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%).
- 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.

In [116]:

```
# Dependencies and Setup
import pandas as pd

# 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",
school_data_complete

# school_data_complete.count()
```

Out[116]:

	Student ID	student_name	gender	grade	school_name	reading_score	math_score	Scho I
0	0	Paul Bradley	M	9th	Huang High School	66	79	
1	1	Victor Smith	M	12th	Huang High School	94	61	
2	2	Kevin Rodriguez	M	12th	Huang High School	90	60	
3	3	Dr. Richard Scott	M	12th	Huang High School	67	58	
4	4	Bonnie Ray	F	9th	Huang High School	97	84	
...	...	...	...	...	...	...	...	
39165	39165	Donna Howard	F	12th	Thomas High School	99	90	-
39166	39166	Dawn Bell	F	10th	Thomas High School	95	70	-
39167	39167	Rebecca Tanner	F	9th	Thomas High School	73	84	-
39168	39168	Desiree Kidd	F	10th	Thomas High School	99	90	-
39169	39169	Carolyn Jackson	F	11th	Thomas High School	95	75	-

39170 rows × 11 columns

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

In [10]:

```
# Create a District Summary
Dist_Sum = pd.DataFrame({'Total Number of Schools': [total_schools],
                       'Total Students': [total_students],
                       'Total Budget': [total_budget],
                       'Average Math Score': [avg_math_score],
                       'Average Reading Score': [avg_read_score],
                       'Overall Average Score': [avg_ovr_score],
                       'Percent Passing Math': [pass_math_perc],
                       'Percent Passing Reading':[pass_read_perc]})

Dist_Sum.style.format({"Total Budget": "${:,d}"})
```

Out[10]:

	Total Number of Schools	Total Students	Total Budget	Average Math Score	Average Reading Score	Overall Average Score	Percent Passing Math	Percent Passing Reading
0	15	39170	$24,649,428	78.985371	81.877840	160.863212	0.749809	0.858055

In [3]:

```
# Total number of schools
total_schools = school_data_complete.school_name.nunique()
total_schools
```

Out[3]:

```
15
```

In [4]:
```
# Total number of students
total_students = school_data_complete['Student ID'].count()
total_students
```

Out[4]:
```
39170
```

In [5]:
```
# Total budget

total_budget = school_data["budget"].sum()
total_budget
```

Out[5]:
```
24649428
```

In [6]:
```
# Average math score

avg_math_score = school_data_complete['math_score'].mean()
avg_math_score
```

Out[6]:
```
78.98537145774827
```

In [7]:
```
# Average reading score

avg_read_score = school_data_complete['reading_score'].mean()
avg_read_score
```

Out[7]:
```
81.87784018381414
```

In [8]:
```
# Overall average score

avg_ovr_score = avg_read_score + avg_math_score
avg_ovr_score
```

Out[8]:
```
160.86321164156243
```

In [9]:
```
# Percentage of passing math (70 or greater)
pass_math_perc = (school_data_complete['math_score']>=70).sum()/total_
print(pass_math_perc)

pass_read_perc = (school_data_complete['reading_score']>=70).sum()/tot
print(pass_read_perc)
```

```
0.749808526933878
0.8580546336482001
```

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

Top Performing Schools (By Passing Rate)

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

In [74]:
```
#  Sort and display the top five schools in overall passing rate
topfive = indi_sch.sort_values('Overall Passing Rate', ascending = Fal

topfive.head(5)
```

Out[74]:

	Type of School	Student Count	Budget of School	Budget Per Student	Average Math Score	Average Reading Score	Math Passing Rate	Reading Passing Rate	F
Cabrera High School	Charter	1858	1081356.0	582.0	83.061895	83.975780	94.133477	97.039828	95.
Thomas High School	Charter	1635	1043130.0	638.0	83.418349	83.848930	93.272171	97.308869	95.
Pena High School	Charter	962	585858.0	609.0	83.839917	84.044699	94.594595	95.945946	95.
Griffin High School	Charter	1468	917500.0	625.0	83.351499	83.816757	93.392371	97.138965	95.
Wilson High School	Charter	2283	1319574.0	578.0	83.274201	83.989488	93.867718	96.539641	95.

