

4.12.3 Group by School Size

Now that you have grouped the schools based on size, Maria would like you to calculate the following, based on the school sizes: average math and reading scores, the average percentage of students who passed math and reading, and the average overall percentage. After you gather this data, you'll add it to a new DataFrame to show how school size affects score averages and passing rates.

The final DataFrame will contain the average math and reading scores, the percentage of students who passed math and reading, and the overall percentage of students who passed for each school-size bin. Just as we did when we created the DataFrame for school spending per student, we'll use the `groupby()` function to group the data on the size bins in the index column, School Size.

First, let's create four new Series. We need to get the following data:

- Average Math Score
- Average Reading Score
- Average % Passing Math
- Average % Passing Reading

- Average % Overall Passing

Create each Series using the `groupby()` function on the School Size column. The data for each column will be calculated by using the `mean()` method to get the averages of the following columns: Average Math Score, Average Reading Score, % Passing Math, and % Passing Reading. For the average "% Overall Passing," we'll add the % Passing Math and % Passing Reading columns, and then divide by 2.

To get these averages, run the following code in a new cell.

```
# Calculate averages for the desired columns.  
size_math_scores = per_school_summary_df.groupby(["School Size"]).mean()["Av  
size_reading_scores = per_school_summary_df.groupby(["School Size"]).mean()  
size_passing_math = per_school_summary_df.groupby(["School Size"]).mean()["%  
size_passing_reading = per_school_summary_df.groupby(["School Size"]).mean()  
size_overall_passing = per_school_summary_df.groupby(["School Size"]).mean()
```

Now we can add these averages to a new DataFrame!