Module 4 Challenge

Blueprint

Background

The school board has notified Maria and her supervisor that the students\_complete.csv file shows evidence of academic dishonesty; specifically, reading and math grades for Thomas High School ninth graders appear to have been altered. Although the school board does not know the full extent of the academic dishonesty, they want to uphold state-testing standards and have turned to Maria for help. She has asked you to replace the math and reading scores for Thomas High School with NaNs while keeping the rest of the data intact. Once you’ve replaced the math and reading scores, Maria would like you to repeat the school district analysis that you did in this module and write up a report to describe how these changes affected the overall analysis.

What You're Creating

This new assignment consists of two technical analysis deliverables and a written report to present your results. You will submit the following:

Deliverable 1: Replace ninth-grade reading and math scores

Deliverable 2: Repeat the school district analysis

Deliverable 3: A written report for the school district analysis (README.md)

Files

Use the following link to download the Challenge starter code:

Download challenge starter code (Links to an external site.)

Before You Start

Before you get started, follow these steps:

Make a copy of your PyCitySchools.ipynb file and rename it PyCitySchools\_Challenge\_testing.ipynb.

Download the PyCitySchools\_Challenge\_starter\_code.ipynb file, copy the code, and paste it at the top of your PyCitySchools\_Challenge\_testing.ipynb file.

You’ll use this file to test your code as you work through the challenge.

Once your code is working, you'll make a copy of the PyCitySchools\_Challenge\_testing.ipynb file and rename it PyCitySchools\_Challenge.ipynb.

When you're ready to submit, be sure to check that all DataFrames created for Deliverables 1 and 2 are visible in your outputs. Do not include any unnecessary print statements in your code.

Deliverable 1: Replace Ninth-Grade Reading and Math Scores (50 points)

Deliverable 1 Instructions

Using the Pandas loc method with conditional statements and comparison and logical operators, select the ninth-grade reading and math scores for Thomas High School. Then, use the Pandas NumPy module to change the reading and math scores to NaN.

REWIND

For this deliverable, you’ve already done the following in this module:

Lesson 4.4.3: Open and read CSV files.

Lesson 4.5.6: Replace substrings to fix the students’ names.

Lesson 4.7.6: Use logical operators to select specific data from a DataFrame column.

Use the instructions below to add code where indicated by the numbered-step comments in the starter code file.

IMPORTANT

Before you get started, open up your command line and use either of the following commands to install the NumPy module:

conda install numpy or pip install numpy

Use the code snippet provided in Step 1 to import the NumPy module: import numpy as np.

Use the code snippet provided in Step 2 for the Pandas loc method.

If you’d like a hint on using the loc method, that’s totally okay. If not, that’s great too. You can always revisit this later if you change your mind.

HINT

To select all the ninth-grade reading scores at Thomas High School, use the following steps to write code inside the brackets of the loc method:

a) Add an opening parenthesis, then use a comparison operator to retrieve all the rows with Thomas High School from the "school\_name" column of the student\_data\_df, then close the parenthesis.

b) Add a logical operator then another opening parenthesis, then use a comparison operator to retrieve all the rows with ninth grade from the "grade" column of the student\_data\_df, then close the parenthesis.

c) To change the reading scores only, add a comma after the last closing parenthesis then add the "reading\_score" column.

d) Outside of the closing brackets of the loc method, set the ninth-grade reading scores from Thomas High School equal to np.nan.

NOTE

If your student\_data\_df looks like the image below, you have not completed 3c above. In the image below, all the ninth-grade student data for Thomas High School was replaced with NaN.

The student data DataFrame with all the ninth grade student data for Thomas High School replaced with NaN

In Step 3, refactor the code from Step 2 to replace the math scores with NaNs.

In Step 4, check the student data to make sure the grades were replaced with NaNs.

After you run Step 4 in your PyCitySchools\_Challenge\_testing.ipynb file, confirm that the DataFrame looks like the image below, where the ninth-grade reading and math scores from Thomas High School have been replaced with NaNs. Then, make a copy of the PyCitySchools\_Challenge\_testing.ipynb file and rename it PyCitySchools\_Challenge.ipynb.

