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 [1]:
```python
# 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_data_complete.head()

# school_data_complete.count()
```

Out[1]:

	Student ID	student_name	gender	grade	school_name	reading_score	math_score	School ID	type	size	bu
0	0	Paul Bradley	M	9th	Huang High School	66	79	0	District	2917	191
1	1	Victor Smith	M	12th	Huang High School	94	61	0	District	2917	191
2	2	Kevin Rodriguez	M	12th	Huang High School	90	60	0	District	2917	191
3	3	Dr. Richard Scott	M	12th	Huang High School	67	58	0	District	2917	191
4	4	Bonnie Ray	F	9th	Huang High School	97	84	0	District	2917	191

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

```python
In [2]:  # Total number of schools
         school_count = school_data_complete['School ID'].nunique()
         school_count
```

Out[2]: 15

```python
In [3]:  # Total number of students
         student_count = school_data_complete['Student ID'].count()
         student_count
```

Out[3]: 39170

```python
In [4]:  # Total budget
         total_budget = school_data_complete['budget'].sum()
         total_budget
```

Out[4]: 82932329558

```python
In [5]:  # Average math score
         avg_math = school_data_complete['math_score'].mean()
         avg_math
```

Out[5]: 78.98537145774827

```python
In [6]:  # Average reading score
         avg_reading = school_data_complete['reading_score'].mean()
         avg_reading
```

Out[6]: 81.87784018381414

```python
In [7]:  # Overall average score
         overall_avg = (school_data_complete['math_score'].mean() \
             + school_data_complete['reading_score'].mean()) / 2
         overall_avg
```

Out[7]: 80.43160582078121

```python
In [8]:  # Percentage of passing math (70 or greater)
         math_pass = school_data_complete[school_data_complete['math_score'] >= 70] \
             .shape[0] / school_data_complete.shape[0] * 100
         math_pass
```

```
Out[8]:    74.9808526933878

In [9]:    # Percentage of passing reading (70 or greater)
           read_pass = school_data_complete[school_data_complete['reading_score'] >= 70] \
               .shape[0] / school_data_complete.shape[0] * 100
           read_pass

Out[9]:    85.80546336482001

In [10]:   # Overall passing rate
           overall_pass = (math_pass + read_pass) / 2
           overall_pass

Out[10]:   80.39315802910392

In [ ]:    # District Summary
           district_summary = pd.DataFrame({'Number of Schools':[school_count],
                                            'Number of Students':[student_count],
                                            'Total Budget':[total_budget],
                                            'Avg Math Score':[avg_math],
                                            'Avg Reading Score':[avg_reading],
                                            'Overall Avg Score':[overall_avg],
                                            'Math Pass Rate':[math_pass],
                                            'Reading Pass Rate':[read_pass],
                                            'Overall Pass Rate': [overall_pass]})
           district_summary
```

