

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

# File to Load (Remember to Change These)
schools_complete = "schools_complete.csv"
students_complete = "students_complete.csv"

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

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

# school_data_complete.count()
school_data_complete.count()
```

```
Out[86]: Student ID      39170
          student_name    39170
          gender           39170
          grade            39170
          school_name      39170
          reading_score     39170
          math_score        39170
          School ID         39170
          type              39170
          size              39170
          budget            39170
          dtype: int64
```

In [87]: ► `print(school_data)`

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

In [88]: ► `print(student_data)`

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

[39170 rows x 7 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 [89]: ► # Create a District Summary

```
school_data_complete.info()
school_data_complete.head()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 39170 entries, 0 to 39169
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Student ID      39170 non-null   int64  
 1   student_name    39170 non-null   object  
 2   gender          39170 non-null   object  
 3   grade           39170 non-null   object  
 4   school_name     39170 non-null   object  
 5   reading_score   39170 non-null   int64  
 6   math_score      39170 non-null   int64  
 7   School ID       39170 non-null   int64  
 8   type            39170 non-null   object  
 9   size             39170 non-null   int64  
 10  budget          39170 non-null   int64  
dtypes: int64(6), object(5)
memory usage: 3.6+ MB
```

Out[89]:

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

In [90]: # Total number of schools
total_school=school_data.groupby('type')['school_name'].count()
total_school

Out[90]: type
Charter 8
District 7
Name: school_name, dtype: int64

In [91]: # Total number of student
total_student=school_data_complete.groupby('type')['student_name'].nunique()
total_student

Out[91]: type
Charter 11340
District 23558
Name: student_name, dtype: int64

In [92]: # Total budget
total_budget=school_data.groupby('type')['budget'].sum()
total_budget

Out[92]: type
Charter 7301505
District 17347923
Name: budget, dtype: int64

In [93]: # Average math score
average_math=school_data_complete.groupby('type')['math_score'].mean()
average_math

Out[93]: type
Charter 83.406183
District 76.987026
Name: math_score, dtype: float64

In [94]: # Average reading score
average_reading=school_data_complete.groupby('type')['reading_score'].mean()
average_reading

Out[94]: type
Charter 83.902821
District 80.962485
Name: reading_score, dtype: float64

In [95]: # Overall average score

```
average_score=school_data_complete.mean()
average_score
```

C:\Users\Aungkyaw\AppData\Local\Temp/ipykernel_24736/1699624738.py:4: FutureWarning: The default value of numeric_only in DataFrame.mean is deprecated. In a future version, it will default to False. In addition, specifying 'numeric_only=None' is deprecated. Select only valid columns or specify the value of numeric_only to silence this warning.

```
average_score=school_data_complete.mean()
```

Out[95]:

Student ID	1.958450e+04
reading_score	8.187784e+01
math_score	7.898537e+01
School ID	6.978172e+00
size	3.332957e+03
budget	2.117241e+06
dtype:	float64

In [96]: # Percentage of passing math (70 or greater)

```
Total_students=school_data_complete['student_name'].count()
```

```
Math_70_greater=school_data_complete[school_data_complete['math_score']>=70]
passing_student_math=Math_70_greater['student_name'].count()
```

```
percentage_passing_student_math=(passing_student_math/Total_students)*100
print("Percentage of passing student is ",round(percentage_passing_student_ma
```

Percentage of passing student is 74.98

In [97]: # District Summary

```
summary_district=pd.DataFrame({'Total Students':total_student,
                               "Total school": total_school,
                               "Total budget":total_budget,
                               "Average Math Score":average_math,
                               "Average Reading Score":average_reading})
summary_district
```

Out[97]:

	Total Students	Total school	Total budget	Average Math Score	Average Reading Score
type					
Charter	11340	8	7301505	83.406183	83.902821
District	23558	7	17347923	76.987026	80.962485

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 [98]: # Percentage of student who pass math
Total_students=school_data_complete.groupby('school_name')[['student_name']].co
Math_70_greater=school_data_complete[school_data_complete['math_score']>=70]
passing_school_math=Math_70_greater.groupby('school_name')[['student_name']].co
percentage_passing_school_math=(passing_school_math/Total_students)*100
percentage_passing_school_math.sort_values(ascending=False)
```

```
Out[98]: school_name
Pena High School      94.594595
Cabrera High School   94.133477
Wilson High School    93.867718
Shelton High School   93.867121
Griffin High School   93.392371
Wright High School    93.333333
Thomas High School    93.272171
Holden High School    92.505855
Ford High School      68.309602
Hernandez High School 66.752967
Bailey High School    66.680064
Rodriguez High School 66.366592
Johnson High School   66.057551
Figueroa High School   65.988471
Huang High School     65.683922
Name: student_name, dtype: float64
```