The student data DataFrame after replacing the ninth

grade math and reading scores for Thomas High School with NaN

Deliverable 1 Requirements

You will earn a perfect score for Deliverable 1 by completing all requirements below:

The loc method is used to select all the reading and math scores from the ninth grade at Thomas High School. Inside the loc method, the following are completed:

A comparison operator is used to retrieve all the rows with Thomas High School in the "school\_name" column of the student\_data\_df (10 pt).

A comparison operator is used to retrieve all the rows with the ninth grade in the "grade" column of the student\_data\_df (10 pt).

Logical and comparison operators are used to retrieve all the rows with the "reading\_score" column for Thomas High School ninth graders from the student\_data\_df (10 pt).

Logical and comparison operators are used to retrieve all the rows with the "math\_score" column for Thomas High School ninth graders from the student\_data\_df (10 pt).

The reading and math scores for the ninth graders in Thomas High school are replaced with NaNs (10 pt).

Deliverable 2: Repeat the School District Analysis (25 points)

Deliverable 2 Instructions

Repeat the school district analysis you did in this module, and recreate the following metrics:

The district summary

The school summary

The top 5 and bottom 5 performing schools, based on the overall passing rate

The average math score for each grade level from each school

The average reading score for each grade level from each school

The scores by school spending per student, by school size, and by school type

In Steps 1-4, you’ll update the district summary. For this task, you’ll recalculate the total student count by subtracting the number of ninth-grade students in Thomas High School from the total student count, then you'll recalculate the passing math and passing reading percentages, and the overall passing percentage with the recalculated total student count.

In Steps 5-14, you’ll execute the code from this module that creates and formats the School Summary DataFrame, then update the school summary using the 10th-12th graders from Thomas High School as follows:

First, you’ll calculate the number of 10th-12th graders in Thomas High School.

Create three new DataFrames for the 10th-12th graders from Thomas High School: students who passed math, students who passed reading, and students who passed both math and reading.

Using these DataFrames, you'll recalculate the percentage of students who passed math, passed reading, and passed both math and reading for Thomas High School only.

Finally, you'll replace the % Passing Math, % Passing Reading, and % Overall Passing scores in the current School Summary DataFrame with the new passing percentages for Thomas High School.

REWIND

For this deliverable, you’ve already completed the school district analysis in this module:

Lessons 7-13

Use the instructions below to add code where indicated by the numbered-step comments in the starter code file to update the District Summary DataFrame.

In Step 1, using the loc method with logical and comparison operators, retrieve the student count for Thomas High School ninth graders in the school\_data\_complete\_df DataFrame.

In Step 2, subtract the number of students retrieved from Step 1 from the total student count to get the new total student count.

In Step 3, calculate the math and reading passing percentages based on the new total student count.

In Step 4, calculate the overall passing percentage with the new total student count.

Before moving on, confirm that that your District Summary DataFrame looks like this image:

The updated district summary DataFrame

Use the instructions below to add code where indicated by the numbered-step comments in the starter code file to update the School Summary DataFrame.

Run the code from this module that creates and formats the School Summary DataFrame.

Before moving on, confirm that the metrics for Thomas High School look like this image.

Thomas High School’s metrics in the School Summary DataFrame

In Step 5, get the number of 10th-12th grade students from Thomas High School.

In Step 6, use the loc method to create a new DataFrame that has all the students passing math from Thomas High School.

In Step 7, use the loc method to create a new DataFrame that has all the students passing reading from Thomas High School.

In Step 8, use the loc method to create a new DataFrame that has all the students passing math and reading from Thomas High School.

In Step 9, calculate the percentage of 10th-12th grade students passing math from Thomas High School.

In Step 10, calculate the percentage of 10th-12th grade students passing reading from Thomas High School.

In Step 11, calculate the overall passing percentage of 10th-12th grade students from Thomas High School.

In Step 12, use the loc method to replace the % Passing Math score for Thomas High School with the new math passing percentage you calculated in Step 9.

In Step 13, use the loc method to replace the % Passing Reading score for Thomas High School with the new reading passing percentage you calculated in Step 10.