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 [62]:   #  Sort and display the top five schools in overall passing rate
           top_school = school_summary.sort_values('overall_pass_rate', ascending=False) \
               .reset_index()
           top_school.iloc[:, [1, 2, -1]].head(5)
```

Out[62]:

school_name	type	overall_pass_rate

0	Cabrera High School	Charter	95.586652
1	Thomas High School	Charter	95.290520
2	Pena High School	Charter	95.270270
3	Griffin High School	Charter	95.265668
4	Wilson High School	Charter	95.203679

In [26]:
```python
# Calculate total school budget
table0 = school_data_complete.groupby(['school_name', 'budget', 'type']).count().reset_i
    .sort_values('school_name', ascending=True).iloc[:,0:4]
table0
```

Out[26]:

	school_name	budget	type	Student ID
0	Bailey High School	3124928	District	4976
1	Cabrera High School	1081356	Charter	1858
2	Figueroa High School	1884411	District	2949
3	Ford High School	1763916	District	2739
4	Griffin High School	917500	Charter	1468
5	Hernandez High School	3022020	District	4635
6	Holden High School	248087	Charter	427
7	Huang High School	1910635	District	2917
8	Johnson High School	3094650	District	4761
9	Pena High School	585858	Charter	962
10	Rodriguez High School	2547363	District	3999
11	Shelton High School	1056600	Charter	1761
12	Thomas High School	1043130	Charter	1635
13	Wilson High School	1319574	Charter	2283
14	Wright High School	1049400	Charter	1800

In [29]:
```python
# Calculate per student budget
table1 = table0
table1['per_student_budget'] = table1['budget'] / table1['Student ID']
table1
```

Out[29]:

	school_name	budget	type	Student ID	per_student_budget
0	Bailey High School	3124928	District	4976	628.0
1	Cabrera High School	1081356	Charter	1858	582.0
2	Figueroa High School	1884411	District	2949	639.0
3	Ford High School	1763916	District	2739	644.0
4	Griffin High School	917500	Charter	1468	625.0
5	Hernandez High School	3022020	District	4635	652.0
6	Holden High School	248087	Charter	427	581.0
7	Huang High School	1910635	District	2917	655.0
8	Johnson High School	3094650	District	4761	650.0

9	Pena High School	585858	Charter	962	609.0
10	Rodriguez High School	2547363	District	3999	637.0
11	Shelton High School	1056600	Charter	1761	600.0
12	Thomas High School	1043130	Charter	1635	638.0
13	Wilson High School	1319574	Charter	2283	578.0
14	Wright High School	1049400	Charter	1800	583.0

In [14]:
```python
# Calcluate the avg math and reading score
table2 = school_data_complete.groupby(['school_name']).sum(numeric_only=True)\
    .reset_index().sort_values('school_name', ascending=True).iloc[:,[0,2,3]]
table2['avg_reading_score'] = table2['reading_score'] / table1['Student ID']
table2['avg_math_score'] = table2['math_score'] / table1['Student ID']
table2 = table2.iloc[:,[0,3,4]]
table2
```

Out[14]:

	school_name	avg_reading_score	avg_math_score
0	Bailey High School	81.033963	77.048432
1	Cabrera High School	83.975780	83.061895
2	Figueroa High School	81.158020	76.711767
3	Ford High School	80.746258	77.102592
4	Griffin High School	83.816757	83.351499
5	Hernandez High School	80.934412	77.289752
6	Holden High School	83.814988	83.803279
7	Huang High School	81.182722	76.629414
8	Johnson High School	80.966394	77.072464
9	Pena High School	84.044699	83.839917
10	Rodriguez High School	80.744686	76.842711
11	Shelton High School	83.725724	83.359455
12	Thomas High School	83.848930	83.418349
13	Wilson High School	83.989488	83.274201
14	Wright High School	83.955000	83.682222

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

In [15]:
```python
# Find the total counts of math result
table3 = school_data_complete.groupby(['school_name', 'math_score']).count()\
    .reset_index().sort_values('school_name', ascending=True).iloc[:,:]

# Find the counts for math result in each school that pass 70 or higher
table3['math_pass'] = 0
table3.loc[table3['math_score'] >= 70, 'math_pass'] = 1
table3['math_pass_count'] = table3['Student ID'] * table3['math_pass']

table3 = table3.groupby(['school_name']).sum()\
    .reset_index().sort_values('school_name', ascending=True).iloc[:,[0,2,-1]]

# Calculate the math passing rate
table3['math_pass_rate'] = table3['math_pass_count'] / table3['Student ID']\
```

```
    * 100
table3
```

Out[15]:

	school_name	Student ID	math_pass_count	math_pass_rate
0	Bailey High School	4976	3318	66.680064
1	Cabrera High School	1858	1749	94.133477
2	Figueroa High School	2949	1946	65.988471
3	Ford High School	2739	1871	68.309602
4	Griffin High School	1468	1371	93.392371
5	Hernandez High School	4635	3094	66.752967
6	Holden High School	427	395	92.505855
7	Huang High School	2917	1916	65.683922
8	Johnson High School	4761	3145	66.057551
9	Pena High School	962	910	94.594595
10	Rodriguez High School	3999	2654	66.366592
11	Shelton High School	1761	1653	93.867121
12	Thomas High School	1635	1525	93.272171
13	Wilson High School	2283	2143	93.867718
14	Wright High School	1800	1680	93.333333

In [16]:
```python
# Find the total counts of read result
table4 = school_data_complete.groupby(['school_name', 'reading_score']).count()\
    .reset_index().sort_values('school_name', ascending=True).iloc[:,:]

# Find the counts for read result in each school that pass 70 or higher
table4['reading_pass'] = 0
table4.loc[table4['reading_score'] >= 70, 'reading_pass'] = 1
table4['reading_pass_count'] = table4['Student ID'] * table4['reading_pass']

table4 = table4.groupby(['school_name']).sum()\
    .reset_index().sort_values('school_name', ascending=True).iloc[:,[0,2,-1]]