In [99]: school_data_complete.school_name.unique()

Out[99]: array(['Huang High School', 'Figueroa High School', 'Shelton High School', 'Hernandez High School', 'Griffin High School', 'Wilson High School', 'Cabrera High School', 'Bailey High School', 'Holden High School', 'Pena High School', 'Wright High School', 'Rodriguez High School', 'Johnson High School', 'Ford High School', 'Thomas High School'], dtype=object)

In [100]: # School type

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

Out[100]:

Huang High School	District
Figueroa High School	District
Shelton High School	Charter
Hernandez High School	District
Griffin High School	Charter
Wilson High School	Charter
Cabrera High School	Charter
Bailey High School	District
Holden High School	Charter
Pena High School	Charter
Wright High School	Charter
Rodriguez High School	District
Johnson High School	District
Ford High School	District
Thomas High School	Charter

Name: type, dtype: object

In [101]: # Calculate total school budget

```
total_school_budget=school_data.set_index('school_name')['budget']
total_school_budget
```

Out[101]:

Huang High School	1910635
Figueroa High School	1884411
Shelton High School	1056600
Hernandez High School	3022020
Griffin High School	917500
Wilson High School	1319574
Cabrera High School	1081356
Bailey High School	3124928
Holden High School	248087
Pena High School	585858
Wright High School	1049400
Rodriguez High School	2547363
Johnson High School	3094650
Ford High School	1763916
Thomas High School	1043130

Name: budget, dtype: int64

In [102]: # Calculate per student budget

```
per_student_budget = (school_data.set_index('school_name')['budget']/school_data['size']).reset_index()
```

Out[102]: school_name

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

dtype: float64

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

```
avg_math_score=school_data_complete.groupby('school_name')['math_score'].mean()
print('Average math score in school')
print(avg_math_score.sort_values(ascending=False))
avg_reading_score=school_data_complete.groupby('school_name')['reading_score']
print('Average reading score in school')
print(avg_reading_score.sort_values(ascending=False))
```

Average math score in school

school_name	
Pena High School	83.839917
Holden High School	83.803279
Wright High School	83.682222
Thomas High School	83.418349
Shelton High School	83.359455
Griffin High School	83.351499
Wilson High School	83.274201
Cabrera High School	83.061895
Hernandez High School	77.289752
Ford High School	77.102592
Johnson High School	77.072464
Bailey High School	77.048432
Rodriguez High School	76.842711
Figueroa High School	76.711767
Huang High School	76.629414
Name: math_score, dtype: float64	

Average reading score in school

school_name	
Pena High School	84.044699
Wilson High School	83.989488
Cabrera High School	83.975780
Wright High School	83.955000
Thomas High School	83.848930
Griffin High School	83.816757
Holden High School	83.814988
Shelton High School	83.725724
Huang High School	81.182722
Figueroa High School	81.158020
Bailey High School	81.033963
Johnson High School	80.966394
Hernandez High School	80.934412
Ford High School	80.746258
Rodriguez High School	80.744686
Name: reading_score, dtype: float64	

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

In [104]: # Calculate the math passing rate

```
Total_students=school_data_complete.groupby('school_name')['student_name'].count()
Math_70_greater=school_data_complete[school_data_complete['math_score']>=70]
passing_school_math=Math_70_greater.groupby('school_name')['student_name'].count()
percentage_passing_school_math=(passing_school_math/Total_students)*100
percentage_passing_school_math.sort_values(ascending=False)
```

Out[104]: school_name

Pena High School	94.594595
Cabrera High School	94.133477
Wilson High School	93.867718
Shelton High School	93.867121
Griffin High School	93.392371
Wright High School	93.333333
Thomas High School	93.272171
Holden High School	92.505855
Ford High School	68.309602
Hernandez High School	66.752967
Bailey High School	66.680064
Rodriguez High School	66.366592
Johnson High School	66.057551
Figueroa High School	65.988471
Huang High School	65.683922

