

Python_Lab_Assignment

October 2, 2022

1 PyCity Schools Analysis

- As a whole, schools with higher budgets, did not yield better test results. By contrast, schools with higher spending per student actually (\\$645 - 675) underperformed compared to schools with smaller budgets (\\$585 per student).
- As a whole, smaller and medium sized schools dramatically out-performed large sized schools on passing math performances (89-91% passing vs 67%).
-

1.1 As a whole, charter schools out-performed the public district schools across all metrics. However, more analysis will be required to glean if the effect is due to school practices or the fact that charter schools tend to serve smaller student populations per school.

Note: Instructions have been included for each segment. You do not have to follow them exactly, but they are included to help you think through the steps.

```
[90]: # Dependencies and Setup
import pandas as pd
import numpy as np

# File to Load (Remember to Change These)
school_data_to_load = "data/schools_complete.csv"
student_data_to_load = "data/students_complete.csv"

# Read School and Student Data File and store into Pandas Data Frames
school_data = pd.read_csv(school_data_to_load)
student_data = pd.read_csv(student_data_to_load)

# Combine the data into a single dataset
school_data_complete = pd.merge(student_data, school_data, how="left",
    on=["school_name", "school_name"])
school_data_complete.head()

# school_data_complete.count()
```

```
[90]:
```

	Student ID	student_name	gender	grade	school_name \
0	0	Paul Bradley	M	9th	Huang High School
1	1	Victor Smith	M	12th	Huang High School
2	2	Kevin Rodriguez	M	12th	Huang High School
3	3	Dr. Richard Scott	M	12th	Huang High School
4	4	Bonnie Ray	F	9th	Huang High School

	reading_score	math_score	School ID	type	size	budget
0	66	79	0	District	2917	1910635
1	94	61	0	District	2917	1910635
2	90	60	0	District	2917	1910635
3	67	58	0	District	2917	1910635
4	97	84	0	District	2917	1910635

1.2 District Summary

- Calculate the total number of schools
- Calculate the total number of students
- Calculate the total budget
- Calculate the average math score
- Calculate the average reading score
- Calculate the overall passing rate (overall average score), i.e. (avg. math score + avg. reading score)/2
- Calculate the percentage of students with a passing math score (70 or greater)
- Calculate the percentage of students with a passing reading score (70 or greater)
- Create a dataframe to hold the above results
- Optional: give the displayed data cleaner formatting

```
[91]: # Create a District Summary

# Total number of schools
total_number_of_schools = len(school_data_complete['school_name'].unique())
# Total number of students
total_number_of_students = school_data_complete['student_name'].count()
# Total budget
total_budget = sum(school_data_complete['budget'].unique())
# Average math score
average_math_score = school_data_complete['math_score'].mean()
# Average reading score
average_reading_score = school_data_complete['reading_score'].mean()
# Overall average score
overall_average_score = (average_math_score + average_reading_score)/2
# Percentage of passing math (70 or greater)
```

```

passing_math_percent =
    (school_data_complete[school_data_complete['math_score']>=70]['student_name'].
    count()/total_number_of_students)*100
# Percentage of passing reading
passing_reading_percent =
    (school_data_complete[school_data_complete['reading_score']>=70]['student_name'].
    count()/total_number_of_students)*100
# Percentage of passing overall
overall_passing_percent =
    (passing_math_percent+passing_reading_percent)/2

district_dataframe = pd.DataFrame(
    {
        'Total Schools':total_number_of_schools,
        'Total Student': '{:,}'.format(total_number_of_students),
        'Total Budget': '${:,.2f}'.format(total_budget),
        'Average Math Score':average_math_score,
        'Average Reading Score':average_reading_score,
        'Overall Score':[overall_average_score],
        '% Passing Math':passing_math_percent,
        '% Passing Reading':passing_reading_percent,
        '% Overall Passing':overall_passing_percent,
    }
)
district_dataframe

```

```

[91]: Total Schools Total Student    Total Budget Average Math Score \
0          15          39,170 $24,649,428.00          78.985371

      Average Reading Score Overall Score % Passing Math % Passing Reading \
0          81.87784          80.431606          74.980853          85.805463

      % Overall Passing
0          80.393158