# Calculate the read passing rate
table4['reading_pass_rate'] = table4['reading_pass_count'] / table4['Student ID']\
    * 100
table4
```

Out[16]:

	school_name	Student ID	reading_pass_count	reading_pass_rate
0	Bailey High School	4976	4077	81.933280
1	Cabrera High School	1858	1803	97.039828
2	Figueroa High School	2949	2381	80.739234
3	Ford High School	2739	2172	79.299014
4	Griffin High School	1468	1426	97.138965
5	Hernandez High School	4635	3748	80.862999
6	Holden High School	427	411	96.252927
7	Huang High School	2917	2372	81.316421
8	Johnson High School	4761	3867	81.222432

9	Pena High School	962	923	95.945946
10	Rodriguez High School	3999	3208	80.220055
11	Shelton High School	1761	1688	95.854628
12	Thomas High School	1635	1591	97.308869
13	Wilson High School	2283	2204	96.539641
14	Wright High School	1800	1739	96.611111

In [17]:
```python
# Calculate the overall passing rate (average of the math and reading passing rate)
table5 = table4
table5['math_pass_rate'] = table3['math_pass_rate']
table5['overall_pass_rate'] = (table3['math_pass_rate'] + \
    table5['reading_pass_rate']) / 2
table5.iloc[:,[0, 3, 4, 5]]
```

Out[17]:

	school_name	reading_pass_rate	math_pass_rate	overall_pass_rate
0	Bailey High School	81.933280	66.680064	74.306672
1	Cabrera High School	97.039828	94.133477	95.586652
2	Figueroa High School	80.739234	65.988471	73.363852
3	Ford High School	79.299014	68.309602	73.804308
4	Griffin High School	97.138965	93.392371	95.265668
5	Hernandez High School	80.862999	66.752967	73.807983
6	Holden High School	96.252927	92.505855	94.379391
7	Huang High School	81.316421	65.683922	73.500171
8	Johnson High School	81.222432	66.057551	73.639992
9	Pena High School	95.945946	94.594595	95.270270
10	Rodriguez High School	80.220055	66.366592	73.293323
11	Shelton High School	95.854628	93.867121	94.860875
12	Thomas High School	97.308869	93.272171	95.290520
13	Wilson High School	96.539641	93.867718	95.203679
14	Wright High School	96.611111	93.333333	94.972222

In [36]:
```python
# Merge above tables
table6 = table1
table6 = pd.merge(table1, table2, how="left", on=["school_name", "school_name"])
table6
```

Out[36]:

	school_name	budget	type	Student ID	per_student_budget	avg_reading_score	avg_math_score
0	Bailey High School	3124928	District	4976	628.0	81.033963	77.048432
1	Cabrera High School	1081356	Charter	1858	582.0	83.975780	83.061895
2	Figueroa High School	1884411	District	2949	639.0	81.158020	76.711767
3	Ford High School	1763916	District	2739	644.0	80.746258	77.102592
4	Griffin High School	917500	Charter	1468	625.0	83.816757	83.351499

	school_name		Student ID	type		budget	avg_reading_score	avg_math_score
5	Hernandez High School	3022020	District	4635	652.0	80.934412	77.289752
6	Holden High School	248087	Charter	427	581.0	83.814988	83.803279
7	Huang High School	1910635	District	2917	655.0	81.182722	76.629414
8	Johnson High School	3094650	District	4761	650.0	80.966394	77.072464
9	Pena High School	585858	Charter	962	609.0	84.044699	83.839917
10	Rodriguez High School	2547363	District	3999	637.0	80.744686	76.842711
11	Shelton High School	1056600	Charter	1761	600.0	83.725724	83.359455
12	Thomas High School	1043130	Charter	1635	638.0	83.848930	83.418349
13	Wilson High School	1319574	Charter	2283	578.0	83.989488	83.274201
14	Wright High School	1049400	Charter	1800	583.0	83.955000	83.682222

In [51]:
```python
# School Summary Overview Table
school_summary = pd.merge(table6, table5, how="left", on=["school_name", "school_name"])
school_summary = school_summary.iloc[:, [0, 2, 3, 1, 4, 5, 6, 9, 10, 11]]
school_summary.rename(columns = {'Student ID_x':'students'}, inplace = True)
school_summary
```

Out[51]:

	school_name	type	students	budget	per_student_budget	avg_reading_score	avg_math_score	reading
0	Bailey High School	District	4976	3124928	628.0	81.033963	77.048432	
1	Cabrera High School	Charter	1858	1081356	582.0	83.975780	83.061895	
2	Figueroa High School	District	2949	1884411	639.0	81.158020	76.711767	
3	Ford High School	District	2739	1763916	644.0	80.746258	77.102592	
4	Griffin High School	Charter	1468	917500	625.0	83.816757	83.351499	
5	Hernandez High School	District	4635	3022020	652.0	80.934412	77.289752	
6	Holden High School	Charter	427	248087	581.0	83.814988	83.803279	
7	Huang High School	District	2917	1910635	655.0	81.182722	76.629414	
8	Johnson High School	District	4761	3094650	650.0	80.966394	77.072464	
9	Pena High School	Charter	962	585858	609.0	84.044699	83.839917	
10	Rodriguez High School	District	3999	2547363	637.0	80.744686	76.842711	
11	Shelton High School	Charter	1761	1056600	600.0	83.725724	83.359455	
12	Thomas High	Charter	1635	1043130	638.0	83.848930	83.418349	

					School			
13	Wilson High School	Charter	2283	1319574		578.0	83.989488	83.274201
14	Wright High School	Charter	1800	1049400		583.0	83.955000	83.682222

Bottom Performing Schools (By Passing Rate)

- Sort and display the five worst-performing schools

In [64]:
```python
# Sort and display the worst five schools in overall passing rate
bottom_school = school_summary.sort_values('overall_pass_rate', ascending=True)\
    .reset_index()
bottom_school.iloc[:, [1, 2, -1]].head(5)
```

Out[64]:

	school_name	type	students	budget	per_student_budget	avg_reading_score	avg_math_score	reading
0	Bailey High School	District	4976	3124928	628.0	81.033963	77.048432	
1	Cabrera High School	Charter	1858	1081356	582.0	83.975780	83.061895	
2	Figueroa High School	District	2949	1884411	639.0	81.158020	76.711767	
3	Ford High School	District	2739	1763916	644.0	80.746258	77.102592	
4	Griffin High School	Charter	1468	917500	625.0	83.816757	83.351499	
5	Hernandez High School	District	4635	3022020	652.0	80.934412	77.289752	
6	Holden High School	Charter	427	248087	581.0	83.814988	83.803279	
7	Huang High School	District	2917	1910635	655.0	81.182722	76.629414	
8	Johnson High School	District	4761	3094650	650.0	80.966394	77.072464	
9	Pena High School	Charter	962	585858	609.0	84.044699	83.839917	
10	Rodriguez High School	District	3999	2547363	637.0	80.744686	76.842711	
11	Shelton High School	Charter	1761	1056600	600.0	83.725724	83.359455	
12	Thomas High School	Charter	1635	1043130	638.0	83.848930	83.418349	
13	Wilson High School	Charter	2283	1319574	578.0	83.989488	83.274201	
14	Wright High School	Charter	1800	1049400	583.0	83.955000	83.682222	

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