Name: student_name, dtype: float64

In [105]: # Calculate the read passing rate

```
Total_students=school_data_complete.groupby('school_name')['student_name'].count()
Reading_70_greater=school_data_complete[school_data_complete['reading_score']>=70]
passing_school_reading=Math_70_greater.groupby('school_name')['student_name'].count()
percentage_passing_school_reading=(passing_school_reading/Total_students)*100
percentage_passing_school_reading.sort_values(ascending=False)
```

Out[105]:

school_name	student_name
Pena High School	94.594595
Cabrera High School	94.133477
Wilson High School	93.867718
Shelton High School	93.867121
Griffin High School	93.392371
Wright High School	93.333333
Thomas High School	93.272171
Holden High School	92.505855
Ford High School	68.309602
Hernandez High School	66.752967
Bailey High School	66.680064
Rodriguez High School	66.366592
Johnson High School	66.057551
Figueroa High School	65.988471
Huang High School	65.683922

Name: student_name, dtype: float64

In [106]: # Calculate the overall passing rate (average of the math and reading passing rates)

```
overall_pass=(percentage_passing_school_reading+percentage_passing_school_math)/2
overall_pass.sort_values(ascending=False)
```

Out[106]:

school_name	student_name
Pena High School	94.594595
Cabrera High School	94.133477
Wilson High School	93.867718
Shelton High School	93.867121
Griffin High School	93.392371
Wright High School	93.333333
Thomas High School	93.272171
Holden High School	92.505855
Ford High School	68.309602
Hernandez High School	66.752967
Bailey High School	66.680064
Rodriguez High School	66.366592
Johnson High School	66.057551
Figueroa High School	65.988471
Huang High School	65.683922

Name: student_name, dtype: float64

Bottom Performing Schools (By Passing Rate)

- Sort and display the five worst-performing schools

In [107]: ► # Sort and display the worst five schools in overall passing rate

```
worst_overall_pass=(percentage_passing_school_reading+percentage_passing_scho  
worst_overall_pass.sort_values()
```

Out[107]: school_name

Huang High School	65.683922
Figueroa High School	65.988471
Johnson High School	66.057551
Rodriguez High School	66.366592
Bailey High School	66.680064
Hernandez High School	66.752967
Ford High School	68.309602
Holden High School	92.505855
Thomas High School	93.272171
Wright High School	93.333333
Griffin High School	93.392371
Shelton High School	93.867121
Wilson High School	93.867718
Cabrera High School	94.133477
Pena High School	94.594595

Name: student_name, dtype: float64

In [108]: ► # Sort and display the best five schools in overall passing rate

```
best_overall_pass=(percentage_passing_school_reading+percentage_passing_scho  
best_overall_pass.sort_values(ascending=False).head(5)
```

Out[108]: school_name

Pena High School	94.594595
Cabrera High School	94.133477
Wilson High School	93.867718
Shelton High School	93.867121
Griffin High School	93.392371

Name: student_name, dtype: float64

```
In [109]: summary_school = pd.DataFrame({ "Total Students": Total_students,
                                         "School Type": school_type,
                                         "Total School Budget": total_school_budget,
                                         "Per Student Budget": per_student_budget,
                                         "Average Math Score": avg_math_score,
                                         "Average Reading Score": avg_reading_score,
                                         "% Passing Math": percentage_passing_school_mat,
                                         "% Passing Reading": percentage_passing_school_
                                         "Overall Passing Rate": overall_pass})
summary_school.reset_index()
```

Out[109]:

	school_name	Total Students	School_Type	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math
0	Bailey High School	4976	District	3124928	628.0	77.048432	81.033963	66.680064
1	Cabrera High School	1858	Charter	1081356	582.0	83.061895	83.975780	94.133477
2	Figueroa High School	2949	District	1884411	639.0	76.711767	81.158020	65.988471
3	Ford High School	2739	District	1763916	644.0	77.102592	80.746258	68.309602
4	Griffin High School	1468	Charter	917500	625.0	83.351499	83.816757	93.392371
5	Hernandez High School	4635	District	3022020	652.0	77.289752	80.934412	66.752967
6	Holden High School	427	Charter	248087	581.0	83.803279	83.814988	92.505855
7	Huang High School	2917	District	1910635	655.0	76.629414	81.182722	65.683922
8	Johnson High School	4761	District	3094650	650.0	77.072464	80.966394	66.057551
9	Pena High School	962	Charter	585858	609.0	83.839917	84.044699	94.594595
10	Rodriguez High School	3999	District	2547363	637.0	76.842711	80.744686	66.366592
11	Shelton High School	1761	Charter	1056600	600.0	83.359455	83.725724	93.867121
12	Thomas High School	1635	Charter	1043130	638.0	83.418349	83.848930	93.272171
13	Wilson High School	2283	Charter	1319574	578.0	83.274201	83.989488	93.867718
14	Wright High School	1800	Charter	1049400	583.0	83.682222	83.955000	93.333333