In Step 14, use the loc method to replace the % Overall Passing score for Thomas High School with the new overall passing percentage you calculated in Step 11.

If you’d like a hint on using the loc method to select an index and column, that’s totally okay. If not, that’s great too. You can always revisit this later if you change your mind.

HINT

Before moving on, confirm that the updated metrics for Thomas High School look like this image:

Thomas High School’s updated metrics in the School Summary DataFrame

Next, complete the following steps for school district analysis using the remaining steps that are provided in the starter code.

The top 5 and bottom 5 performing schools, based on the overall passing rate

The average math score for each grade level from each school

The average reading score for each grade level from each school

The scores by school spending per student, by school size, and by school type

Deliverable 2 Requirements

You will earn a perfect score for Deliverable 2 by repeating the school district analysis and updating the following required metrics in the PyCitySchools\_Challenge.ipynb file:

The district summary DataFrame (3 pt)

The school summary DataFrame (3 pt)

The top 5 performing schools, based on the overall passing rate (2 pt)

The bottom 5 performing schools, based on the overall passing rate (2 pt)

The average math score for each grade level from each school (3 pt)

The average reading score for each grade level from each school (3 pt)

The scores by school spending per student (3 pt)

The scores by school size (3 pt)

The scores by school type (3 pt)

Deliverable 3: A Written Report for the School District Analysis (25 points)

Deliverable 3 Instructions

For this part of the Challenge, write a report that summarizes your updated analysis and compares it with the results from the module.

The analysis should contain the following:

Overview of the school district analysis: Explain the purpose of this analysis.

Results: Using bulleted lists and images of DataFrames as support, address the following questions.

How is the district summary affected?

How is the school summary affected?

How does replacing the ninth graders’ math and reading scores affect Thomas High School’s performance relative to the other schools?

How does replacing the ninth-grade scores affect the following:

Math and reading scores by grade

Scores by school spending

Scores by school size

Scores by school type

Summary: Summarize four changes in the updated school district analysis after reading and math scores for the ninth grade at Thomas High School have been replaced with NaNs.

Deliverable 3 Requirements

Structure, Organization, and Formatting (7 points)

The written analysis has the following structure, organization, and formatting:

There is a title, and there are multiple sections (2 pt).

Each section has a heading and subheading (3 pt).

Links to images are working, and code is formatted and displayed correctly (2 pt).

Analysis (18 points)

The written analysis has the following:

Overview of the school district analysis:

The purpose of this analysis is well defined (3 pt).

Results:

There is a bulleted list that addresses how each of the seven school district metrics was affected by the changes in the data (10 pt).

Summary:

There is a statement summarizing four changes to the school district analysis after reading and math scores have been replaced (5 pt).

Submission

Once you’re ready to submit, make sure to check your work against the rubric to ensure you are meeting the requirements for this Challenge one final time. It’s easy to overlook items when you’re in the zone!

As a reminder, the deliverables for this Challenge are as follows:

Deliverable 1: Replace ninth-grade reading and math scores

Deliverable 2: Repeat the school district analysis

Deliverable 3: A written report for the school district analysis (README.md)

Upload the following to your School\_District\_Analysis GitHub repository:

The PyCitySchools\_Challenge.ipynb file.

The Resources folder with the schools\_complete.csv and students\_complete.csv files.

An updated README.md that has your written analysis.

To submit your challenge assignment for grading in Bootcamp Spot, click Start Assignment, click the Website URL tab, then provide the URL of your School\_District\_Analysis GitHub repository, and then click Submit. Comments are disabled for graded submissions in BootCampSpot. If you have questions about your feedback, please notify your instructional staff or the Student Success Manager. If you would like to resubmit your work for an improved grade, you can use the Re-Submit Assignment button to upload new links. You may resubmit up to 3 times for a total of 4 submissions.

IMPORTANT

Once you receive feedback on your Challenge, make any suggested updates or adjustments to your work. Then, add this week’s Challenge to your professional portfolio.

NOTE

You are allowed to miss up to two Challenge assignments and still earn your certificate. If you complete all Challenge assignments, your lowest two grades will be dropped. If you wish to skip this assignment, click Next, and move on to the next Module.