Module 4 Challenge

New Attempt

Due Jan 31 by 12:59am **Points** 100 **Submitting** a text entry box or a website url

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 (https://2u-data-curriculum-team.s3.amazonaws.com/dataviz-online/module_4/PyCitySchools_Challenge_starter_code.ipynb)

Before You Start

Before you get started, follow these steps:

1. Make a copy of your `PyCitySchools.ipynb` file and rename it `PyCitySchools_Challenge_testing.ipynb`.

2. 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.

3. Once your code is working, you'll make a copy of the `PyCitySchools_Challenge_testing.ipynb` file and rename it `PyCitySchools_Challenge.ipynb`.
4. 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

> 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`

1. Use the code snippet provided in Step 1 to import the NumPy module: `import numpy as np`.
2. 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.

SHOW HINT

3. 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.

	Student ID	student_name	gender	grade	school_name	reading_score	math_score
39160	39160.0	Katie Weaver	F	11th	Thomas High School	89.0	86.0
39161	39161.0	April Reyes	F	10th	Thomas High School	70.0	84.0
39162	39162.0	Derek Weeks	M	12th	Thomas High School	94.0	77.0
39163	39163.0	John Reese	M	11th	Thomas High School	90.0	75.0
39164	NaN	NaN	NaN	NaN	NaN	NaN	NaN
39165	39165.0	Donna Howard	F	12th	Thomas High School	99.0	90.0
39166	39166.0	Dawn Bell	F	10th	Thomas High School	95.0	70.0
39167	NaN	NaN	NaN	NaN	NaN	NaN	NaN
39168	39168.0	Desiree Kidd	F	10th	Thomas High School	99.0	90.0
39169	39169.0	Carolyn Jackson	F	11th	Thomas High School	95.0	75.0

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

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

6. 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`.

```
#  Check the student data for NaN's.
student_data_df.tail(10)
```

	Student ID	student_name	gender	grade	school_name	reading_score	math_score
39160	39160	Katie Weaver	F	11th	Thomas High School	89.0	86.0
39161	39161	April Reyes	F	10th	Thomas High School	70.0	84.0
39162	39162	Derek Weeks	M	12th	Thomas High School	94.0	77.0
39163	39163	John Reese	M	11th	Thomas High School	90.0	75.0
39164	39164	Joseph Anthony	M	9th	Thomas High School	NaN	NaN
39165	39165	Donna Howard	F	12th	Thomas High School	99.0	90.0
39166	39166	Dawn Bell	F	10th	Thomas High School	95.0	70.0
39167	39167	Rebecca Tanner	F	9th	Thomas High School	NaN	NaN
39168	39168	Desiree Kidd	F	10th	Thomas High School	99.0	90.0
39169	39169	Carolyn Jackson	F	11th	Thomas High School	95.0	75.0

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.

1. 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.
2. In Step 2, subtract the number of students retrieved from Step 1 from the total student count to get the new total student count.
3. In Step 3, calculate the math and reading passing percentages based on the new total student count.
4. 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:

	Total Schools	Total Students	Total Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing
0	15	39,170	$24,649,428.00	78.9	81.9	74.8	85.7	64.9

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.

5. 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.

	School Type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing
Thomas High School	Charter	1635	$1,043,130.00	$638.00	83.350937	83.896082	66.911315	69.663609	65.076453

6. In Step 5, get the number of 10th-12th grade students from Thomas High School.
7. In Step 6, use the `loc` method to create a new DataFrame that has all the students passing math from Thomas High School.
8. In Step 7, use the `loc` method to create a new DataFrame that has all the students passing reading from Thomas High School.
9. 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.
10. In Step 9, calculate the percentage of 10th-12th grade students passing math from Thomas High School.
11. In Step 10, calculate the percentage of 10th-12th grade students passing reading from Thomas High School.
12. In Step 11, calculate the overall passing percentage of 10th-12th grade students from Thomas High School.

13. 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.
14. 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

percentage you calculated in Step 10.

15. 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.

SHOW HINT

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

	School Type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing
Thomas High School	Charter	1635	$1,043,130.00	$638.00	83.350937	83.896082	93.18569	97.018739	90.630324

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:

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

2. **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

3. **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

- 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)

(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 in Canvas, click Submit, then provide the URL of your School_District_Analysis GitHub repository for grading. 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.