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 [130]: ► school_data_complete.head()

Out[130]:

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

In [131]: ┏ # Create table that lists the average math score for each school of each grade

```
school_data_complete.groupby(['grade','school_name'])['math_score'].mean()
```

Out[131]:

	grade	school_name	math_score
10th	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	
	11th	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	
12th		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	
	9th	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 [132]: # Calculate the average math score for 9th grade in each school

```
ninth_grade=school_data_complete[school_data_complete['grade']=='9th']
average_math_9th=ninth_grade.groupby('school_name')['math_score'].mean()
average_math_9th.sort_values(ascending=False)
```

Out[132]: school_name

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

In [133]: ┏ # Calculate the average math score for 10th grade in each school

```
ten_grade=school_data_complete[school_data_complete['grade']=='10th']
average_math_10th=ten_grade.groupby('school_name')['math_score'].mean()
average_math_10th.sort_values(ascending=False)
```

Out[133]: school_name

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

In [134]: ┏ # Calculate the average math score for 11th grade in each school

```
eleven_grade=school_data_complete[school_data_complete['grade']=='11th']
average_math_11th=eleven_grade.groupby('school_name')['math_score'].mean()
average_math_11th.sort_values(ascending=False)
```

Out[134]: school_name

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

In [135]: ┏ # Calculate the average math score for 12th grade in each school

```
twelve_grade=school_data_complete[school_data_complete['grade']=='12th']
average_math_12th=twelve_grade.groupby('school_name')['math_score'].mean()
average_math_12th.sort_values(ascending=False)
```

Out[135]: school_name

Peña High School	84.121547
Shelton High School	83.778976
Wright High School	83.644986
Thomas High School	83.497041
Griffin High School	83.356164
Cabrera High School	83.277487
Wilson High School	83.035794
Holden High School	82.855422
Rodriguez High School	77.690748
Huang High School	77.225641
Hernandez High School	77.186567
Figueroa High School	77.151369
Johnson High School	76.863248
Bailey High School	76.492218
Ford High School	76.179963

Name: math_score, dtype: float64

In [136]: # Math score summary

```
Average_math_score_grade=pd.DataFrame({ "9th grade": average_math_9th,  
                                         "10th grade":average_math_10th,  
                                         "11th grade": average_math_11th,  
                                         "12th": average_math_12th})  
Average_math_score_grade.reset_index()
```

Out[136]:

	school_name	9th grade	10th grade	11th grade	12th
0	Bailey High School	77.083676	76.996772	77.515588	76.492218
1	Cabrera High School	83.094697	83.154506	82.765560	83.277487
2	Figueroa High School	76.403037	76.539974	76.884344	77.151369
3	Ford High School	77.361345	77.672316	76.918058	76.179963
4	Griffin High School	82.044010	84.229064	83.842105	83.356164
5	Hernandez High School	77.438495	77.337408	77.136029	77.186567
6	Holden High School	83.787402	83.429825	85.000000	82.855422
7	Huang High School	77.027251	75.908735	76.446602	77.225641
8	Johnson High School	77.187857	76.691117	77.491653	76.863248
9	Peña High School	83.625455	83.372000	84.328125	84.121547
10	Rodriguez High School	76.859966	76.612500	76.395626	77.690748
11	Shelton High School	83.420755	82.917411	83.383495	83.778976
12	Thomas High School	83.590022	83.087886	83.498795	83.497041
13	Wilson High School	83.085578	83.724422	83.195326	83.035794
14	Wright High School	83.264706	84.010288	83.836782	83.644986

