

4.8.5 Get the Passing Percentages Per School

Great work on finding the average math and reading scores for every high school. Now Maria would like you to continue gathering key data by calculating the passing percentages for math and reading for each school, as well as get the overall passing percentage for each school. She goes over what you will need in order to perform these calculations, writing the pseudocode on a whiteboard.

Here's the pseudocode from Maria:

```
# To get the passing percentages, we need to:  
# 1. Determine what is the passing grade.  
# 2. Get the number of students who passed math and reading.  
# 3. Get the students who passed math and passed reading
```

We're in luck! We have already determined the passing grade as well as calculated the number of students who passed math and reading.

Determine the Passing Grade

REWIND

Remember, for assessment tests, the passing score is 70, with the `>= 70` statement to filter the grades that are passing.

Get the Number of Students Who Passed Math and Reading

REWIND

We determined the number of students who passed math and reading for the district summary using the following code:

```
passing_math =  
school_data_complete_df[(school_data_complete_df["math_score"] >=  
70)]
```

and

```
passing_reading =  
school_data_complete_df[(school_data_complete_df["reading_score"]  
>= 70)]
```

So variables don't get reassigned, we'll use the same code but assign each calculation to new variables that reflect the students who passed math or

reading for each school.

Add the following code to a new cell and run the cell.

```
# Calculate the passing scores by creating a filtered DataFrame.  
per_school_passing_math = school_data_complete_df[(school_data_complete_df["  
  
per_school_passing_reading = school_data_complete_df[(school_data_complete_d
```

When we execute this cell and print the `per_school_passing_math`, we get a DataFrame that looks like this:



However, we need to get the average reading and math scores for each school. So, the index needs to be the `school_name`, and we need to get the number of students in the `per_school_passing_math` and the `per_school_passing_reading` DataFrames.



Calculating score. This might take a while. Please wait...

Your assignment has been successfully submitted.

You may close this window or continue to wait for your final summary.

To create a Series that has `school_name` as the index, we'll use `groupby(["school_name"])` on the `per_school_passing_math` and the `per_school_passing_math` DataFrames.

REWIND

Using the `groupby()` function will split an object (like a DataFrame) and apply a mathematical operation.

Next, we need to perform a mathematical operation on the `groupby()` object. Remember, we need to get the number of students who passed, so we'll use the `count()` method on the `student_name` column. Add the following code to a new cell and run the cell.

```
# Calculate the number of students passing math and passing reading by school
per_school_passing_math = per_school_passing_math.groupby(["school_name"]).c
per_school_passing_reading = per_school_passing_reading.groupby(["school_name"]).c
```

When we print the `per_school_passing_math`, the output will be a Series where the index is the school name and the column is the number of students who passed math for each school:

per_school_passing_math	
school_name	
Bailey High School	3318
Cabrera High School	1749
Figueroa High School	1946
Ford High School	1871
Griffin High School	1371
Hernandez High School	3094
Holden High School	395
Huang High School	1916
Johnson High School	3145
Pena High School	910
Rodriguez High School	2654
Shelton High School	1653
Thomas High School	1525
Wilson High School	2143
Wright High School	1680
Name: student_name, dtype: int64	

The `per_school_passing_reading` results will have the same format, with the column being the number of students who passed reading for each school.

Determine the Percentage of Students Passing Math and Reading

To determine the percentage of students passing math and reading, we must divide `per_school_passing_math` and `per_school_passing_reading` by the `per_school_counts`, and then multiply by 100, as shown in the following code:

```
# Calculate the percentage of passing math and reading scores per school.  
per_school_passing_math = per_school_passing_math / per_school_counts * 100  
  
per_school_passing_reading = per_school_passing_reading / per_school_counts
```

This time, when we print the `per_school_passing_math`, the output will be a Series where the index is the school name and the column is the percentage of students who passed math for each school.



The `per_school_passing_reading` results will have the same format, with the column being the percentage of students who passed reading.

Get the Overall Passing Percentage for All Students for Each School

REWIND

To get the overall passing percentage, we need to get all the students who passed both math and reading and then divide by the total number of students.

