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 `os.path.join()` to load CSV files, you need to import the `os` 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:

	city	driver_count	type
0	Richardfort	38	Urban
1	Williamsstad	59	Urban
2	Port Angela	67	Urban
3	Rodneyfort	34	Urban
4	West Robert	39	Urban
5	West Anthony	70	Urban
6	West Angela	48	Urban
7	Martinezhaven	25	Urban
8	Karenberg	22	Urban
9	Barajasview	26	Urban

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:

	city	date	fare	ride_id
0	Lake Jonathanshire	2019-01-14 10:14:22	13.83	5739410935873
1	South Michelleport	2019-03-04 18:24:09	30.24	2343912425577
2	Port Samanthamouth	2019-02-24 04:29:00	33.44	2005065760003
3	Rodneyfort	2019-02-10 23:22:03	23.44	5149245426178
4	South Jack	2019-03-06 04:28:35	34.58	3908451377344
5	South Latoya	2019-03-11 12:26:48	9.52	1994999424437
6	New Paulville	2019-02-27 11:17:56	43.25	793208410091
7	Simpsonburgh	2019-04-26 00:43:24	35.98	111953927754
8	South Karenland	2019-01-08 03:28:48	35.09	7995623208694
9	North Jasmine	2019-03-09 06:26:29	42.81	5327642267789