

## 4.2.5

## Activity: Heads Up

---

**Maria** has a file of student records from all the schools in the district. The school board wants to get some analysis done on the file, and she's brought the data to you so that you can use Python and Pandas to do it! Eventually, she'll ask for detailed information about specific subsets of the data. But, she's not yet sure what questions the school board will have. So, Maria asks you to collect the data for now as a first step.

## Files

Download the following files to help you get started:

**Heads Up files** ([https://2u-data-curriculum-team.s3.amazonaws.com/dataviz-online/v2/module\\_4/4-2-Student\\_Data\\_Starter\\_Code.zip](https://2u-data-curriculum-team.s3.amazonaws.com/dataviz-online/v2/module_4/4-2-Student_Data_Starter_Code.zip))

Using the Jupyter Notebook file in the **Unsolved** folder, navigate to the "Step 1" section to write your code. Alternatively, you can use the Jupyter Notebook file that you created in the previous lesson to write your code.

## Instructions

1. In Jupyter Notebook, open the **Student\_Data\_Analysis.ipynb** file, and use it to write your code. Note that the **Resources** folder contains the CSV file that you'll import.
2. Import the **pandas** and **os** libraries, as the following code shows:

```
import pandas as pd
import os
```

3. Using the Pandas **read\_csv** function and the **os.path.join** function, import the data from the **new\_student\_data.csv** file into a DataFrame named **student\_df**, as the following code shows:

```
student_data = os.path.join('../Resources/new_student_data.csv')  
student_df = pd.read_csv(student_data)
```

4. Use the `head` function (and/or the `tail` function) to check that Pandas correctly imported the data, as the following code shows:

```
student_df.head()
```

## Solution

How did you do?

You can refer to the solution file in the `Solved` folder, which is in the zipped folder that you downloaded for this activity.

## What's Next?

You've learned how to use two powerful tools—Pandas and Jupyter Notebook—to collect data. That makes up the first phase of our three-phase data analysis process.

In this activity, you might have noticed some duplicated or missing values. Next, you'll learn how to use Pandas to address those issues and more. Doing so is part of the second phase of the data analysis process: preparing the data.

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