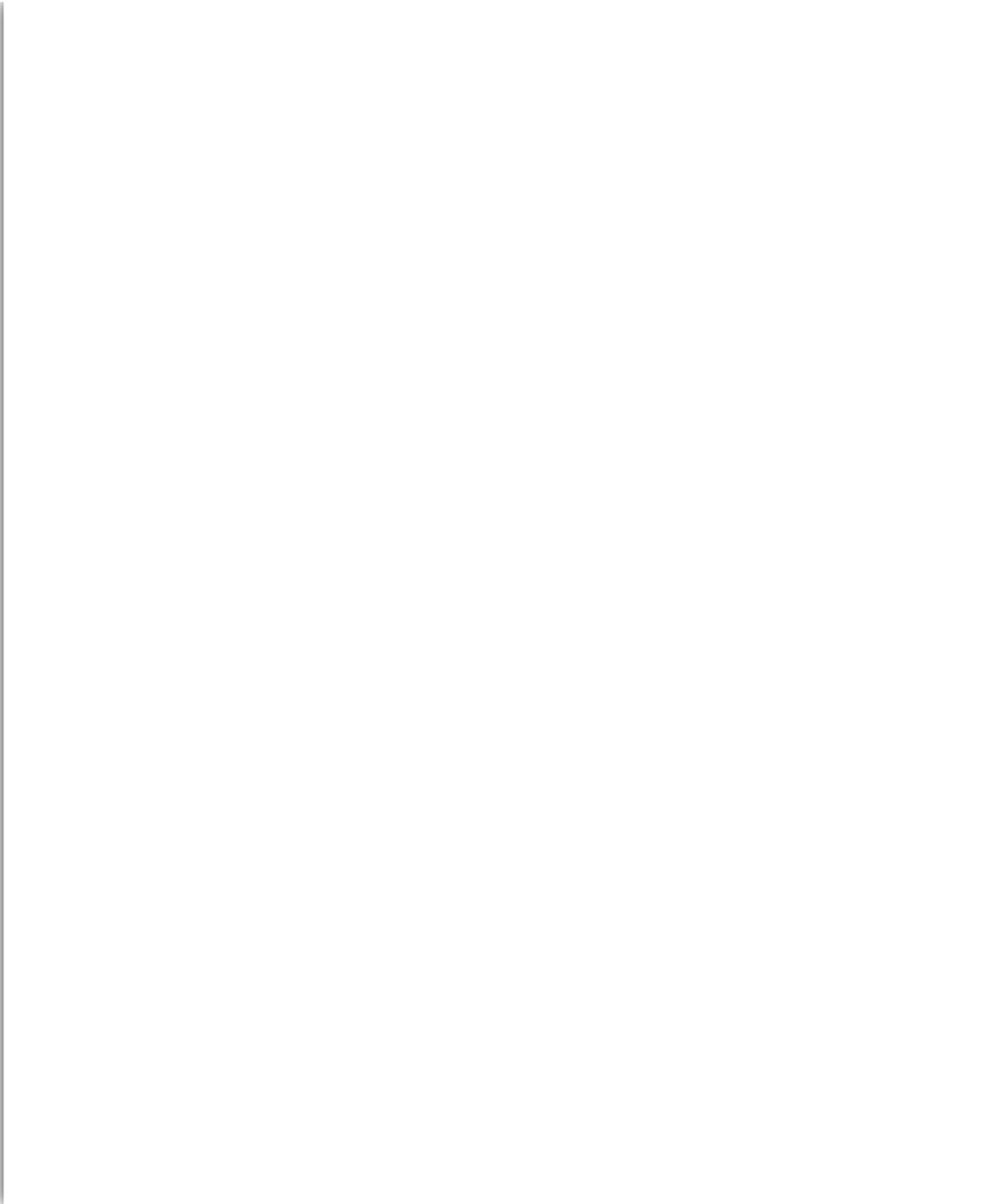
The code to calculate the overall passing percentage is as follows:

```
# Calculate the students who passed both math and reading.  
per_passing_math_reading = school_data_complete_df[(school_data_complete_df[  
  
per_passing_math_reading.head()
```

When we run this cell, the output will be a new DataFrame with all the columns that contain students who passed both math and reading.



Next, we'll get the total number of students from each school who passed both math and reading. To do this, we need to set the index as `school_name`, and then we need to get the number of students.





To get the total number of students who passed both math and reading on the `per_passing_math_reading` DataFrame, we use the following code, which sets the index to `school_name`, and then use the `count()` method for the `student_name`.

```
# Calculate the number of students who passed both math and reading.  
per_passing_math_reading = per_passing_math_reading.groupby(["school_name"])
```

Finally, we calculate the percentage of students who passed math and reading by dividing the the total number of students and multiplying by 100, using the following code.

```
# Calculate the overall passing percentage.  
per_overall_passing_percentage = per_passing_math_reading / per_school_count
```

When we execute this cell, we get a Series with the overall passing percentage for each school.

per_overall_passing_percentage	
Bailey High School	54.642283
Cabrera High School	61.334760

Congratulations—you've gathered all of the data you need! Now we can add the data to a new DataFrame.

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