Module-4 Rubric

Criteria	Ratings					Pts
Deliverable 1: Replace Ninth Grade Reading and Math Scores	**50 to >47.0 pts Demonstrating Proficiency** The Deliverable Fulfills "Approaching Proficiency" Required Criteria and meets this requirement: ✓The reading and math scores are replaced with NaN.	**47 to >42.0 pts Approaching Proficiency** The Deliverable Fulfills "Developing Proficiency" Required Criteria and meets these requirements: PLUS ✓ Logical AND comparison operators are used to retrieve all the rows with the math scores forThomas High School ninth graders. AND does this: ✓ Either the reading OR math scores are replaced with NaN.	**42 to >36.0 pts Developing Proficiency** The Deliverable Fulfills "Emerging" Required Criteria and meets these requirements: ✓ A comparison operator is used to retrieve all the rows with ninth grade in the "grade" column ✓ Logical AND comparison operators are used to retrieve all the rows with the reading scores for Thomas High School ninth graders. AND does these: ✓ Logical AND comparison operators are used to retrieve the math scores from ALL the grades at Thomas High School. ✓ There is an attempt to replace reading and/or math scores with NaN, OR all the rows from Thomas High School are	**36 to >0.0 pts Emerging** REQUIRED: The Deliverable does the following: ✓A comparison operator is used to retrieve all the rows with Thomas High School in the "school_name" column. AND does these: ✓A comparison operator is used to retrieve all the rows from the "grade" column. ✓Logical AND comparison operators are used to retrieve the reading scores from all grades of Thomas High School. ✓Logical AND comparison operators are used to retrieve the math scores from all grades of Thomas High School. ✓There is an attempt to replace reading and math scores with NaN, OR all the	**0 pts Incomplete**	50 pts

Criteria	Ratings					Pts
Deliverable 2: Repeat the School District Analysis	**25 to >24.0 pts** **Demonstrating Proficiency** The reading and math scores are replaced with NaN and all the following are completed with no errors: ✓ There is a new district summary DataFrame. ✓There is a new school summary DataFrame. ✓ The bottom 5 performing schools are shown. ✓ The average math scores for each grade level are shown. ✓ The average math scores for each grade level are shown. ✓ The average reading scores for each grade level are shown. ✓ The scores by school spending per student is shown. ✓ The scores by school size is shown. ✓ The scores by school type are shown.	**24 to >23.0 pts** **Approaching Proficiency** The reading and math scores are replaced with NaN and all the following are completed with some errors: ✓ There is a new district summary DataFrame. ✓There is a new school summary DataFrame. ✓ The bottom 5 performing schools are shown. ✓ The average math scores for each grade level are shown. ✓ The average math scores for each grade level are shown. ✓ The average reading scores for each grade level are shown. ✓ The scores by school spending per student is shown. ✓ The scores by school size is shown. ✓ The scores by school type are shown.	**23 to >20.0 pts** **Developing Proficiency** Either the reading OR math scores are replaced with NaN and all the following are completed: ✓ There is a new district summary DataFrame. ✓There is a new school summary DataFrame. ✓ The bottom 5 performing schools are shown. ✓ The average math scores for each grade level are shown. ✓ The average math scores for each grade level are shown. ✓ The average reading scores for each grade level are shown. ✓ The scores by school spending per student is shown. ✓ The scores by school size is shown. ✓ The scores by school type are shown.	**20 to >0.0 pts** **Emerging** The reading and math scores are not replaced with NaN but all the following are completed: ✓ There is a district summary DataFrame. ✓ There is a new school summary DataFrame. ✓ The top 5 performing schools are shown. ✓ The bottom 5 performing schools are shown. ✓ The average math scores for each grade level are shown. ✓ The average reading scores for each grade level are shown. ✓ The scores by school spending per student is shown. ✓ The scores by school size is shown. ✓ The scores by school type are shown.	**0 pts** **Incomplete**	25 pts

Criteria	Ratings					Pts
Deliverable 3: Structure, Organization, and Formatting	**7 to >6.0 pts Demonstrating Proficiency** The written analysis has ALL of the following: ✓ There is a title, and there are multiple sections. ✓ Each section has a heading and subheading. ✓ There are images and references to code, and they are formatted and displayed correctly.	**6 to >4.0 pts Approaching Proficiency** The written analysis has ALL of the following: ✓ There is a title, and there are multiple sections. ✓ Each section has a heading and subheading. ✓ There are images and references to code, and they are formatted and displayed correctly with one or two minor errors.	**4 to >3.0 pts Developing Proficiency** The written analysis has ALL of the following: ✓ There is a title, and there are multiple sections. AND ONE of the following: ✓ Each section may have a heading and subheading. ✓ There are images and references to code, and they are formatted and displayed correctly with one or two minor errors.	**3 to >0.0 pts Emerging** The written analysis has ALL of the following: ✓ There is a title. ✓ There may be a subheading for a section. ✓ There are no headings for each section, but there are three sections.	**0 pts Incomplete**	7 pts
Deliverable 3: Analysis	**18 to >15.0 pts Demonstrating Proficiency** ✓ The purpose is well defined. ✓ SIX to SEVEN metrics are addressed. ✓ THREE to FOUR major changes are summarized for the school district analysis.	**15 to >13.0 pts Approaching Proficiency** ✓ The purpose is well defined. ✓ FIVE to SIX of the SEVEN metrics are addressed. ✓ TWO to THREE major changes are summarized for the school district analysis.	**13 to >10.0 pts Developing Proficiency** ✓ The purpose is well defined. ✓ THREE to FOUR of the SEVEN metrics are addressed. ✓ ONE to TWO major changes are summarized for the school district analysis.	**10 to >0.0 pts Emerging** ✓ The purpose is well defined. ✓ Less than THREE of the SEVEN metrics are addressed. ✓ Only ONE major change is summarized or the summary does not adequately address the major changes to the school district analysis.	**0 pts Incomplete**	18 pts
						Total Points: 100