4.4.1 Import and Inspect CSV Files

Now that you have learned about the Pandas library and what it can do, Maria is going to share the datasets with you, which are in CSV format. After you import these CSV files, you will need to inspect them. Then you'll be given a set of tasks to complete for the analysis.

Import the Data

Create a folder named "Resources" in your "School_District_Analysis" folder. Click the following links to download the datasets into the Resources folder.

<u>Download schools_complete.csv (https://2u-data-curriculum-team.s3.amazonaws.com/dataviz-online/module_4/schools_complete.csv)</u>

<u>Download students_complete.csv (https://2u-data-curriculum-team.s3.amazonaws.com/dataviz-online/module_4/students_complete.csv)</u>

You should now have two CSV files in your Resources folder:

- (schools_complete.csv)
- students_complete.csv

Inspect the Data

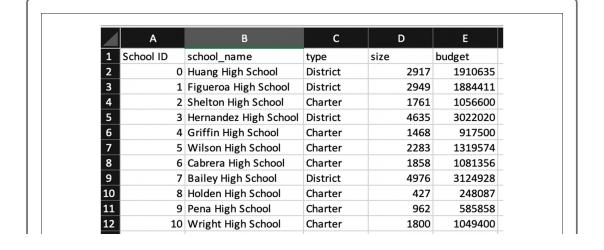
Now that the CSV files downloaded and imported into the correct folder, let's inspect the data in the files.

REWIND

Remember, when inspecting the data, focus on the following questions:

- How many columns and rows are there?
- What types of data are present?
- Is the data readable, or does it need to be converted in some way?

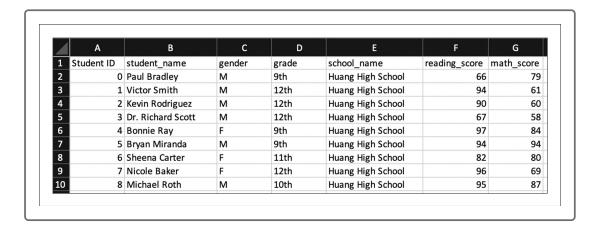
When we open schools_complete.csv, we can see that there is a top header row followed by 15 rows of data, labeled with the names "School ID," "school_name," "type," "size," and "budget." See the following image for reference.



	11 Rodriguez High School 12 Johnson High School	District District	3999 4761	2547363 3094650
	L3 Ford High School	District	2739	1763916
16	L4 Thomas High School	Charter	1635	1043130

Take a few moments to read through the data. There are no misspellings, changes of case, duplicates, or special characters; the dataset looks clean! This means that this dataset will not require any cleaning once loaded into the DataFrame, which is convenient.

Next, open <u>students_complete.csv</u>. This is a large file! Here are the first 10 rows:



Now go to the last row of the file.

REWIND

To get to the last row of an Excel file, place the cursor in a column that doesn't have any empty cells and press Command + the down-arrow key (on a Mac) or CTRL + the down-arrow key (on Windows).

39167	39165 Donna Howard	F	12th	Thomas High School	99	90
39168	39166 Dawn Bell	F	10th	Thomas High School	95	70
39169	39167 Rebecca Tanne	r F	9th	Thomas High School	73	84
39170	39168 Desiree Kidd	F	10th	Thomas High School	99	90
39171	39169 Carolyn Jacksor	F	11th	Thomas High School	95	75

When we get to the last row, we see there are 39,170 rows (the first row is the header row, so it's not counted). Also, each row contains a student ID that is associated with a student's name, gender, grade in school, the name of the school they attend, and their reading and math scores. Students are grouped by school. Like schools_complete.csv, this file also includes the "school_name" in the header.

C Retake		

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