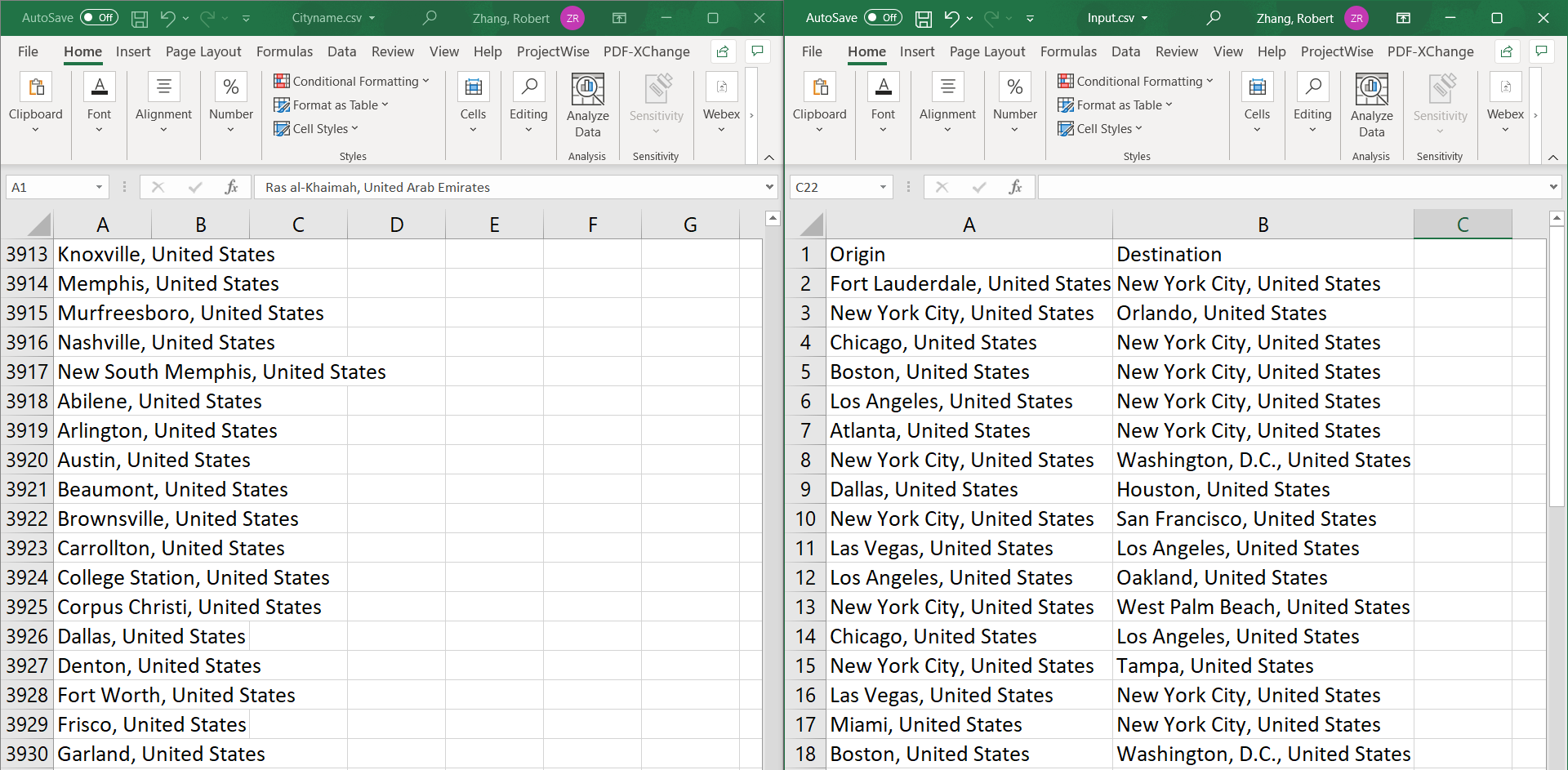
Select the cities from 'cityname.csv' and create the 'Input.csv' to store the city pair for the program. The 'Input.csv' should have two columns with the heading row **'Origin' (in cell A1)** and **'Destination' (in cell A2)**. Most importantly, the city name in the input spreadsheet should be **the same** as the city name spreadsheet. For example, 'New York' should be 'New York City, United States', 'Washington DC' should be 'Washington, D.C., United States', 'Ft. Lauderdale' should be 'Fort Lauderdale, United States', etc. Also, the city pair **should not** be blocked by the ocean because the website does not have the travel time information for city pairs such as New York to London and Tokyo to Los Angeles.

Figure 2 Sample of Cityname.csv (Left) and Input.csv (Right)

1. Have the compiler, Google Chrome, and web driver tool ready for the program:

There are two types of codes in the folder, .py and .ipynb. The functions are the same, and users can choose either Python or Jupyter Notebook to run the program. The tool uses Google Chrome for scraping, and users need to know the browser's version number. Wed driver tools are open-source tools for automated testing of web apps, and Chrome drivers can be downloaded here (<https://chromedriver.chromium.org/downloads>). Remember to replace the correct release for the version of Google Chrome with the 'chromedriver.exe'. (The driver in the folder is version 103 Chrome driver)

1. Run the code:

Before running the code, ensure all files are in the same folder, including chromedriver.exe, Cityname.csv, and Input.csv.

1. Run the .ipynb code using the Jupyter notebook:

Have the Jupyter notebook ready (I highly recommend this Youtube tutorial: <https://www.youtube.com/watch?v=2WL-XTl2QYI&ab_channel=DaveGray>), and then open the Hyperloop. ipynb. Hit run and wait for the web driver to collect all the datasets. The program will show the validation process and which city pair it is exploring. Once "Completed" comes out, the output CSV file is ready.

1. Run the .py code using any compiler:

Run Hyperloop.py and wait for the web driver to collect all the datasets. The program will show the validation process and which city pair it is exploring. Once "Completed" comes out, the output CSV file is ready.