In [12]:
```
# Calculate total school budget
individual_budg = school_data_complete.groupby(['school_name'])['budge
individual_budg
```

Out[12]:
```
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
Hernandez High School    3022020.0
Holden High School        248087.0
Huang High School        1910635.0
Johnson High School      3094650.0
Pena High School          585858.0
Rodriguez High School    2547363.0
Shelton High School      1056600.0
Thomas High School       1043130.0
Wilson High School       1319574.0
Wright High School       1049400.0
Name: budget, dtype: float64
```

In [27]:
```
# Calculate per student budget
per_stu_budg = individual_budg/students_per_sch

per_stu_budg
```

Out[27]:
```
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
```

In [15]:
```
# Cacluate the avg math and reading score

avgmath_per_sch = school_data_complete.groupby(['school_name'])['math_
avgread_per_sch = school_data_complete.groupby(['school_name'])['readi
```

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

In [184]:

```
# Find the total counts of math result
counts_math = school_data_complete.groupby(['school_name'])['math_scor

# Find the counts for math result in each school that pass 70 or highe
pass_math = school_data_complete[(school_data_complete['math_score']>=
pass_math2 = pass_math.groupby(['school_name'])['math_score'].count()
# Calculate the math passing rate
pass_mathcount = pass_math2/counts_math
pass_mathcount
```

Out[184]:

```
school_name
Bailey High School       0.666801
Cabrera High School      0.941335
Figueroa High School     0.659885
Ford High School         0.683096
Griffin High School      0.933924
Hernandez High School    0.667530
Holden High School       0.925059
Huang High School        0.656839
Johnson High School      0.660576
Pena High School         0.945946
Rodriguez High School    0.663666
Shelton High School      0.938671
Thomas High School       0.932722
Wilson High School       0.938677
Wright High School       0.933333
Name: math_score, dtype: float64
```

In [185]:

```
# Find the total counts of read result
counts_read = school_data_complete.groupby(['school_name'])['reading_s
# Find the counts for read result in each school that pass 70 or highe
pass_read = school_data_complete[(school_data_complete['reading_score'
pass_read2 = pass_read.groupby(['school_name'])['reading_score'].count
# Calculate the read passing rate
pass_readcount = pass_read2/counts_read
pass_readcount
```

Out[185]:

```
school_name
Bailey High School       0.819333
Cabrera High School      0.970398
Figueroa High School     0.807392
Ford High School         0.792990
Griffin High School      0.971390
Hernandez High School    0.808630
Holden High School       0.962529
Huang High School        0.813164
Johnson High School      0.812224
Pena High School         0.959459
Rodriguez High School    0.802201
Shelton High School      0.958546
Thomas High School       0.973089
Wilson High School       0.965396
Wright High School       0.966111
Name: reading_score, dtype: float64
```

In [186]:

```
# Calculate the overall passing rate (average of the math and reading

ovrpass_rate = (pass_readcount+pass_mathcount)/2
ovrpass_rate
```

Out[186]:

```
school_name
Bailey High School       0.743067
Cabrera High School      0.955867
Figueroa High School     0.733639
Ford High School         0.738043
Griffin High School      0.952657
Hernandez High School    0.738080
Holden High School       0.943794
Huang High School        0.735002
Johnson High School      0.736400
Pena High School         0.952703
Rodriguez High School    0.732933
Shelton High School      0.948609
Thomas High School       0.952905
Wilson High School       0.952037
Wright High School       0.949722
dtype: float64
```

In [187]:

```

school_types = school_data.set_index(['school_name'])['type']
```

In [188]:

```
students_per_sch = student_data['school_name'].value_counts()
students_per_sch
```

Out[188]:

```
Bailey High School       4976
Johnson High School      4761
Hernandez High School    4635
Rodriguez High School    3999
Figueroa High School     2949
Huang High School        2917
Ford High School         2739
Wilson High School       2283
Cabrera High School      1858
Wright High School       1800
Shelton High School      1761
Thomas High School       1635
Griffin High School      1468
Pena High School          962
Holden High School        427
Name: school_name, dtype: int64
```

In [189]:

```
indi_sch=pd.DataFrame({'Type of School': school_types,
                       'Student Count': students_per_sch,
                       'Budget of School': individual_budg,
                       'Budget Per Student': per_stu_budg,
                       'Average Math Score': avgmath_per_sch,
                       'Average Reading Score': avgread_per_sch,
                       'Math Passing Rate': pass_mathcount,
                       'Reading Passing Rate': pass_readcount,
                       'Overall Passing Rate': ovrpass_rate })
indi_sch
```