Reading Score by Grade

- Perform the same operations as above for reading scores

In [137]: # Create table that lists the average reading score for each school of each grade

```
school_data_complete.groupby(['grade','school_name'])['reading_score'].mean()
```

Out[137]:

grade	school_name	reading_score	
10th	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	
	11th	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	
12th		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	
	9th	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 [138]: ► *# Calculate the average reading score for 9th grade in each school*

```
ninth_grade=school_data_complete[school_data_complete['grade']=='9th']
average_reading_9th=ninth_grade.groupby('school_name')['reading_score'].mean()
average_reading_9th.sort_values(ascending=False)
```

Out[138]: school_name

```
Shelton High School      84.122642
Wilson High School       83.939778
Wright High School       83.833333
Pena High School         83.807273
Thomas High School        83.728850
Holden High School        83.677165
Cabrera High School      83.676136
Griffin High School      83.369193
Bailey High School        81.303155
Huang High School         81.290284
Johnson High School       81.260714
Figueroa High School      81.198598
Rodriguez High School     80.993127
Hernandez High School     80.866860
Ford High School          80.632653
Name: reading_score, dtype: float64
```

In [139]: ┏ # Calculate the average reading score for 10th grade in each school

```
ten_grade=school_data_complete[school_data_complete['grade']=='10th']
average_reading_10th=ten_grade.groupby('school_name')['reading_score'].mean()
average_reading_10th.sort_values(ascending=False)
```

Out[139]: school_name

Thomas High School	84.254157
Cabrera High School	84.253219
Wilson High School	84.021452
Wright High School	83.812757
Griffin High School	83.706897
Pena High School	83.612000
Shelton High School	83.441964
Holden High School	83.324561
Huang High School	81.512386
Figueroa High School	81.408912
Ford High School	81.262712
Bailey High School	80.907183
Johnson High School	80.773431
Hernandez High School	80.660147
Rodriguez High School	80.629808

Name: reading_score, dtype: float64

In [140]: ┏ # Calculate the average reading score for 11th grade in each school

```
eleven_grade=school_data_complete[school_data_complete['grade']=='11th']
average_reading_11th=eleven_grade.groupby('school_name')['reading_score'].mean()
average_reading_11th.sort_values(ascending=False)
```

Out[140]: school_name

Shelton High School	84.373786
Pena High School	84.335938
Griffin High School	84.288089
Wright High School	84.156322
Holden High School	83.815534
Cabrera High School	83.788382
Wilson High School	83.764608
Thomas High School	83.585542
Huang High School	81.417476
Hernandez High School	81.396140
Bailey High School	80.945643
Rodriguez High School	80.864811
Figueroa High School	80.640339
Johnson High School	80.616027
Ford High School	80.403642

Name: reading_score, dtype: float64

In [141]: # Calculate the average reading score for 12th grade in each school

```
twelve_grade=school_data_complete[school_data_complete['grade']=='12th']
average_reading_12th=twelve_grade.groupby('school_name')['reading_score'].mean()
average_reading_12th.sort_values(ascending=False)
```

Out[141]:

school_name	reading_score
Holden High School	84.698795
Pena High School	84.591160
Wilson High School	84.317673
Cabrera High School	84.287958
Wright High School	84.073171
Griffin High School	84.013699
Thomas High School	83.831361
Shelton High School	82.781671
Figueroa High School	81.384863
Johnson High School	81.227564
Bailey High School	80.912451
Hernandez High School	80.857143
Ford High School	80.662338
Rodriguez High School	80.376426
Huang High School	80.305983

Name: reading_score, dtype: float64

In [142]: ► Average_reading_score_grade=pd.DataFrame({ "9th grade": average_reading_9th,
"10th grade":average_reading_10th,
"11th grade": average_reading_11th,
"12th": average_reading_12th})
Average_reading_score_grade.reset_index()

