4.4.3 Load and Read CSV Files

One of the first things you will need to do to get started is to load the datasets into a Jupyter Notebook file. You recall how to do this using Python, but you're not sure how to load datasets using Jupyter Notebook. Say hello to your good friend Google! You're going to research how to import a CSV file into Jupyter Notebook. Maria will be available to assist when you need it.

So far, we've practiced using Jupyter Notebook as well as the Pandas library. Now we'll put these skills together to perform our analysis of school and student data.

To get started, activate the PythonData environment. If you're using a Mac, use the command line to navigate to the "School_District_Analysis" folder and activate the PythonData environment. If you're on a Windows machine, open the PythonData Anaconda Prompt for the PythonData environment and navigate to the "School_District_Analysis" folder.

Create a new Jupyter Notebook file in the "School_District_Analysis" folder and rename it `PyCitySchools.ipynb`.

Load the CSV Files

In the first cell of your `PyCitySchools.ipynb` file, import the Pandas library as the dependency and run the cell.

```
# Add the Pandas dependency.
import pandas as pd
```

In the next cell, declare two variables: one assigned to the `schools_complete.csv` file and one assigned to the `students_complete.csv` file. (These files are located in the Resources folder.) Your code should look like this:

```
# Files to load
school_data_to_load = "Resources/schools_complete.csv"
student_data_to_load = "Resources/students_complete.csv"
```

Alternatively, you can use the indirect path method to access the `schools_complete.csv` and `students_complete.csv` files.

REWIND

When we want to get the indirect path to a file, we use `os.path.join()` to load a file from somewhere in our directory.

If you decide to use this approach, you will need to import the `os` module with your Pandas dependency using the following code:

```
# Add the dependencies.
import pandas as pd
import os
```

Then, use the `os.path.join()` method to connect to the CSV files:

```
# Files to load
school_data_to_load = os.path.join("Resources", "schools_complete.csv")
student_data_to_load = os.path.join("Resources", "students_complete.csv")
```

Read the School Data File

Now we'll read each CSV file with the Pandas function `read_csv()`. Inside this function, we'll add the file we want to read, which is one of many parameters that we can add to this function.

NOTE

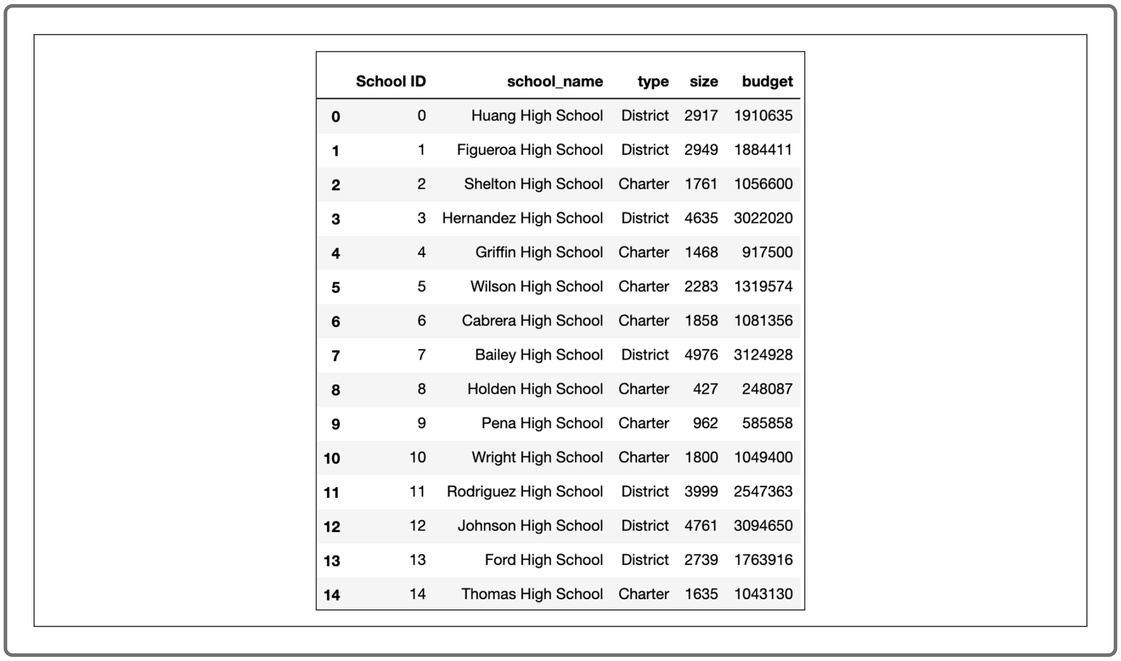
> For more information, see the **Pandas documentation on the read_csv() function (https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.read_csv.html)**.