Out[189]:

	Type of School	Student Count	Budget of School	Budget Per Student	Average Math Score	Average Reading Score	Math Passing Rate	Reading Passing Rate	O P
Bailey High School	District	4976	3124928.0	628.0	77.048432	81.033963	0.666801	0.819333	0.7
Cabrera High School	Charter	1858	1081356.0	582.0	83.061895	83.975780	0.941335	0.970398	0.9
Figueroa High School	District	2949	1884411.0	639.0	76.711767	81.158020	0.659885	0.807392	0.7
Ford High School	District	2739	1763916.0	644.0	77.102592	80.746258	0.683096	0.792990	0.7
Griffin High School	Charter	1468	917500.0	625.0	83.351499	83.816757	0.933924	0.971390	0.9
Hernandez High School	District	4635	3022020.0	652.0	77.289752	80.934412	0.667530	0.808630	0.7
Holden High School	Charter	427	248087.0	581.0	83.803279	83.814988	0.925059	0.962529	0.9
Huang High School	District	2917	1910635.0	655.0	76.629414	81.182722	0.656839	0.813164	0.7
Johnson High School	District	4761	3094650.0	650.0	77.072464	80.966394	0.660576	0.812224	0.7
Pena High School	Charter	962	585858.0	609.0	83.839917	84.044699	0.945946	0.959459	0.9
Rodriguez High School	District	3999	2547363.0	637.0	76.842711	80.744686	0.663666	0.802201	0.7

Shelton High School	Charter	1761	1056600.0	600.0	83.359455	83.725724	0.938671	0.958546	0.9
Thomas High School	Charter	1635	1043130.0	638.0	83.418349	83.848930	0.932722	0.973089	0.9
Wilson High School	Charter	2283	1319574.0	578.0	83.274201	83.989488	0.938677	0.965396	0.9
Wright High School	Charter	1800	1049400.0	583.0	83.682222	83.955000	0.933333	0.966111	0.9

Bottom Performing Schools (By Passing Rate)

- Sort and display the five worst-performing schools

In [190]:
```
#  Sort and display the worst five schools in overall passing rate
lowfive = indi_sch.sort_values('Overall Passing Rate', ascending = Tru
lowfive.head(5)
```

Out[190]:

	Type of School	Student Count	Budget of School	Budget Per Student	Average Math Score	Average Reading Score	Math Passing Rate	Reading Passing Rate	O Pa
Rodriguez High School	District	3999	2547363.0	637.0	76.842711	80.744686	0.663666	0.802201	0.7
Figueroa High School	District	2949	1884411.0	639.0	76.711767	81.158020	0.659885	0.807392	0.7
Huang High School	District	2917	1910635.0	655.0	76.629414	81.182722	0.656839	0.813164	0.7
Johnson High School	District	4761	3094650.0	650.0	77.072464	80.966394	0.660576	0.812224	0.7
Ford High School	District	2739	1763916.0	644.0	77.102592	80.746258	0.683096	0.792990	0.7

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

In [191]:

```
# Create table that lists the average math score for each school of ea
math_by_grade = pd.DataFrame({'9th': freshmen,
                            '10th': sophomore,
                             '11th': junior,
                            '12th': senior})
math_by_grade
```

Out[191]:

school_name	9th	10th	11th	12th
Bailey High School	77.083676	80.907183	77.515588	76.492218
Cabrera High School	83.094697	84.253219	82.765560	83.277487
Figueroa High School	76.403037	81.408912	76.884344	77.151369
Ford High School	77.361345	81.262712	76.918058	76.179963
Griffin High School	82.044010	83.706897	83.842105	83.356164
Hernandez High School	77.438495	80.660147	77.136029	77.186567
Holden High School	83.787402	83.324561	85.000000	82.855422
Huang High School	77.027251	81.512386	76.446602	77.225641
Johnson High School	77.187857	80.773431	77.491653	76.863248
Pena High School	83.625455	83.612000	84.328125	84.121547
Rodriguez High School	76.859966	80.629808	76.395626	77.690748
Shelton High School	83.420755	83.441964	83.383495	83.778976
Thomas High School	83.590022	84.254157	83.498795	83.497041
Wilson High School	83.085578	84.021452	83.195326	83.035794
Wright High School	83.264706	83.812757	83.836782	83.644986