Out[142]:

	school_name	9th grade	10th grade	11th grade	12th
0	Bailey High School	81.303155	80.907183	80.945643	80.912451
1	Cabrera High School	83.676136	84.253219	83.788382	84.287958
2	Figueroa High School	81.198598	81.408912	80.640339	81.384863
3	Ford High School	80.632653	81.262712	80.403642	80.662338
4	Griffin High School	83.369193	83.706897	84.288089	84.013699
5	Hernandez High School	80.866860	80.660147	81.396140	80.857143
6	Holden High School	83.677165	83.324561	83.815534	84.698795
7	Huang High School	81.290284	81.512386	81.417476	80.305983
8	Johnson High School	81.260714	80.773431	80.616027	81.227564
9	Pena High School	83.807273	83.612000	84.335938	84.591160
10	Rodriguez High School	80.993127	80.629808	80.864811	80.376426
11	Shelton High School	84.122642	83.441964	84.373786	82.781671
12	Thomas High School	83.728850	84.254157	83.585542	83.831361
13	Wilson High School	83.939778	84.021452	83.764608	84.317673
14	Wright High School	83.833333	83.812757	84.156322	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 [123]: ► # 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 [124]: school_data_complete.head()

Out[124]:

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



In [125]: school_data.head()

Out[125]:

	School ID	school_name	type	size	budget
0	0	Huang High School	District	2917	1910635
1	1	Figueroa High School	District	2949	1884411
2	2	Shelton High School	Charter	1761	1056600
3	3	Hernandez High School	District	4635	3022020
4	4	Griffin High School	Charter	1468	917500

In [144]: ⏮ summary_school

Out[144]:

	Total Students	School_Type	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	P% Re
school_name								
Bailey High School	4976	District	3124928	628.0	77.048432	81.033963	66.680064	66.6
Cabrera High School	1858	Charter	1081356	582.0	83.061895	83.975780	94.133477	94.1
Figueroa High School	2949	District	1884411	639.0	76.711767	81.158020	65.988471	65.9
Ford High School	2739	District	1763916	644.0	77.102592	80.746258	68.309602	68.3
Griffin High School	1468	Charter	917500	625.0	83.351499	83.816757	93.392371	93.3
Hernandez High School	4635	District	3022020	652.0	77.289752	80.934412	66.752967	66.7
Holden High School	427	Charter	248087	581.0	83.803279	83.814988	92.505855	92.5
Huang High School	2917	District	1910635	655.0	76.629414	81.182722	65.683922	65.6
Johnson High School	4761	District	3094650	650.0	77.072464	80.966394	66.057551	66.0
Pena High School	962	Charter	585858	609.0	83.839917	84.044699	94.594595	94.5
Rodriguez High School	3999	District	2547363	637.0	76.842711	80.744686	66.366592	66.3
Shelton High School	1761	Charter	1056600	600.0	83.359455	83.725724	93.867121	93.8
Thomas High School	1635	Charter	1043130	638.0	83.418349	83.848930	93.272171	93.2
Wilson High School	2283	Charter	1319574	578.0	83.274201	83.989488	93.867718	93.8
Wright High School	1800	Charter	1049400	583.0	83.682222	83.955000	93.333333	93.3



In [145]: # Create a new column to define the spending ranges per student

```
def condition(x):
    if x<585:
        return "<$585"
    elif x>=585 and x<=615:
        return "$585-615"
    elif x>615 and x<=645:
        return "$615-645"
    else:
        return '$645-675'
summary_school['spending ranges per student'] = summary_school["Per Student B
summary_school
```

Out[145]:

	Total Students	School_Type	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math
school_name							
Bailey High School	4976	District	3124928	628.0	77.048432	81.033963	66.680064
Cabrera High School	1858	Charter	1081356	582.0	83.061895	83.975780	94.133477
Figueroa High School	2949	District	1884411	639.0	76.711767	81.158020	65.988471
Ford High School	2739	District	1763916	644.0	77.102592	80.746258	68.309602
Griffin High School	1468	Charter	917500	625.0	83.351499	83.816757	93.392371

In [146]: # Calculate the average math score within each spending range

```
spending_ranges_average_math=summary_school.groupby('spending ranges per stu
```

In [147]: # Calculate the average reading score within each spending range

```
spending_ranges_average_reading=summary_school.groupby('spending ranges per s
```

In [148]: # Calculate the percentage passing rate for math in each spending range

```
spending_ranges_passingrate_math=summary_school.groupby('spending ranges per  
spending_ranges_passingrate_math
```

Out[148]: spending ranges per student
\$585-615 94.230858
\$615-645 75.668212
\$645-675 66.164813
<\$585 93.460096
Name: % Passing Math, dtype: float64

In [149]: # Calculate the percentage passing rate for reading in each spending range

```
spending_ranges_passingrate_reading=summary_school.groupby('spending ranges p  
spending_ranges_passingrate_reading
```