```

```

[92]: # Total number of schools
total_number_of_schools

```

```

[92]: 15

```

```

[93]: # Total number of students
total_number_of_students

```

```

[93]: 39170

```

```

[94]: # Total budget
total_budget

```

[94]: 24649428

```
[95]: # Average math score
average_math_score
```

[95]: 78.98537145774827

```
[96]: # Average reading score
average_reading_score
```

[96]: 81.87784018381414

```
[97]: # Overall average score
overall_average_score
```

[97]: 80.43160582078121

```
[98]: # Percentage of passing math (70 or greater)
passing_math_percent
```

[98]: 74.9808526933878

1.3 School Summary

- Create an overview table that summarizes key metrics about each school, including:
 - School Name
 - School Type
 - Total Students
 - Total School Budget
 - Per Student Budget
 - Average Math Score
 - Average Reading Score
 - % Passing Math
 - % Passing Reading
 - Overall Passing Rate (Average of the above two)
- Create a dataframe to hold the above results

```
[99]: grouped_by_school = school_data_complete.groupby(['school_name'])
# School types
school_type = grouped_by_school['type'].first()
# Total number of students per school
total_students_count = grouped_by_school.size()
# Total school budget
total_budget = grouped_by_school['budget'].first()
# Total budget per school
total_budget_per_student = total_budget/total_students_count
# Average math score
average_math_score = grouped_by_school['math_score'].mean()
```

```

# Average reading score
average_reading_score = grouped_by_school['reading_score'].mean()
# Percentage of passing math (70 or greater)
grouped_passing_math = □
    ↳ school_data_complete[school_data_complete['math_score'] >= 70].
    ↳ groupby(['school_name']).size()
percent_passing_math = (grouped_passing_math/total_students_count)*100
# Percentage of passing reading
grouped_passing_reading = □
    ↳ school_data_complete[school_data_complete['reading_score'] >= 70].
    ↳ groupby(['school_name']).size()
percent_passing_reading = (grouped_passing_reading/total_students_count)*100
# Percentage of passing overall
percent_overall_passing = (percent_passing_math + percent_passing_reading)/2

school_dataframe = pd.DataFrame(
    {
        'School Type': school_type,
        'Total Students': total_students_count,
        'Total School Budget': total_budget,
        'Per Student Budget': total_budget_per_student,
        'Average Math Score': average_math_score,
        'Average Reading Score': average_reading_score,
        '% Passing Math': percent_passing_math,
        '% Passing Reading': percent_passing_reading,
        '% Overall Passing Rate': percent_overall_passing,
    }
)
school_dataframe.head()

```

```

[99]:

```

	School Type	Total Students	Total School Budget	\
school_name				
Bailey High School	District	4976	3124928	
Cabrera High School	Charter	1858	1081356	
Figueroa High School	District	2949	1884411	
Ford High School	District	2739	1763916	
Griffin High School	Charter	1468	917500	

	Per Student Budget	Average Math Score	\
school_name			
Bailey High School	628.0	77.048432	
Cabrera High School	582.0	83.061895	
Figueroa High School	639.0	76.711767	
Ford High School	644.0	77.102592	
Griffin High School	625.0	83.351499	

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

school_name		
Bailey High School	81.033963	66.680064
Cabrera High School	83.975780	94.133477
Figueroa High School	81.158020	65.988471
Ford High School	80.746258	68.309602
Griffin High School	83.816757	93.392371

	% Passing Reading	% Overall Passing Rate
school_name		
Bailey High School	81.933280	74.306672
Cabrera High School	97.039828	95.586652
Figueroa High School	80.739234	73.363852
Ford High School	79.299014	73.804308
Griffin High School	97.138965	95.265668

1.3.1 Top Performing Schools (By Passing Rate)

- Sort and display the top five schools in overall passing rate

```
[100]: # Sort and display the top five schools in overall passing rate
top_five_performing_schools = school_dataframe.sort_values(by='% Overall_
↳Passing Rate',ascending=False)
top_five_performing_schools.head()
```

```
[100]:
```

	School Type	Total Students	Total School Budget \
school_name			
Cabrera High School	Charter	1858	1081356
Thomas High School	Charter	1635	1043130
Pena High School	Charter	962	585858
Griffin High School	Charter	1468	917500
Wilson High School	Charter	2283	1319574

	Per Student Budget	Average Math Score \
school_name		
Cabrera High School	582.0	83.061895
Thomas High School	638.0	83.418349
Pena High School	609.0	83.839917
Griffin High School	625.0	83.351499
Wilson High School	578.0	83.274201