In [192]:

```
# Calculate the average math score for 9th grade in each school
freshmen = student_data.loc[student_data['grade']== '9th'].groupby(['s
freshmen
```

Out[192]:

```
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
```

In [193]:

```
# Calculate the average math score for 10th grade in each school
sophomore = student_data.loc[student_data['grade']=='10th'].groupby(['
sophomore
```

Out[193]:

```
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
```

In [194]:

```
# Calculate the average math score for 11th grade in each school
junior = student_data.loc[student_data['grade']== '11th'].groupby(['sc
junior
```

Out[194]:

```
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
```

In [195]:

```
# Calculate the average math score for 12th grade in each school
senior = student_data.loc[student_data['grade'] == '12th'].groupby(['s
senior
```

Out[195]:

```
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
```

Reading Score by Grade

- Perform the same operations as above for reading scores

In [196]:

```
# Create table that lists the average reading score for each school of
read_by_grade = pd.DataFrame({'9th': freshmen1,
                              '10th': sophomore1,
                              '11th': junior1,
                              '12th': senior1})
read_by_grade
```

Out[196]:

	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

In [197]:

```
# Calculate the average reading score for 9th grade in each school

freshmen1 = student_data.loc[student_data['grade']=='9th'].groupby(['s
freshmen1
```

Out[197]:

```
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
```

In [198]:

```
# Calculate the average reading score for 10th grade in each school
sophomore1 = student_data.loc[student_data['grade']=='10th'].groupby([
sophomore1
```

Out[198]:

```
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
```

In [199]:

```
# Calculate the average reading score for 11th grade in each school
junior1 = student_data.loc[student_data['grade']=='11th'].groupby(['sc
junior1
```

Out[199]:

```
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
```

In [200]:

```
# Calculate the average reading score for 12th grade in each school
senior1 = student_data.loc[student_data['grade']=='12th'].groupby(['sc
senior1
```

Out[200]:

```
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
```

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)

In [201]:

```
# 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"]
```

In [202]:

```
# Create a new column to show budget per student in each row
#column is already there
indi_sch
```

Out[202]:

	Type of School	Student Count	Budget of School	Budget Per Student	Average Math Score	Average Reading Score	Math Passing Rate	Reading Passing Rate	(P
Bailey High School	District	4976	3124928.0	628.0	77.048432	81.033963	0.666801	0.819333	0.
Cabrera High School	Charter	1858	1081356.0	582.0	83.061895	83.975780	0.941335	0.970398	0.
Figueroa High School	District	2949	1884411.0	639.0	76.711767	81.158020	0.659885	0.807392	0.
Ford High School	District	2739	1763916.0	644.0	77.102592	80.746258	0.683096	0.792990	0.
Griffin High School	Charter	1468	917500.0	625.0	83.351499	83.816757	0.933924	0.971390	0.
Hernandez High School	District	4635	3022020.0	652.0	77.289752	80.934412	0.667530	0.808630	0.
Holden High School	Charter	427	248087.0	581.0	83.803279	83.814988	0.925059	0.962529	0.
Huang High School	District	2917	1910635.0	655.0	76.629414	81.182722	0.656839	0.813164	0.

Johnson High School	District	4761	3094650.0	650.0	77.072464	80.966394	0.660576	0.812224	0.7
Pena High School	Charter	962	585858.0	609.0	83.839917	84.044699	0.945946	0.959459	0.9
Rodriguez High School	District	3999	2547363.0	637.0	76.842711	80.744686	0.663666	0.802201	0.7
Shelton High School	Charter	1761	1056600.0	600.0	83.359455	83.725724	0.938671	0.958546	0.9
Thomas High School	Charter	1635	1043130.0	638.0	83.418349	83.848930	0.932722	0.973089	0.9
Wilson High School	Charter	2283	1319574.0	578.0	83.274201	83.989488	0.938677	0.965396	0.9
Wright High School	Charter	1800	1049400.0	583.0	83.682222	83.955000	0.933333	0.966111	0.9

In [203]:

```
# Create a new column to define the spending ranges per student
indi_sch['Spending Range Per Student'] = pd.cut(per_stu_budg,spending_
indi_sch
```