Add the following code to a new cell and run the cell. This will allow us to read `schools_complete.csv` and store it in a DataFrame.

```
# Read the school data file and store it in a Pandas DataFrame.
school_data_df = pd.read_csv(school_data_to_load)
school_data_df
```

Previously, we used the `pd.DataFrame()` function to convert a list of dictionaries, as we have in these CSV files. The `read_csv()` function makes it easier for us by converting the CSV file to a DataFrame.

Your results should look something like this:



	School ID	school_name	type	size	budget
0	0	Huang High School	District	2917	1910635
1	1	Figueroa High School	District	2949	1884411
2	2	Shelton High School	Charter	1761	1056600
3	3	Hernandez High School	District	4635	3022020
4	4	Griffin High School	Charter	1468	917500
5	5	Wilson High School	Charter	2283	1319574
6	6	Cabrera High School	Charter	1858	1081356
7	7	Bailey High School	District	4976	3124928
8	8	Holden High School	Charter	427	248087
9	9	Pena High School	Charter	962	585858
10	10	Wright High School	Charter	1800	1049400
11	11	Rodriguez High School	District	3999	2547363
12	12	Johnson High School	District	4761	3094650
13	13	Ford High School	District	2739	1763916
14	14	Thomas High School	Charter	1635	1043130

CAUTION

> If you see the error `FileNotFoundError` in your output, this means that the CSV file was not found in the Resources subfolder inside the School_District_Analysis folder.
>
> To fix this error, add the CSV file to the Resources subfolder. Make sure the Resources subfolder is located in the School_District_Analysis folder, or you can use the indirect path approach with `os.path.join()` method.

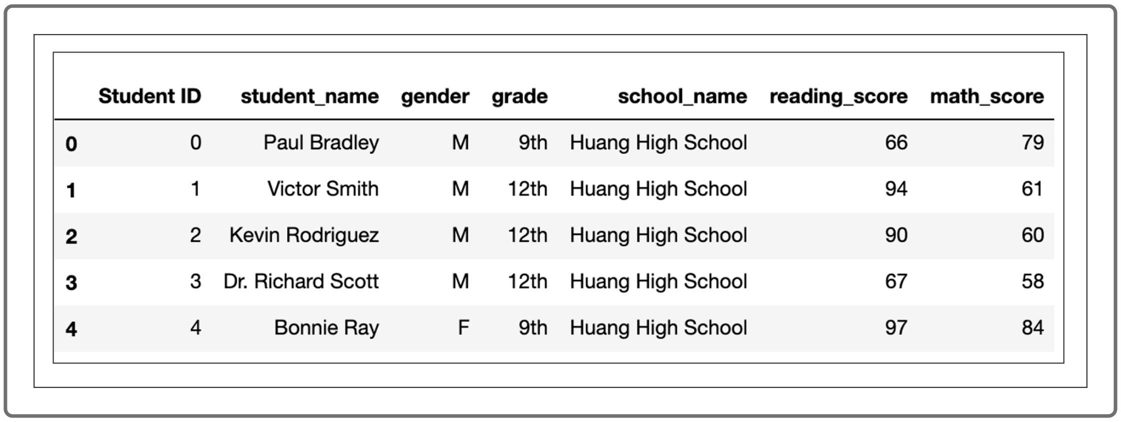
SHOW PRO TIP

Read the Student Data File

Now we'll read the student data file and store it in a Pandas DataFrame by adding the following code to a new cell:

```
# Read the student data file and store it in a Pandas DataFrame.
student_data_df = pd.read_csv(student_data_to_load)
student_data_df.head()
```

After running the code, the output should look like this:



	Student ID	student_name	gender	grade	school_name	reading_score	math_score
0	0	Paul Bradley	M	9th	Huang High School	66	79
1	1	Victor Smith	M	12th	Huang High School	94	61
2	2	Kevin Rodriguez	M	12th	Huang High School	90	60
3	3	Dr. Richard Scott	M	12th	Huang High School	67	58
4	4	Bonnie Ray	F	9th	Huang High School	97	84

You have now loaded and read the CSV files—nice work! Now is a good time to save your work in your GitHub repository.