	Average Reading Score	% Passing Math	% Passing Reading \
school_name			
Cabrera High School	83.975780	94.133477	97.039828
Thomas High School	83.848930	93.272171	97.308869
Pena High School	84.044699	94.594595	95.945946
Griffin High School	83.816757	93.392371	97.138965
Wilson High School	83.989488	93.867718	96.539641

	% Overall Passing Rate
school_name	
Cabrera High School	95.586652
Thomas High School	95.290520
Pena High School	95.270270
Griffin High School	95.265668
Wilson High School	95.203679

```
[101]: # Calculate total school budget
school_budget = school_data_complete.groupby(["school_name"])["budget"].mean()
school_budget.head()
```

```
[101]: school_name
Bailey High School    3124928.0
Cabrera High School   1081356.0
Figueroa High School  1884411.0
Ford High School      1763916.0
Griffin High School    917500.0
Name: budget, dtype: float64
```

```
[102]: # Calculate per student budget
total_budget_per_student
```

```
[102]: school_name
Bailey High School    628.0
Cabrera High School   582.0
Figueroa High School  639.0
Ford High School      644.0
Griffin High School   625.0
Hernandez High School 652.0
Holden High School    581.0
Huang High School     655.0
Johnson High School  650.0
Pena High School      609.0
Rodriguez High School 637.0
Shelton High School   600.0
Thomas High School    638.0
Wilson High School    578.0
Wright High School    583.0
dtype: float64
```

```
[103]: # Calculate the avg math and reading score
average_math_score
```

```
[103]: school_name
Bailey High School    77.048432
```

Cabrera High School	83.061895
Figueroa High School	76.711767
Ford High School	77.102592
Griffin High School	83.351499
Hernandez High School	77.289752
Holden High School	83.803279
Huang High School	76.629414
Johnson High School	77.072464
Pena High School	83.839917
Rodriguez High School	76.842711
Shelton High School	83.359455
Thomas High School	83.418349
Wilson High School	83.274201
Wright High School	83.682222

Name: math_score, dtype: float64

[104]: average_reading_score

[104]:

school_name	
Bailey High School	81.033963
Cabrera High School	83.975780
Figueroa High School	81.158020
Ford High School	80.746258
Griffin High School	83.816757
Hernandez High School	80.934412
Holden High School	83.814988
Huang High School	81.182722
Johnson High School	80.966394
Pena High School	84.044699
Rodriguez High School	80.744686
Shelton High School	83.725724
Thomas High School	83.848930
Wilson High School	83.989488
Wright High School	83.955000

Name: reading_score, dtype: float64

Find the passing rate for math and reading (above 70 points)

[105]: percent_passing_math

[105]:

school_name	
Bailey High School	66.680064
Cabrera High School	94.133477
Figueroa High School	65.988471
Ford High School	68.309602
Griffin High School	93.392371
Hernandez High School	66.752967
Holden High School	92.505855

Huang High School	65.683922
Johnson High School	66.057551
Pena High School	94.594595
Rodriguez High School	66.366592
Shelton High School	93.867121
Thomas High School	93.272171
Wilson High School	93.867718
Wright High School	93.333333

dtype: float64

```
[106]: percent_passing_reading
```

```
[106]: school_name
Bailey High School      81.933280
Cabrera High School     97.039828
Figueroa High School    80.739234
Ford High School        79.299014
Griffin High School     97.138965
Hernandez High School   80.862999
Holden High School      96.252927
Huang High School       81.316421
Johnson High School     81.222432
Pena High School        95.945946
Rodriguez High School   80.220055
Shelton High School     95.854628
Thomas High School      97.308869
Wilson High School      96.539641
Wright High School      96.611111
dtype: float64
```

```
[107]: # Calculate the overall passing rate (average of the math and reading passing
      ↪rate)
percent_overall_passing
```

```
[107]: school_name
Bailey High School      74.306672
Cabrera High School     95.586652
Figueroa High School    73.363852
Ford High School        73.804308
Griffin High School     95.265668
Hernandez High School   73.807983
Holden High School      94.379391
Huang High School       73.500171
Johnson High School     73.639992
Pena High School        95.270270
Rodriguez High School   73.293323
Shelton High School     94.860875
```