```python
# Create table that lists the average math score for each school of each grade level.
table910 = pd.merge(ninth_grade_table, tenth_grade_table, how='outer')

table91011 = pd.merge(table910, eleventh_grade_table, how='outer')
table91011

table9101112 = pd.merge(table91011, twelfth_grade_table, how='outer')

math_scores_by_grade = table9101112 \
    .sort_values(['school_name', 'grade'], ascending=True)
math_scores_by_grade.iloc[:, [0,1,4,5,6]]
```

Out[160]:

	school_name	grade	students	avg_math_score	avg_reading_score
15	Bailey High School	10th	1239	76.996772	80.907183
30	Bailey High School	11th	1251	77.515588	80.945643
45	Bailey High School	12th	1028	76.492218	80.912451
0	Bailey High School	9th	1458	77.083676	81.303155
16	Cabrera High School	10th	466	83.154506	84.253219
31	Cabrera High School	11th	482	82.765560	83.788382
46	Cabrera High School	12th	382	83.277487	84.287958
1	Cabrera High School	9th	528	83.094697	83.676136
17	Figueroa High School	10th	763	76.539974	81.408912
32	Figueroa High School	11th	709	76.884344	80.640339
47	Figueroa High School	12th	621	77.151369	81.384863
2	Figueroa High School	9th	856	76.403037	81.198598
18	Ford High School	10th	708	77.672316	81.262712
33	Ford High School	11th	659	76.918058	80.403642
48	Ford High School	12th	539	76.179963	80.662338
3	Ford High School	9th	833	77.361345	80.632653
19	Griffin High School	10th	406	84.229064	83.706897
34	Griffin High School	11th	361	83.842105	84.288089
49	Griffin High School	12th	292	83.356164	84.013699
4	Griffin High School	9th	409	82.044010	83.369193
20	Hernandez High School	10th	1227	77.337408	80.660147
35	Hernandez High School	11th	1088	77.136029	81.396140
50	Hernandez High School	12th	938	77.186567	80.857143

5	Hernandez High School	9th	1382	77.438495	80.866860
21	Holden High School	10th	114	83.429825	83.324561
36	Holden High School	11th	103	85.000000	83.815534
51	Holden High School	12th	83	82.855422	84.698795
6	Holden High School	9th	127	83.787402	83.677165
22	Huang High School	10th	767	75.908735	81.512386
37	Huang High School	11th	721	76.446602	81.417476
52	Huang High School	12th	585	77.225641	80.305983
7	Huang High School	9th	844	77.027251	81.290284
23	Johnson High School	10th	1227	76.691117	80.773431
38	Johnson High School	11th	1198	77.491653	80.616027
53	Johnson High School	12th	936	76.863248	81.227564
8	Johnson High School	9th	1400	77.187857	81.260714
24	Pena High School	10th	250	83.372000	83.612000
39	Pena High School	11th	256	84.328125	84.335938
54	Pena High School	12th	181	84.121547	84.591160
9	Pena High School	9th	275	83.625455	83.807273
25	Rodriguez High School	10th	1040	76.612500	80.629808
40	Rodriguez High School	11th	1006	76.395626	80.864811
55	Rodriguez High School	12th	789	77.690748	80.376426
10	Rodriguez High School	9th	1164	76.859966	80.993127
26	Shelton High School	10th	448	82.917411	83.441964
41	Shelton High School	11th	412	83.383495	84.373786
56	Shelton High School	12th	371	83.778976	82.781671
11	Shelton High School	9th	530	83.420755	84.122642
27	Thomas High School	10th	421	83.087886	84.254157
42	Thomas High School	11th	415	83.498795	83.585542
57	Thomas High School	12th	338	83.497041	83.831361
12	Thomas High School	9th	461	83.590022	83.728850
28	Wilson High School	10th	606	83.724422	84.021452
43	Wilson High School	11th	599	83.195326	83.764608
58	Wilson High School	12th	447	83.035794	84.317673
13	Wilson High School	9th	631	83.085578	83.939778
29	Wright High School	10th	486	84.010288	83.812757
44	Wright High School	11th	435	83.836782	84.156322
59	Wright High School	12th	369	83.644986	84.073171
14	Wright High School	9th	510	83.264706	83.833333

```
In [95]:  # Calculate the average math score for 9th grade in each school
```

```
all_grades_counts = school_data_complete.groupby(['school_name', 'grade']) \
    .count().reset_index() \
    .sort_values(['school_name','grade'], ascending=True).iloc[:,:]

all_grades_sum_table = school_data_complete.groupby(['school_name', 'grade']) \
    .sum(numeric_only=True).reset_index() \
    .sort_values(['school_name','grade'], ascending=True).iloc[:,[0,1,3,4]]
all_grades_sum_table['students'] = all_grades_counts['Student ID']

ninth_grade_table = all_grades_sum_table[all_grades_sum_table['grade'] == '9th']
ninth_grade_table['avg_math_score'] = ninth_grade_table['math_score'] \
    / ninth_grade_table['students']
ninth_grade_table
```

Out[95]:

	school_name	grade	reading_score	math_score	students	avg_math_score
3	Bailey High School	9th	118540	112388	1458	77.083676
7	Cabrera High School	9th	44181	43874	528	83.094697
11	Figueroa High School	9th	69506	65401	856	76.403037
15	Ford High School	9th	67167	64442	833	77.361345
19	Griffin High School	9th	34098	33556	409	82.044010
23	Hernandez High School	9th	111758	107020	1382	77.438495
27	Holden High School	9th	10627	10641	127	83.787402
31	Huang High School	9th	68609	65011	844	77.027251
35	Johnson High School	9th	113765	108063	1400	77.187857
39	Pena High School	9th	23047	22997	275	83.625455
43	Rodriguez High School	9th	94276	89465	1164	76.859966
47	Shelton High School	9th	44585	44213	530	83.420755
51	Thomas High School	9th	38599	38535	461	83.590022
55	Wilson High School	9th	52966	52427	631	83.085578
59	Wright High School	9th	42755	42465	510	83.264706

In [96]:
```
# Calculate the average math score for 10th grade in each school
tenth_grade_table = all_grades_sum_table[all_grades_sum_table['grade'] == '10th']
tenth_grade_table['avg_math_score'] = tenth_grade_table['math_score'] \
    / tenth_grade_table['students']
tenth_grade_table
```

```
Out[96]:
```

	school_name	grade	reading_score	math_score	students	avg_math_score
0	Bailey High School	10th	100244	95399	1239	76.996772
4	Cabrera High School	10th	39262	38750	466	83.154506
8	Figueroa High School	10th	62115	58400	763	76.539974
12	Ford High School	10th	57534	54992	708	77.672316
16	Griffin High School	10th	33985	34197	406	84.229064
20	Hernandez High School	10th	98970	94893	1227	77.337408
24	Holden High School	10th	9499	9511	114	83.429825
28	Huang High School	10th	62520	58222	767	75.908735
32	Johnson High School	10th	99109	94100	1227	76.691117
36	Pena High School	10th	20903	20843	250	83.372000
40	Rodriguez High School	10th	83855	79677	1040	76.612500
44	Shelton High School	10th	37382	37147	448	82.917411
48	Thomas High School	10th	35471	34980	421	83.087886
52	Wilson High School	10th	50917	50737	606	83.724422
56	Wright High School	10th	40733	40829	486	84.010288

```
In [97]: # Calculate the average math score for 11th grade in each school
         eleventh_grade_table = all_grades_sum_table[all_grades_sum_table['grade'] == '11th']
         eleventh_grade_table['avg_math_score'] = eleventh_grade_table['math_score'] \
             / eleventh_grade_table['students']
         eleventh_grade_table
```

```
/var/folders/my/1fd3hsfs2gj1zs1vyd2xwpcc0000gn/T/ipykernel_49677/3781187322.py:3: Settin
gWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_
guide/indexing.html#returning-a-view-versus-a-copy
  eleventh_grade_table['avg_math_score'] = eleventh_grade_table['math_score'] \
```

```
Out[97]:
```

	school_name	grade	reading_score	math_score	students	avg_math_score
1	Bailey High School	11th	101263	96972	1251	77.515588
5	Cabrera High School	11th	40386	39893	482	82.765560
9	Figueroa High School	11th	57174	54511	709	76.884344
13	Ford High School	11th	52986	50689	659	76.918058
17	Griffin High School	11th	30428	30267	361	83.842105
21	Hernandez High School	11th	88559	83924	1088	77.136029
25	Holden High School	11th	8633	8755	103	85.000000
29	Huang High School	11th	58702	55118	721	76.446602
33	Johnson High School	11th	96578	92835	1198	77.491653
37	Pena High School	11th	21590	21588	256	84.328125
41	Rodriguez High School	11th	81350	76854	1006	76.395626
45	Shelton High School	11th	34762	34354	412	83.383495
49	Thomas High School	11th	34688	34652	415	83.498795

	school_name	grade				
53	Wilson High School	11th	50175	49834	599	83.195326
57	Wright High School	11th	36608	36469	435	83.836782

In [98]:
```python
# Calculate the average math score for 12th grade in each school
twelfth_grade_table = all_grades_sum_table[all_grades_sum_table['grade'] == '12th']
twelfth_grade_table['avg_math_score'] = twelfth_grade_table['math_score'] \
    / twelfth_grade_table['students']
twelfth_grade_table
```

/var/folders/my/1fd3hsfs2gj1zs1vyd2xwpcc0000gn/T/ipykernel_49677/3266250069.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  twelfth_grade_table['avg_math_score'] = twelfth_grade_table['math_score'] \

Out[98]:

	school_name	grade	reading_score	math_score	students	avg_math_score
2	Bailey High School	12th	83178	78634	1028	76.492218
6	Cabrera High School	12th	32198	31812	382	83.277487
10	Figueroa High School	12th	50540	47911	621	77.151369
14	Ford High School	12th	43477	41061	539	76.179963
18	Griffin High School	12th	24532	24340	292	83.356164
22	Hernandez High School	12th	75844	72401	938	77.186567
26	Holden High School	12th	7030	6877	83	82.855422
30	Huang High School	12th	46979	45177	585	77.225641
34	Johnson High School	12th	76029	71944	936	76.863248
38	Pena High School	12th	15311	15226	181	84.121547
42	Rodriguez High School	12th	63417	61298	789	77.690748
46	Shelton High School	12th	30712	31082	371	83.778976
50	Thomas High School	12th	28335	28222	338	83.497041
54	Wilson High School	12th	37690	37117	447	83.035794
58	Wright High School	12th	31023	30865	369	83.644986

Reading Score by Grade

- Perform the same operations as above for reading scores

In [157…
```python
# Create table that lists the average reading score for each school of each grade level.
reading_scores_by_grade = table9101112 \
    .sort_values(['school_name', 'grade'], ascending=True)
reading_scores_by_grade.iloc[:, [0,1,4,5,6]]
```

Out[157]:

	school_name	grade	students	avg_math_score	avg_reading_score
15	Bailey High School	10th	1239	76.996772	80.907183
30	Bailey High School	11th	1251	77.515588	80.945643

45	Bailey High School	12th	1028	76.492218	80.912451
0	Bailey High School	9th	1458	77.083676	81.303155
16	Cabrera High School	10th	466	83.154506	84.253219
31	Cabrera High School	11th	482	82.765560	83.788382
46	Cabrera High School	12th	382	83.277487	84.287958
1	Cabrera High School	9th	528	83.094697	83.676136
17	Figueroa High School	10th	763	76.539974	81.408912
32	Figueroa High School	11th	709	76.884344	80.640339
47	Figueroa High School	12th	621	77.151369	81.384863
2	Figueroa High School	9th	856	76.403037	81.198598
18	Ford High School	10th	708	77.672316	81.262712
33	Ford High School	11th	659	76.918058	80.403642
48	Ford High School	12th	539	76.179963	80.662338
3	Ford High School	9th	833	77.361345	80.632653
19	Griffin High School	10th	406	84.229064	83.706897
34	Griffin High School	11th	361	83.842105	84.288089
49	Griffin High School	12th	292	83.356164	84.013699
4	Griffin High School	9th	409	82.044010	83.369193
20	Hernandez High School	10th	1227	77.337408	80.660147
35	Hernandez High School	11th	1088	77.136029	81.396140
50	Hernandez High School	12th	938	77.186567	80.857143
5	Hernandez High School	9th	1382	77.438495	80.866860
21	Holden High School	10th	114	83.429825	83.324561
36	Holden High School	11th	103	85.000000	83.815534
51	Holden High School	12th	83	82.855422	84.698795
6	Holden High School	9th	127	83.787402	83.677165
22	Huang High School	10th	767	75.908735	81.512386
37	Huang High School	11th	721	76.446602	81.417476
52	Huang High School	12th	585	77.225641	80.305983
7	Huang High School	9th	844	77.027251	81.290284
23	Johnson High School	10th	1227	76.691117	80.773431
38	Johnson High School	11th	1198	77.491653	80.616027
53	Johnson High School	12th	936	76.863248	81.227564
8	Johnson High School	9th	1400	77.187857	81.260714
24	Pena High School	10th	250	83.372000	83.612000
39	Pena High School	11th	256	84.328125	84.335938
54	Pena High School	12th	181	84.121547	84.591160
9	Pena High School	9th	275	83.625455	83.807273
25	Rodriguez High School	10th	1040	76.612500	80.629808

	school_name	grade		students		
40	Rodriguez High School	11th	1006	76.395626	80.864811
55	Rodriguez High School	12th	789	77.690748	80.376426
10	Rodriguez High School	9th	1164	76.859966	80.993127
26	Shelton High School	10th	448	82.917411	83.441964
41	Shelton High School	11th	412	83.383495	84.373786
56	Shelton High School	12th	371	83.778976	82.781671
11	Shelton High School	9th	530	83.420755	84.122642
27	Thomas High School	10th	421	83.087886	84.254157
42	Thomas High School	11th	415	83.498795	83.585542
57	Thomas High School	12th	338	83.497041	83.831361
12	Thomas High School	9th	461	83.590022	83.728850
28	Wilson High School	10th	606	83.724422	84.021452
43	Wilson High School	11th	599	83.195326	83.764608
58	Wilson High School	12th	447	83.035794	84.317673
13	Wilson High School	9th	631	83.085578	83.939778
29	Wright High School	10th	486	84.010288	83.812757
44	Wright High School	11th	435	83.836782	84.156322
59	Wright High School	12th	369	83.644986	84.073171
14	Wright High School	9th	510	83.264706	83.833333

In [101…
```python
# Calculate the average reading score for 9th grade in each school
ninth_grade_table['avg_reading_score'] = ninth_grade_table['reading_score'] \
    / ninth_grade_table['students']
ninth_grade_table
```

```
/var/folders/my/1fd3hsfs2gj1zs1vyd2xwpcc0000gn/T/ipykernel_49677/3788121873.py:2: Settin
gWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_
guide/indexing.html#returning-a-view-versus-a-copy
  ninth_grade_table['avg_reading_score'] = ninth_grade_table['reading_score'] \
```

Out[101]:

	school_name	grade	reading_score	math_score	students	avg_math_score	avg_reading_score
3	Bailey High School	9th	118540	112388	1458	77.083676	81.303155
7	Cabrera High School	9th	44181	43874	528	83.094697	83.676136
11	Figueroa High School	9th	69506	65401	856	76.403037	81.198598
15	Ford High School	9th	67167	64442	833	77.361345	80.632653
19	Griffin High School	9th	34098	33556	409	82.044010	83.369193
23	Hernandez High School	9th	111758	107020	1382	77.438495	80.866860
27	Holden High School	9th	10627	10641	127	83.787402	83.677165
31	Huang High School	9th	68609	65011	844	77.027251	81.290284
35	Johnson High School	9th	113765	108063	1400	77.187857	81.260714
39	Pena High School	9th	23047	22997	275	83.625455	83.807273

43	Rodriguez High School	9th	94276	89465	1164	76.859966	80.993127
47	Shelton High School	9th	44585	44213	530	83.420755	84.122642
51	Thomas High School	9th	38599	38535	461	83.590022	83.728850
55	Wilson High School	9th	52966	52427	631	83.085578	83.939778
59	Wright High School	9th	42755	42465	510	83.264706	83.833333

In [102…

```python
# Calculate the average reading score for 10th grade in each school
tenth_grade_table['avg_reading_score'] = tenth_grade_table['reading_score'] \
    / tenth_grade_table['students']
tenth_grade_table
```

/var/folders/my/1fd3hsfs2gj1zs1vyd2xwpcc0000gn/T/ipykernel_49677/1946921372.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  tenth_grade_table['avg_reading_score'] = tenth_grade_table['reading_score'] \

Out[102]:

	school_name	grade	reading_score	math_score	students	avg_math_score	avg_reading_score
0	Bailey High School	10th	100244	95399	1239	76.996772	80.907183
4	Cabrera High School	10th	39262	38750	466	83.154506	84.253219
8	Figueroa High School	10th	62115	58400	763	76.539974	81.408912
12	Ford High School	10th	57534	54992	708	77.672316	81.262712
16	Griffin High School	10th	33985	34197	406	84.229064	83.706897
20	Hernandez High School	10th	98970	94893	1227	77.337408	80.660147
24	Holden High School	10th	9499	9511	114	83.429825	83.324561
28	Huang High School	10th	62520	58222	767	75.908735	81.512386
32	Johnson High School	10th	99109	94100	1227	76.691117	80.773431
36	Pena High School	10th	20903	20843	250	83.372000	83.612000
40	Rodriguez High School	10th	83855	79677	1040	76.612500	80.629808
44	Shelton High School	10th	37382	37147	448	82.917411	83.441964
48	Thomas High School	10th	35471	34980	421	83.087886	84.254157
52	Wilson High School	10th	50917	50737	606	83.724422	84.021452
56	Wright High School	10th	40733	40829	486	84.010288	83.812757

In [103…

```python
# Calculate the average reading score for 11th grade in each school
eleventh_grade_table['avg_reading_score'] = eleventh_grade_table['reading_score'] \
    / eleventh_grade_table['students']
eleventh_grade_table
```

/var/folders/my/1fd3hsfs2gj1zs1vyd2xwpcc0000gn/T/ipykernel_49677/2662804456.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

Out[103]:

	school_name	grade	reading_score	math_score	students	avg_math_score	avg_reading_score
1	Bailey High School	11th	101263	96972	1251	77.515588	80.945643
5	Cabrera High School	11th	40386	39893	482	82.765560	83.788382
9	Figueroa High School	11th	57174	54511	709	76.884344	80.640339
13	Ford High School	11th	52986	50689	659	76.918058	80.403642
17	Griffin High School	11th	30428	30267	361	83.842105	84.288089
21	Hernandez High School	11th	88559	83924	1088	77.136029	81.396140
25	Holden High School	11th	8633	8755	103	85.000000	83.815534
29	Huang High School	11th	58702	55118	721	76.446602	81.417476
33	Johnson High School	11th	96578	92835	1198	77.491653	80.616027
37	Pena High School	11th	21590	21588	256	84.328125	84.335938
41	Rodriguez High School	11th	81350	76854	1006	76.395626	80.864811
45	Shelton High School	11th	34762	34354	412	83.383495	84.373786
49	Thomas High School	11th	34688	34652	415	83.498795	83.585542
53	Wilson High School	11th	50175	49834	599	83.195326	83.764608
57	Wright High School	11th	36608	36469	435	83.836782	84.156322

In [104…

```
# Calculate the average reading score for 12th grade in each school
twelfth_grade_table['avg_reading_score'] = twelfth_grade_table['reading_score'] \
    / twelfth_grade_table['students']
twelfth_grade_table
```

Out[104]:

	school_name	grade	reading_score	math_score	students	avg_math_score	avg_reading_score
2	Bailey High School	12th	83178	78634	1028	76.492218	80.912451
6	Cabrera High School	12th	32198	31812	382	83.277487	84.287958
10	Figueroa High School	12th	50540	47911	621	77.151369	81.384863
14	Ford High School	12th	43477	41061	539	76.179963	80.662338
18	Griffin High School	12th	24532	24340	292	83.356164	84.013699
22	Hernandez High School	12th	75844	72401	938	77.186567	80.857143
26	Holden High School	12th	7030	6877	83	82.855422	84.698795
30	Huang High School	12th	46979	45177	585	77.225641	80.305983
34	Johnson High School	12th	76029	71944	936	76.863248	81.227564
38	Pena High School	12th	15311	15226	181	84.121547	84.591160

42	Rodriguez High School	12th	63417	61298	789	77.690748	80.376426
46	Shelton High School	12th	30712	31082	371	83.778976	82.781671
50	Thomas High School	12th	28335	28222	338	83.497041	83.831361
54	Wilson High School	12th	37690	37117	447	83.035794	84.317673
58	Wright High School	12th	31023	30865	369	83.644986	84.073171

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 [ ]:   # 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 [ ]:   # Create a new column to show budget per student in each row
```

```
In [ ]:   # Create a new column to define the spending ranges per student
```

```
In [ ]:   # Calculate the average math score within each spending range
```

```
In [ ]:   # Calculate the percentage passing rate for math in each spending range
```

```
In [ ]:   # Calculate the percentage passing rate for reading in each spending range
```

```
In [ ]:   # Calculate the percentage overall passing rate in each spending range
```

Scores by School Size

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

```
In [ ]:   # 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)"]
```

```
In [ ]:   # Create a new column for the bin groups
```

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

```
In [ ]:   # math_pass_size
```

```
In [ ]:   # read_pass_size
```

```
In [ ]:   # Calculate the overall passing rate for different school size
```

Scores by School Type

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

```
In [ ]:   # Create bins and groups, school type {'Charter', 'District'}
```

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

```
In [ ]:   # math pass size
```

```
In [ ]:   # reading pass size
```

```
In [ ]:   # Calculate the overall passing rate
```