5.2.3 Load and Read the CSV files

With the project scope approved, you're ready to kick off your analysis by making a new Jupyter Notebook file, loading the CSV files, and inspecting them to make sure you don't need to clean the data or change column names before (finally!) merging the two datasets.

Let's get started on the analysis! First, we need to open Jupyter Notebook and create a file for the project, and then we'll import the required libraries and load the data.

Open Jupyter Notebook on macOS

REWIND

On a Mac, to open Jupyter Notebook in the PyBer_Analysis folder:

- 1. On the command line, navigate to the PyBer_Analysis folder and activate the PythonData environment.
- 2. Type jupyter notebook.

Open Jupyter Notebook on Windows

REWIND

In Windows, to open Jupyter Notebook in the PyBer_Analysis folder:

- 1. Open the PythonData Anaconda Prompt for the PythonData environment.
- 2. Type jupyter notebook.

Next, create a new Jupyter Notebook file in your PyBer_Analysis folder and name it PyBer.

Load the CSV files

In the first cell, add the following code to import the Pandas and Matplotlib libraries with the Pyplot module, and run the cell:

```
# Add Matplotlib inline magic command
%matplotlib inline
# Dependencies and Setup
import matplotlib.pyplot as plt
import pandas as pd
```

In a new cell, declare variables that connect to the CSV files in the Resources folder:

```
# Files to load
city_data_to_load = "Resources/city_data.csv"
ride_data_to_load = "Resources/ride_data.csv"
```

REWIND

If you want to use <code>os.path.join()</code> to load CSV files, you need to import the <code>os</code> module with your dependencies, like this:

import os

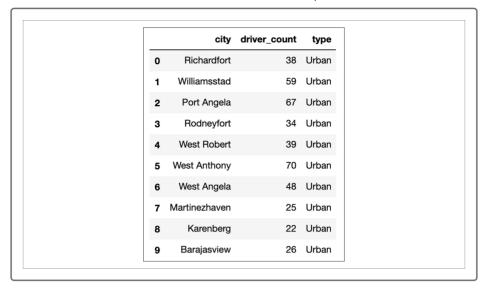
Next, we will read each CSV file in Pandas.

Read the City Data File

To read a CSV file into Pandas, we use pd.read_csv. Add the following code to a new cell:

```
# Read the city data file and store it in a pandas DataFrame.
city_data_df = pd.read_csv(city_data_to_load)
city_data_df.head(10)
```

When you run the cell, the first 10 rows of your city data should look something like this:

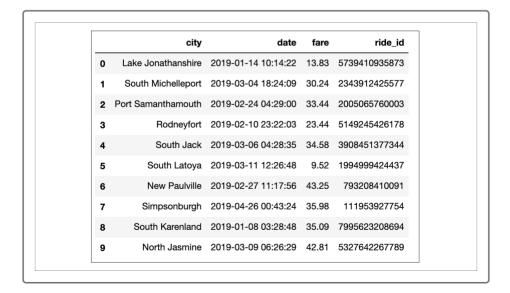


Read the Ride Data File

To load the ride_data.csv file into a Pandas DataFrame, add the following code to a new cell:

```
# Read the ride data file and store it in a pandas DataFrame.
ride_data_df = pd.read_csv(ride_data_to_load)
ride_data_df.head(10)
```

When you run the cell, the first 10 rows of the ride data should look something like this:



© 2020 - 2022 Trilogy Education Services, a 2U, Inc. brand. All Rights Reserved.