Thomas High School	95.290520
Wilson High School	95.203679
Wright High School	94.972222

dtype: float64

1.3.2 Bottom Performing Schools (By Passing Rate)

- Sort and display the five worst-performing schools

```
[108]: # Sort and display the worst five schools in overall passing rate
worst_performing_schools = school_dataframe.sort_values(by='% Overall Passing_
↳Rate',ascending=True)
worst_performing_schools.head()
```

```
[108]:
```

	School Type	Total Students	Total School Budget \
school_name			
Rodriguez High School	District	3999	2547363
Figueroa High School	District	2949	1884411
Huang High School	District	2917	1910635
Johnson High School	District	4761	3094650
Ford High School	District	2739	1763916

	Per Student Budget	Average Math Score \
school_name		
Rodriguez High School	637.0	76.842711
Figueroa High School	639.0	76.711767
Huang High School	655.0	76.629414
Johnson High School	650.0	77.072464
Ford High School	644.0	77.102592

	Average Reading Score	% Passing Math \
school_name		
Rodriguez High School	80.744686	66.366592
Figueroa High School	81.158020	65.988471
Huang High School	81.182722	65.683922
Johnson High School	80.966394	66.057551
Ford High School	80.746258	68.309602

	% Passing Reading	% Overall Passing Rate
school_name		
Rodriguez High School	80.220055	73.293323
Figueroa High School	80.739234	73.363852
Huang High School	81.316421	73.500171
Johnson High School	81.222432	73.639992
Ford High School	79.299014	73.804308

1.4 Math Scores by Grade

- Create a table that lists the average Reading Score for students of each grade level (9th, 10th, 11th, 12th) at each school.
 - Create a pandas series for each grade. Hint: use a conditional statement.
 - Group each series by school
 - Combine the series into a dataframe
 - Optional: give the displayed data cleaner formatting

```
[109]: # Create table that lists the average math score for each school of each grade_
↳level.
math_9th_grade = student_data[student_data["grade"] == "9th"].
↳groupby(["school_name"])["math_score"].mean()
math_10th_grade = student_data[student_data["grade"] == "10th"].
↳groupby(["school_name"])["math_score"].mean()
math_11th_grade = student_data[student_data["grade"] == "11th"].
↳groupby(["school_name"])["math_score"].mean()
math_12th_grade = student_data[student_data["grade"] == "12th"].
↳groupby(["school_name"])["math_score"].mean()

math_by_grade = pd.DataFrame(
    {
        "9th":math_9th_grade,
        "10th":math_10th_grade,
        "11th":math_11th_grade,
        "12th":math_12th_grade
    }
)
math_by_grade
```

```
[109]:
```

	9th	10th	11th	12th
school_name				
Bailey High School	77.083676	76.996772	77.515588	76.492218
Cabrera High School	83.094697	83.154506	82.765560	83.277487
Figueroa High School	76.403037	76.539974	76.884344	77.151369
Ford High School	77.361345	77.672316	76.918058	76.179963
Griffin High School	82.044010	84.229064	83.842105	83.356164
Hernandez High School	77.438495	77.337408	77.136029	77.186567
Holden High School	83.787402	83.429825	85.000000	82.855422
Huang High School	77.027251	75.908735	76.446602	77.225641
Johnson High School	77.187857	76.691117	77.491653	76.863248
Pena High School	83.625455	83.372000	84.328125	84.121547
Rodriguez High School	76.859966	76.612500	76.395626	77.690748
Shelton High School	83.420755	82.917411	83.383495	83.778976
Thomas High School	83.590022	83.087886	83.498795	83.497041
Wilson High School	83.085578	83.724422	83.195326	83.035794

Wright High School 83.264706 84.010288 83.836782 83.644986

```
[110]: # Calculate the average math score for 9th grade in each school
math_9th_grade
```

```
[110]: school_name
Bailey High School      77.083676
Cabrera High School     83.094697
Figueroa High School    76.403037
Ford High School        77.361345
Griffin High School     82.044010
Hernandez High School   77.438495
Holden High School      83.787402
Huang High School       77.027251
Johnson High School    77.187857
Pena High School        83.625455
Rodriguez High School   76.859966
Shelton High School     83.420755
Thomas High School      83.590022
Wilson High School      83.085578
Wright High School      83.264706
Name: math_score, dtype: float64
```

```
[111]: # Calculate the average math score for 10th grade in each school
math_10th_grade
```

```
[111]: school_name
Bailey High School      76.996772
Cabrera High School     83.154506
Figueroa High School    76.539974
Ford High School        77.672316
Griffin High School     84.229064
Hernandez High School   77.337408
Holden High School      83.429825
Huang High School       75.908735
Johnson High School    76.691117
Pena High School        83.372000
Rodriguez High School   76.612500
Shelton High School     82.917411
Thomas High School      83.087886
Wilson High School      83.724422
Wright High School      84.010288
Name: math_score, dtype: float64
```

```
[112]: # Calculate the average math score for 11th grade in each school
math_11th_grade
```

```
[112]: school_name
       Bailey High School      77.515588
       Cabrera High School    82.765560
       Figueroa High School   76.884344
       Ford High School       76.918058
       Griffin High School    83.842105
       Hernandez High School   77.136029
       Holden High School     85.000000
       Huang High School      76.446602
       Johnson High School     77.491653
       Pena High School       84.328125
       Rodriguez High School   76.395626
       Shelton High School    83.383495
       Thomas High School     83.498795
       Wilson High School     83.195326
       Wright High School     83.836782
       Name: math_score, dtype: float64
```

```
[113]: # Calculate the average math score for 12th grade in each school
       math_12th_grade
```

```
[113]: school_name
       Bailey High School      76.492218
       Cabrera High School    83.277487
       Figueroa High School   77.151369
       Ford High School       76.179963
       Griffin High School    83.356164
       Hernandez High School   77.186567
       Holden High School     82.855422
       Huang High School      77.225641
       Johnson High School     76.863248
       Pena High School       84.121547
       Rodriguez High School   77.690748
       Shelton High School    83.778976
       Thomas High School     83.497041
       Wilson High School     83.035794
       Wright High School     83.644986
       Name: math_score, dtype: float64
```

1.4.1 Reading Score by Grade

- Perform the same operations as above for reading scores

```
[114]: # Create table that lists the average reading score for each school of each
       ↪ grade level.
       read_9th_grade = student_data[student_data["grade"] == "9th"].
       ↪ groupby(["school_name"])["reading_score"].mean()
```

```

read_10th_grade = student_data[student_data["grade"] == "10th"].
    ↳groupby(["school_name"])["reading_score"].mean()
read_11th_grade = student_data[student_data["grade"] == "11th"].
    ↳groupby(["school_name"])["reading_score"].mean()
read_12th_grade = student_data[student_data["grade"] == "12th"].
    ↳groupby(["school_name"])["reading_score"].mean()

reading_by_grade = pd.DataFrame(
    {
        "9th":read_9th_grade,
        "10th":read_10th_grade,
        "11th":read_11th_grade,
        "12th":read_12th_grade
    }
)
reading_by_grade

```

```

[114]:

```

	9th	10th	11th	12th
school_name				
Bailey High School	81.303155	80.907183	80.945643	80.912451
Cabrera High School	83.676136	84.253219	83.788382	84.287958
Figueroa High School	81.198598	81.408912	80.640339	81.384863
Ford High School	80.632653	81.262712	80.403642	80.662338
Griffin High School	83.369193	83.706897	84.288089	84.013699
Hernandez High School	80.866860	80.660147	81.396140	80.857143
Holden High School	83.677165	83.324561	83.815534	84.698795
Huang High School	81.290284	81.512386	81.417476	80.305983
Johnson High School	81.260714	80.773431	80.616027	81.227564
Pena High School	83.807273	83.612000	84.335938	84.591160
Rodriguez High School	80.993127	80.629808	80.864811	80.376426
Shelton High School	84.122642	83.441964	84.373786	82.781671
Thomas High School	83.728850	84.254157	83.585542	83.831361
Wilson High School	83.939778	84.021452	83.764608	84.317673
Wright High School	83.833333	83.812757	84.156322	84.073171

```

[115]: # Calculate the average reading score for 9th grade in each school
read_9th_grade

```

```

[115]: school_name
Bailey High School      81.303155
Cabrera High School     83.676136
Figueroa High School    81.198598
Ford High School        80.632653
Griffin High School     83.369193
Hernandez High School   80.866860
Holden High School      83.677165
Huang High School       81.290284

```

Johnson High School	81.260714
Pena High School	83.807273
Rodriguez High School	80.993127
Shelton High School	84.122642
Thomas High School	83.728850
Wilson High School	83.939778
Wright High School	83.833333

Name: reading_score, dtype: float64

```
[116]: # Calculate the average reading score for 10th grade in each school
read_10th_grade
```

```
[116]: school_name
Bailey High School      80.907183
Cabrera High School     84.253219
Figueroa High School    81.408912
Ford High School        81.262712
Griffin High School     83.706897
Hernandez High School   80.660147
Holden High School      83.324561
Huang High School       81.512386
Johnson High School     80.773431
Pena High School         83.612000
Rodriguez High School    80.629808
Shelton High School     83.441964
Thomas High School      84.254157
Wilson High School       84.021452
Wright High School      83.812757
Name: reading_score, dtype: float64
```

```
[117]: # Calculate the average reading score for 11th grade in each school
read_11th_grade
```

```
[117]: school_name
Bailey High School      80.945643
Cabrera High School     83.788382
Figueroa High School    80.640339
Ford High School        80.403642
Griffin High School     84.288089
Hernandez High School   81.396140
Holden High School      83.815534
Huang High School       81.417476
Johnson High School     80.616027
Pena High School         84.335938
Rodriguez High School    80.864811
Shelton High School     84.373786
Thomas High School      83.585542
```

```
Wilson High School      83.764608
Wright High School     84.156322
Name: reading_score, dtype: float64
```

```
[118]: # Calculate the average reading score for 12th grade in each school
read_12th_grade
```

```
[118]: school_name
Bailey High School      80.912451
Cabrera High School     84.287958
Figueroa High School    81.384863
Ford High School        80.662338
Griffin High School     84.013699
Hernandez High School   80.857143
Holden High School      84.698795
Huang High School       80.305983
Johnson High School     81.227564
Pena High School        84.591160
Rodriguez High School   80.376426
Shelton High School     82.781671
Thomas High School      83.831361
Wilson High School      84.317673
Wright High School      84.073171
Name: reading_score, dtype: float64
```

1.5 Scores by School Spending

- Create a table that breaks down school performances based on average Spending Ranges (Per Student). Use 4 reasonable bins to group school spending. Include in the table each of the following:
 - Average Math Score
 - Average Reading Score
 - % Passing Math
 - % Passing Reading
 - Overall Passing Rate (Average of the above two)

```
[119]: # Sample bins. Feel free to create your own bins.
spending_bins = [0, 585, 615, 645, 675]
group_names = ["<$585", "$585-615", "$615-645", "$645-675"]
```

```
[120]: # Create a new column to show budget per student in each row
school_spending = school_dataframe.loc[:,["Average Math Score","Average Reading_
↪Score","% Passing Math","% Passing Reading","% Overall Passing Rate","Per_
↪Student Budget"]]
school_spending
```


[120]:

	Average Math Score	Average Reading Score \
school_name		
Bailey High School	77.048432	81.033963
Cabrera High School	83.061895	83.975780
Figueroa High School	76.711767	81.158020
Ford High School	77.102592	80.746258
Griffin High School	83.351499	83.816757
Hernandez High School	77.289752	80.934412
Holden High School	83.803279	83.814988
Huang High School	76.629414	81.182722
Johnson High School	77.072464	80.966394
Pena High School	83.839917	84.044699
Rodriguez High School	76.842711	80.744686
Shelton High School	83.359455	83.725724
Thomas High School	83.418349	83.848930
Wilson High School	83.274201	83.989488
Wright High School	83.682222	83.955000

	% Passing Math	% Passing Reading \
school_name		
Bailey High School	66.680064	81.933280
Cabrera High School	94.133477	97.039828
Figueroa High School	65.988471	80.739234
Ford High School	68.309602	79.299014
Griffin High School	93.392371	97.138965
Hernandez High School	66.752967	80.862999
Holden High School	92.505855	96.252927
Huang High School	65.683922	81.316421
Johnson High School	66.057551	81.222432
Pena High School	94.594595	95.945946
Rodriguez High School	66.366592	80.220055
Shelton High School	93.867121	95.854628
Thomas High School	93.272171	97.308869
Wilson High School	93.867718	96.539641
Wright High School	93.333333	96.611111

	% Overall Passing Rate	Per Student Budget
school_name		
Bailey High School	74.306672	628.0
Cabrera High School	95.586652	582.0
Figueroa High School	73.363852	639.0
Ford High School	73.804308	644.0
Griffin High School	95.265668	625.0
Hernandez High School	73.807983	652.0
Holden High School	94.379391	581.0
Huang High School	73.500171	655.0
Johnson High School	73.639992	650.0

Pena High School	95.270270	609.0
Rodriguez High School	73.293323	637.0
Shelton High School	94.860875	600.0
Thomas High School	95.290520	638.0
Wilson High School	95.203679	578.0
Wright High School	94.972222	583.0

```
[121]: # Create a new column to define the spending ranges per student
school_spending["Per Student Spending Ranges"] = pd.cut(school_dataframe["Per_Student Budget"], spending_bins, labels=group_names)
school_spending = school_spending.groupby("Per Student Spending Ranges").mean()
school_spending.head()
```

```
[121]:
```

	Average Math Score	Average Reading Score \
Per Student Spending Ranges		
<\$585	83.455399	83.933814
\$585-615	83.599686	83.885211
\$615-645	79.079225	81.891436
\$645-675	76.997210	81.027843

	% Passing Math	% Passing Reading \
Per Student Spending Ranges		
<\$585	93.460096	96.610877
\$585-615	94.230858	95.900287
\$615-645	75.668212	86.106569
\$645-675	66.164813	81.133951

	% Overall Passing Rate	Per Student Budget
Per Student Spending Ranges		
<\$585	95.035486	581.000000
\$585-615	95.065572	604.500000
\$615-645	80.887391	635.166667
\$645-675	73.649382	652.333333

```
[122]: # Calculate the average math score within each spending range
school_spending["Average Math Score"]
```

```
[122]: Per Student Spending Ranges
<$585      83.455399
$585-615   83.599686
$615-645   79.079225
$645-675   76.997210
Name: Average Math Score, dtype: float64
```

```
[123]: # Calculate the percentage passing rate for math in each spending range
school_spending["% Passing Math"]
```

```
[123]: Per Student Spending Ranges
<$585      93.460096
$585-615   94.230858
$615-645   75.668212
$645-675   66.164813
Name: % Passing Math, dtype: float64
```

```
[124]: # Calculate the percentage passing rate for reading in each spending range
school_spending["% Passing Reading"]
```

```
[124]: Per Student Spending Ranges
<$585      96.610877
$585-615   95.900287
$615-645   86.106569
$645-675   81.133951
Name: % Passing Reading, dtype: float64
```

```
[125]: # Calculate the percentage overall passing rate in each spending range
school_spending["% Overall Passing Rate"]
```

```
[125]: Per Student Spending Ranges
<$585      95.035486
$585-615   95.065572
$615-645   80.887391
$645-675   73.649382
Name: % Overall Passing Rate, dtype: float64
```

1.5.1 Scores by School Size

- Perform the same operations as above, based on school size.

```
[126]: # Sample bins. Feel free to create your own bins.
size_bins = [0, 1000, 2000, 5000]
group_names = ["Small (<1000)", "Medium (1000-2000)", "Large (2000-5000)"]
```

```
[127]: # Create a new column for the bin groups
school_size = school_dataframe.loc[:,['Average Math Score','Average Reading_
    ↳Score','% Passing Math','% Passing Reading','% Overall Passing Rate',]]
school_size['School Size']= pd.cut(school_dataframe['Total_
    ↳Students'],size_bins,labels=group_names)
school_size
```

```
[127]:
```

	Average Math Score	Average Reading Score \
school_name		
Bailey High School	77.048432	81.033963
Cabrera High School	83.061895	83.975780
Figueroa High School	76.711767	81.158020
Ford High School	77.102592	80.746258

Griffin High School	83.351499	83.816757
Hernandez High School	77.289752	80.934412
Holden High School	83.803279	83.814988
Huang High School	76.629414	81.182722
Johnson High School	77.072464	80.966394
Pena High School	83.839917	84.044699
Rodriguez High School	76.842711	80.744686
Shelton High School	83.359455	83.725724
Thomas High School	83.418349	83.848930
Wilson High School	83.274201	83.989488
Wright High School	83.682222	83.955000

	% Passing Math	% Passing Reading \
school_name		
Bailey High School	66.680064	81.933280
Cabrera High School	94.133477	97.039828
Figueroa High School	65.988471	80.739234
Ford High School	68.309602	79.299014
Griffin High School	93.392371	97.138965
Hernandez High School	66.752967	80.862999
Holden High School	92.505855	96.252927
Huang High School	65.683922	81.316421
Johnson High School	66.057551	81.222432
Pena High School	94.594595	95.945946
Rodriguez High School	66.366592	80.220055
Shelton High School	93.867121	95.854628
Thomas High School	93.272171	97.308869
Wilson High School	93.867718	96.539641
Wright High School	93.333333	96.611111

	% Overall Passing Rate	School Size
school_name		
Bailey High School	74.306672	Large (2000-5000)
Cabrera High School	95.586652	Medium (1000-2000)
Figueroa High School	73.363852	Large (2000-5000)
Ford High School	73.804308	Large (2000-5000)
Griffin High School	95.265668	Medium (1000-2000)
Hernandez High School	73.807983	Large (2000-5000)
Holden High School	94.379391	Small (<1000)
Huang High School	73.500171	Large (2000-5000)
Johnson High School	73.639992	Large (2000-5000)
Pena High School	95.270270	Small (<1000)
Rodriguez High School	73.293323	Large (2000-5000)
Shelton High School	94.860875	Medium (1000-2000)
Thomas High School	95.290520	Medium (1000-2000)
Wilson High School	95.203679	Large (2000-5000)
Wright High School	94.972222	Medium (1000-2000)

Look for the total count of test scores that pass 70% or higher

```
[158]: math_pass_size = school_size[school_size['% Passing Math']>=70].groupby('School_
↳Size')['% Passing Math'].count()
math_pass_size
```

```
[158]: School Size
Small (<1000)      2
Medium (1000-2000) 5
Large (2000-5000)  1
Name: % Passing Math, dtype: int64
```

```
[151]: # read_pass_size
read_pass_size = school_size[school_size['% Passing Reading']>=70].
↳groupby('School Size')['% Passing Reading'].count()
read_pass_size
```

```
[151]: School Size
Small (<1000)      2
Medium (1000-2000) 5
Large (2000-5000)  8
Name: % Passing Reading, dtype: int64
```

```
[150]: # Calculate the overall passing rate for different school size
by_overall_passing_size = school_size[["School Size", "% Overall Passing Rate"]]
by_overall_passing_size = by_overall_passing_size.groupby(["School Size"]).
↳mean()
by_overall_passing_size
```

```
[150]: % Overall Passing Rate
School Size
Small (<1000)      94.824831
Medium (1000-2000) 95.195187
Large (2000-5000)  76.364998
```

1.5.2 Scores by School Type

- Perform the same operations as above, based on school type.

```
[143]: # Create bins and groups, school type {'Charter', 'District'}
school_type = school_dataframe[['School Type', 'Average Math Score', 'Average_
↳Reading Score', '% Passing Math', '% Passing Reading', '% Overall Passing_
↳Rate',]]
school_type = school_type.groupby('School Type').mean()
school_type.head()
```

```
[143]: Average Math Score  Average Reading Score  % Passing Math  \
School Type
```

Charter	83.473852	83.896421	93.620830
District	76.956733	80.966636	66.548453

	% Passing Reading	% Overall Passing Rate
School Type		
Charter	96.586489	95.103660
District	80.799062	73.673757

Find counts of the passing 70 or higher score for the both test

```
[144]: # math pass size
math_pass_size_type = school_dataframe[school_dataframe['% Passing Math']>=70].
    ↳groupby('School Type')['% Passing Math'].count()
math_pass_size_type
```

```
[144]: School Type
Charter      8
Name: % Passing Math, dtype: int64
```

```
[156]: # reading pass size
read_pass_size_type = school_dataframe[school_dataframe['% Passing_
    ↳Reading']>=70].groupby('School Type')['% Passing Reading'].count()
read_pass_size_type
```

```
[156]: School Type
Charter      8
District     7
Name: % Passing Reading, dtype: int64
```

```
[155]: # Calculate the overall passing rate
by_overall_passing_type = school_dataframe[["School Type", "% Overall Passing_
    ↳Rate"]]
by_overall_passing_type = by_overall_passing_type.groupby(["School Type"]).
    ↳mean()
by_overall_passing_type
```

```
[155]: % Overall Passing Rate
School Type
Charter      95.103660
District     73.673757
```