Out[203]:

	Type of School	Student Count	Budget of School	Budget Per Student	Average Math Score	Average Reading Score	Math Passing Rate	Reading Passing Rate	(P
Bailey High School	District	4976	3124928.0	628.0	77.048432	81.033963	0.666801	0.819333	0.7
Cabrera High School	Charter	1858	1081356.0	582.0	83.061895	83.975780	0.941335	0.970398	0.9
Figueroa High School	District	2949	1884411.0	639.0	76.711767	81.158020	0.659885	0.807392	0.7
Ford High School	District	2739	1763916.0	644.0	77.102592	80.746258	0.683096	0.792990	0.7
Griffin High School	Charter	1468	917500.0	625.0	83.351499	83.816757	0.933924	0.971390	0.9
Hernandez High	District	4635	3022020.0	652.0	77.289752	80.934412	0.667530	0.808630	0.7

High School	District	4635	3022020.0	652.0	77.289752	80.934412	0.667530	0.808630	0.
Holden High School	Charter	427	248087.0	581.0	83.803279	83.814988	0.925059	0.962529	0.9
Huang High School	District	2917	1910635.0	655.0	76.629414	81.182722	0.656839	0.813164	0.7
Johnson High School	District	4761	3094650.0	650.0	77.072464	80.966394	0.660576	0.812224	0.7
Pena High School	Charter	962	585858.0	609.0	83.839917	84.044699	0.945946	0.959459	0.9
Rodriguez High School	District	3999	2547363.0	637.0	76.842711	80.744686	0.663666	0.802201	0.7
Shelton High School	Charter	1761	1056600.0	600.0	83.359455	83.725724	0.938671	0.958546	0.9
Thomas High School	Charter	1635	1043130.0	638.0	83.418349	83.848930	0.932722	0.973089	0.9
Wilson High School	Charter	2283	1319574.0	578.0	83.274201	83.989488	0.938677	0.965396	0.9
Wright High School	Charter	1800	1049400.0	583.0	83.682222	83.955000	0.933333	0.966111	0.9

In [204]:

```
# Calculate the average math and reading score within each spending ra

avg_math_spend = indi_sch.groupby(['Spending Range Per Student']).mear
avg_read_spend = indi_sch.groupby(['Spending Range Per Student']).mear
```

In [205]:

```
# Calculate the percentage passing rate for math in each spending rang

avg_perc_mathpass = indi_sch.groupby(['Spending Range Per Student']).m
avg_perc_mathpass
```

Out[205]:

```
Spending Range Per Student
<$585       0.934601
$585-615    0.942309
$615-645    0.756682
$645-675    0.661648
Name: Math Passing Rate, dtype: float64
```

In [206]:

```
# Calculate the percentage passing rate for reading in each spending r
avg_perc_readpass = indi_sch.groupby(['Spending Range Per Student']).m

avg_perc_mathpass
```

Out[206]:

```
Spending Range Per Student
<$585       0.934601
$585-615    0.942309
$615-645    0.756682
$645-675    0.661648
Name: Math Passing Rate, dtype: float64
```

In [207]:

```
# Calculate the percentage overall passing rate in each spending range

avg_ovr_pass = (avg_perc_readpass + avg_perc_mathpass) /2
avg_ovr_pass
```

Out[207]:

```
Spending Range Per Student
<$585       0.950355
$585-615    0.950656
$615-645    0.808874
$645-675    0.736494
dtype: float64
```

In [208]:

```
pass_by_spending = pd.DataFrame({'Math Score Average':avg_math_spend,
                               'Reading Score Average': avg_read_spenc
                               '% Passing Math': avg_perc_mathpass,
                               '% Passing Reading': avg_perc_readpass,
                                'Overall Passing Rate': avg_ovr_pass})
pass_by_spending
```

Out[208]:

	Math Score Average	Reading Score Average	% Passing Math	% Passing Reading	Overall Passing Rate
Spending Range Per Student					
<$585	83.455399	83.933814	0.934601	0.966109	0.950355
$585-615	83.599686	83.885211	0.942309	0.959003	0.950656
$615-645	79.079225	81.891436	0.756682	0.861066	0.808874
$645-675	76.997210	81.027843	0.661648	0.811340	0.736494