Out[149]: spending ranges per student
\$585-615 94.230858
\$615-645 75.668212
\$645-675 66.164813
<\$585 93.460096
Name: % Passing Reading, dtype: float64

In [150]: # Calculate the percentage overall passing rate in each spending range

```
spending_ranges_passingrate_overall=summary_school.groupby('spending ranges p  
spending_ranges_passingrate_overall
```

Out[150]: spending ranges per student
\$585-615 94.230858
\$615-645 75.668212
\$645-675 66.164813
<\$585 93.460096
Name: Overall Passing Rate, dtype: float64

In [151]:

```
summary_spending=pd.DataFrame({ "average math score":spending_ranges_average_math,
                                 "average Reading score": spending_ranges_average_reading,
                                 "percentage passing rate for math":spending_ranges_passing_rate_math,
                                 "percentage passing rate for reading":spending_ranges_passing_rate_reading,
                                 "percentage overall passing rate":spending_ranges_overall_passing_rate})
```

Out[151]:

spending ranges per student	average math score	average Reading score	percentage passing rate for math	percentage passing rate for reading	percentage overall passing rate
\$585-615	83.599686	83.885211	94.230858	94.230858	94.230858
\$615-645	79.079225	81.891436	75.668212	75.668212	75.668212
\$645-675	76.997210	81.027843	66.164813	66.164813	66.164813
<\$585	83.455399	83.933814	93.460096	93.460096	93.460096

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 []:

```
summary_school
```

In [152]: # Create a new column for the bin groups

```
def condition(x):
    if x<1000:
        return "Small (<1000)"
    elif x>=1000 and x<=2000:
        return "Medium (1000-2000)"
    else:
        return 'Large (2000-5000)'
summary_school['school_size'] = summary_school["Total Students"].apply(condition)
summary_school
```

52]:

	Total Students	School Type	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	Overall Pass Rate
school_name									
Bailey High School	4976	District	3124928	628.0	77.048432	81.033963	66.680064	66.680064	66.680064
Cabrera High School	1858	Charter	1081356	582.0	83.061895	83.975780	94.133477	94.133477	94.133477
Figueroa High School	2949	District	1884411	639.0	76.711767	81.158020	65.988471	65.988471	65.988471
Ford High School	2739	District	1763916	644.0	77.102592	80.746258	68.309602	68.309602	68.309602
Griffin High School	1468	Charter	917500	625.0	83.351499	83.816757	93.392371	93.392371	93.392371

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

In [159]: # math_pass_size

```
math_pass_size=summary_school.groupby('school_size')[ '% Passing Math'].mean()
```

In [161]: # read_pass_size

```
read_pass_size=summary_school.groupby('school_size')[ '% Passing Reading'].mean()
```

In [157]: # Calculate the overall passing rate for different school size

```
overall_passing_rate=summary_school.groupby('school_size')[ 'Overall Passing R
```

In [167]: ► score_by_school_size=pd.DataFrame({ "% Passing Math":math_pass_size,
" % Passing Reading":read_pass_size,
" %Overall Passing":overall_passing_rate})
score_by_school_size

Out[167]:

	% Passing Math	% Passing Reading	%Overall Passing
school_size			
Large (2000-5000)	69.963361	69.963361	69.963361
Medium (1000-2000)	93.599695	93.599695	93.599695
Small (<1000)	93.550225	93.550225	93.550225

Scores by School Type

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

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

In [170]: ► # math pass size

```
math_pass_type=summary_school.groupby('School_Type')[ '% Passing Math'].mean()
```

In [178]: ► # reading pass size

```
reading_pass_type=summary_school.groupby('School_Type')[ '% Passing Reading'].
```

In [177]: ► # Calculate the overall passing rate

```
overall_pass_type=summary_school.groupby('School_Type')[ 'Overall Passing Rate'].
```

```
In [181]: ┆ scores_by_school_type=pd.DataFrame({ "% Passing Math":math_pass_type,  
                                         "% Passing Reading":reading_pass_type,  
                                         "Overall Pass":overall_pass_type})  
scores_by_school_type
```

Out[181]:

	% Passing Math	% Passing Reading	Overall Pass
--	----------------	-------------------	--------------

School_Type			
Charter	93.620830	93.620830	93.620830
District	66.548453	66.548453	66.548453