Scores by School Size

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

In [209]:

```
# Sample bins. Feel free to create your own bins.
size_bins = [0, 1000, 2000, 5000]
group_names = ["Small (<1000)", "Medium (1000-2000)", "Large (2000-500
```

In [229]:
```
# Create a new column for the bin groups
indi_sch['School Size Range']=pd.cut(students_per_sch,size_bins,labels
indi_sch
```

Out[229]:

	Type of School	Student Count	Budget of School	Budget Per Student	Average Math Score	Average Reading Score	Math Passing Rate	Reading Passing Rate	P
Bailey High School	District	4976	3124928.0	628.0	77.048432	81.033963	0.666801	0.819333	0.
Cabrera High School	Charter	1858	1081356.0	582.0	83.061895	83.975780	0.941335	0.970398	0.9
Figueroa High School	District	2949	1884411.0	639.0	76.711767	81.158020	0.659885	0.807392	0.
Ford High School	District	2739	1763916.0	644.0	77.102592	80.746258	0.683096	0.792990	0.
Griffin									

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

In [242]:
```
# math_pass_size
math_pass_size = students_per_sch * pass_mathcount
indi_sch['# of Students passing Math'] = math_pass_size
math_pass_count = indi_sch.groupby(['School Size Range']).sum()['# of
math_pass_count
```

Out[242]:
```
School Size Range
Small (<1000)          1305.0
Medium (1000-2000)     7978.0
Large (2000-5000)     20087.0
Name: # of Students passing Math, dtype: float64
```

In [243]:

```
# read_pass_size
read_pass_size = students_per_sch * pass_readcount
indi_sch['# of Students passing reading'] = read_pass_size
read_pass_count = indi_sch.groupby(['School Size Range']).sum()['# of
read_pass_count
```

Out[243]:

```
School Size Range
Small (<1000)          1334.0
Medium (1000-2000)     8247.0
Large (2000-5000)     24029.0
Name: # of Students passing reading, dtype: float64
```

In [244]:

```
# Calculate the overall passing rate for different school size
size_pass_rate = indi_sch.groupby(['School Size Range']).mean()['Overa
size_pass_rate
```

Out[244]:

```
School Size Range
Small (<1000)         0.948248
Medium (1000-2000)    0.951952
Large (2000-5000)     0.763650
Name: Overall Passing Rate, dtype: float64
```

In [245]:

```
School_Size_Metrics = pd.DataFrame({'# Students Passing Math': math_pa
                                  '# Students Passing Reading': read_p
                                  'Overall Passing Rate':size_pass_rat
School_Size_Metrics
```

Out[245]:

	# Students Passing Math	# Students Passing Reading	Overall Passing Rate
School Size Range			
Small (<1000)	1305.0	1334.0	0.948248
Medium (1000-2000)	7978.0	8247.0	0.951952
Large (2000-5000)	20087.0	24029.0	0.763650

Scores by School Type

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

In [241]:

```
# Create bins and groups, school type {'Charter', 'District'}

name_bins = ['Charter', 'District']
name_values = ["Charter", "District"]
```

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

In [250]:

```
# math pass size

math_type = indi_sch.groupby(['Type of School']).sum()['# of Students
math_type
```

Out[250]:

```
Type of School
Charter     11426.0
District    17944.0
Name: # of Students passing Math, dtype: float64
```

In [253]:

```
# reading pass size
read_type = indi_sch.groupby(['Type of School']).sum()['# of Students
read_type
```

Out[253]:

```
Type of School
Charter     11785.0
District    21825.0
Name: # of Students passing reading, dtype: float64
```

In [256]:

```
# Calculate the overall passing rate
ovr_passtype = indi_sch.groupby(['Type of School']).mean()['Overall Pa
ovr_passtype
```

Out[256]:

```
Type of School
Charter     0.951037
District    0.736738
Name: Overall Passing Rate, dtype: float64
```

In [258]:

```
school_type_metrics = pd.DataFrame({'# of Students Passing Math': math
                                    '# of Students Passing Reading': re
                                    'Overall Passing Rate': ovr_passtyp
school_type_metrics
```

Out[258]:

	# of Students Passing Math	# of Students Passing Reading	Overall Passing Rate
Type of School			
Charter	11426.0	11785.0	0.951037
District	17944.0	21825.0	0.736738

In [ ]: