

Python_Lab_Assignment

October 2, 2022

1 PyCity Schools Analysis

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

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

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

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

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

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

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

# school_data_complete.count()
```

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

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

1.2 District Summary

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

```
[171]: # Create a District Summary
District_Summary = pd.DataFrame({"Total Schools": [total_schools], "Total_
↪Students": [total_students], "Total Budget": [total_budget],
                                "Average Math Score": [avg_math], "Average_
↪Reading Score": [avg_reading], "% Passing Math": [perc_math_pass],
                                "% Overall Passing Rate": [overall_avg]})
District_Summary
```

```
[171]: Total Schools  Total Students  Total Budget  Average Math Score \
0             15         39170         24649428         78.985371

      Average Reading Score  % Passing Math  % Overall Passing Rate
0             81.87784         72.392137         80.431606
```

```
[172]: # Total number of schools
total_schools = school_data["school_name"].count()
total_schools
```

[172]: 15

```
[173]: # Total number of students
total_students = student_data["student_name"].count()
total_students
```

[173]: 39170

```
[174]: # Total budget
total_budget = school_data["budget"].sum()
total_budget
```

[174]: 24649428

```
[175]: # Average math score
avg_math = student_data[("math_score")].mean()
avg_math
```

[175]: 78.98537145774827

```
[176]: # Average reading score
avg_reading = student_data[("reading_score")].mean()
avg_reading
```

[176]: 81.87784018381414

```
[177]: # Overall average score
overall_avg = ((avg_math + avg_reading) / 2)
overall_avg
```

[177]: 80.43160582078121

```
[178]: # Percentage of passing math (70 or greater)
perc_math_pass = (school_data_complete["math_score"]>70).sum() /
↪ school_data_complete["math_score"].count()* 100
perc_math_pass
```

[178]: 72.39213683941792

1.3 School Summary

- Create an overview table that summarizes key metrics about each school, including:
 - School Name
 - School Type
 - Total Students

- Total School Budget
- Per Student Budget
- Average Math Score
- Average Reading Score
- % Passing Math
- % Passing Reading
- Overall Passing Rate (Average of the above two)
- Create a dataframe to hold the above results

1.3.1 Top Performing Schools (By Passing Rate)

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

```
[179]: # Sort and display the top five schools in overall passing rate
schools_count = len(school_data_complete["school_name"].unique())
schools_types = school_data.set_index(["school_name"])["type"]
student_perc_school = school_data_complete["school_name"].value_counts()
student_perc_school.head()
```

```
[179]: Bailey High School      4976
Johnson High School      4761
Hernandez High School    4635
Rodriguez High School     3999
Figueroa High School      2949
Name: school_name, dtype: int64
```

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

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

```
[181]: # Calculate per student budget
perc_student_budget = school_budget / student_perc_school
perc_student_budget
```

```
[181]: 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

```
[182]: # Calculate the avg math and reading score
avg_math_perSchool = school_data_complete.
    ↳groupby(["school_name"])["math_score"].mean()
avg_math_perSchool.head()
```

```
[182]: school_name
Bailey High School      77.048432
Cabrera High School     83.061895
Figueroa High School    76.711767
Ford High School        77.102592
Griffin High School     83.351499
Name: math_score, dtype: float64
```

```
[183]: avg_red_perSchool = school_data_complete.
    ↳groupby(["school_name"])["reading_score"].mean()
avg_red_perSchool.head()
```

```
[183]: school_name
Bailey High School      81.033963
Cabrera High School     83.975780
Figueroa High School    81.158020
Ford High School        80.746258
Griffin High School     83.816757
Name: reading_score, dtype: float64
```

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

```
[184]: passing_math_score = school_data_complete.
    ↳loc[school_data_complete["math_score"]>70]
group_math_school = passing_math_score["school_name"].value_counts()

percent_math = group_math_school / student_perc_school*100
percent_math
```

```
[184]: Bailey High School      64.630225
Cabrera High School         89.558665
Figueroa High School        63.750424
```

Ford High School	65.753925
Griffin High School	89.713896
Hernandez High School	64.746494
Holden High School	90.632319
Huang High School	63.318478
Johnson High School	63.852132
Pena High School	91.683992
Rodriguez High School	64.066017
Shelton High School	89.892107
Thomas High School	90.214067
Wilson High School	90.932983
Wright High School	90.277778

Name: school_name, dtype: float64

```
[185]: read_score = school_data_complete.loc[school_data_complete["reading_score"]>=70]
by_read_per_school = read_score["school_name"].value_counts()
percent_read = by_read_per_school / student_perc_school*100
percent_read
```

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

```
[186]: # Calculate the overall passing rate (average of the math and reading passing_
↪rate)
overall = percent_math + percent_read / student_perc_school
overall
```

```
[186]: Bailey High School      64.646691
Cabrera High School      89.610893
Figueroa High School     63.777802
Ford High School         65.782877
Griffin High School      89.780067
Hernandez High School    64.763940
```

Holden High School	90.857735
Huang High School	63.346355
Johnson High School	63.869192
Pena High School	91.783728
Rodriguez High School	64.086077
Shelton High School	89.946539
Thomas High School	90.273583
Wilson High School	90.975269
Wright High School	90.331451

Name: school_name, dtype: float64

1.3.2 Bottom Performing Schools (By Passing Rate)

- Sort and display the five worst-performing schools

```
[187]: # Sort and display the worst five schools in overall passing rate
school_summary_data = pd.DataFrame({"School Type":schools_types, "Total_
↳Students":student_perc_school, "Total School Budget":school_budget
, "Per Student Budget":perc_student_budget_
↳, "Average Math Score" :avg_math_perSchool,
"Average Reading Score":avg_red_perSchool,
"% Passing Math":percent_math, "% Passing Reading":
↳percent_read,
"% Overall Passing Rate":overall})
bottom_school_data_overall = school_summary_data.sort_values("% Overall Passing_
↳Rate")
bottom_school_data_overall.head()
```

```
[187]:
```

	School Type	Total Students	Total School Budget \
Huang High School	District	2917	1910635.0
Figueroa High School	District	2949	1884411.0
Johnson High School	District	4761	3094650.0
Rodriguez High School	District	3999	2547363.0
Bailey High School	District	4976	3124928.0

	Per Student Budget	Average Math Score \
Huang High School	655.0	76.629414
Figueroa High School	639.0	76.711767
Johnson High School	650.0	77.072464
Rodriguez High School	637.0	76.842711
Bailey High School	628.0	77.048432

	Average Reading Score	% Passing Math \
Huang High School	81.182722	63.318478
Figueroa High School	81.158020	63.750424
Johnson High School	80.966394	63.852132
Rodriguez High School	80.744686	64.066017

Bailey High School	81.033963	64.630225
	% Passing Reading	% Overall Passing Rate
Huang High School	81.316421	63.346355
Figueroa High School	80.739234	63.777802
Johnson High School	81.222432	63.869192
Rodriguez High School	80.220055	64.086077
Bailey High School	81.933280	64.646691

1.4 Math Scores by Grade

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

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

#create a DF for Math Scores by Grade
math_by_grade = pd.DataFrame({"9th":math_9th_grade, "10th": math_10th_grade,
        ↪"11th":math_11th_grade, "12th":math_12th_grade })
math_by_grade
```

```
[188]:
```

	9th	10th	11th	12th
school_name				
Bailey High School	77.083676	76.996772	77.515588	76.492218
Cabrera High School	83.094697	83.154506	82.765560	83.277487
Figueroa High School	76.403037	76.539974	76.884344	77.151369
Ford High School	77.361345	77.672316	76.918058	76.179963
Griffin High School	82.044010	84.229064	83.842105	83.356164
Hernandez High School	77.438495	77.337408	77.136029	77.186567
Holden High School	83.787402	83.429825	85.000000	82.855422
Huang High School	77.027251	75.908735	76.446602	77.225641
Johnson High School	77.187857	76.691117	77.491653	76.863248

Pena High School	83.625455	83.372000	84.328125	84.121547
Rodriguez High School	76.859966	76.612500	76.395626	77.690748
Shelton High School	83.420755	82.917411	83.383495	83.778976
Thomas High School	83.590022	83.087886	83.498795	83.497041
Wilson High School	83.085578	83.724422	83.195326	83.035794
Wright High School	83.264706	84.010288	83.836782	83.644986

```
[189]: # Calculate the average math score for 9th grade in each school
math_9th_grade.mean()
```

```
[189]: 80.35161671308565
```

```
[190]: # Calculate the average math score for 10th grade in each school
math_10th_grade.mean()
```

```
[190]: 80.378948279893
```

```
[191]: # Calculate the average math score for 11th grade in each school
math_11th_grade.mean()
```

```
[191]: 80.57587250651258
```

```
[192]: # Calculate the average math score for 12th grade in each school
math_12th_grade.mean()
```

```
[192]: 80.42381140836758
```

1.4.1 Reading Score by Grade

- Perform the same operations as above for reading scores

```
[193]: # Create table that lists the average reading score for each school of each
      ↪ grade level.
read_9th_grade = student_data.loc[student_data["grade"] == "9th" ].
      ↪ groupby(["school_name"])["reading_score"].mean()
read_10th_grade = student_data.loc[student_data["grade"] == "10th" ].
      ↪ groupby(["school_name"])["reading_score"].mean()
read_11th_grade = student_data.loc[student_data["grade"] == "11th" ].
      ↪ groupby(["school_name"])["reading_score"].mean()
read_12th_grade = student_data.loc[student_data["grade"] == "12th" ].
      ↪ groupby(["school_name"])["reading_score"].mean()
reading_by_grade = pd.DataFrame({"9th":read_9th_grade, "10th":read_10th_grade,
      ↪ "11th":read_11th_grade, "12th":read_12th_grade })
reading_by_grade
```

```
[193]:
```

	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

```
[194]: # Calculate the average reading score for 9th grade in each school
read_9th_grade.mean()
```

```
[194]: 82.51331750091344
```

```
[195]: # Calculate the average reading score for 10th grade in each school
read_10th_grade.mean()
```

```
[195]: 82.50543906336743
```

```
[196]: # Calculate the average reading score for 11th grade in each school
read_11th_grade.mean()
```

```
[196]: 82.55948514074527
```

```
[197]: # Calculate the average reading score for 12th grade in each school
read_12th_grade.mean()
```

```
[197]: 82.55481708213321
```

1.5 Scores by School Spending

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

```
[198]: # 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"]
```

```
[199]: # Create a new column to show budget per student in each row
school_summary_data.head()
school_summary_data["Spending Ranges (Per Student)"] = pd.
    ↪cut(school_summary_data["Total School Budget"]/school_summary_data["Total_
    ↪Students"]
    ,spending_bins , labels =_
    ↪group_names)
school_summary_data.head()
scores_by_school = school_summary_data.drop(columns = ["Total Students", "Total_
    ↪School Budget","School Type", "Per Student Budget"])
scores_by_school
```

```
[199]:
```

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

	% Passing Math	% Passing Reading \
Bailey High School	64.630225	81.933280
Cabrera High School	89.558665	97.039828
Figueroa High School	63.750424	80.739234
Ford High School	65.753925	79.299014
Griffin High School	89.713896	97.138965
Hernandez High School	64.746494	80.862999
Holden High School	90.632319	96.252927
Huang High School	63.318478	81.316421
Johnson High School	63.852132	81.222432
Pena High School	91.683992	95.945946
Rodriguez High School	64.066017	80.220055
Shelton High School	89.892107	95.854628
Thomas High School	90.214067	97.308869
Wilson High School	90.932983	96.539641
Wright High School	90.277778	96.611111

	% Overall Passing Rate	Spending Ranges (Per Student)
Bailey High School	64.646691	\$615-645
Cabrera High School	89.610893	<\$585
Figueroa High School	63.777802	\$615-645
Ford High School	65.782877	\$615-645
Griffin High School	89.780067	\$615-645
Hernandez High School	64.763940	\$645-675
Holden High School	90.857735	<\$585
Huang High School	63.346355	\$645-675
Johnson High School	63.869192	\$645-675
Pena High School	91.783728	\$585-615
Rodriguez High School	64.086077	\$615-645
Shelton High School	89.946539	\$585-615
Thomas High School	90.273583	\$615-645
Wilson High School	90.975269	<\$585
Wright High School	90.331451	<\$585

```
[200]: # Create a new column to define the spending ranges per student
by_spending = scores_by_school.groupby(["Spending Ranges (Per Student)"])
by_spending.mean()
```

```
[200]:
```

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

Spending Ranges (Per Student)	% Passing Math	% Passing Reading \
<\$585	90.350436	96.610877
\$585-615	90.788049	95.900287
\$615-645	73.021426	86.106569
\$645-675	63.972368	81.133951

Spending Ranges (Per Student)	% Overall Passing Rate
<\$585	90.443837
\$585-615	90.865133
\$615-645	73.057850
\$645-675	63.993162

```
[201]: # Calculate the average math score within each spending range
by_average_math = scores_by_school[["Spending Ranges (Per Student)", "Average_
↳Math Score"]]
```

```
by_average_math = by_average_math.groupby(["Spending Ranges (Per Student)"]).
    ↪mean()
by_average_math
```

```
[201]:
```

Spending Ranges (Per Student)	Average Math Score
<\$585	83.455399
\$585-615	83.599686
\$615-645	79.079225
\$645-675	76.997210

```
[202]: # Calculate the percentage passing rate for math in each spending range
by_average_math_passing = scores_by_school[["Spending Ranges (Per Student)", "%
    ↪Passing Math"]]
by_average_math_passing = by_average_math_passing.groupby(["Spending Ranges
    ↪(Per Student)"]).mean()
by_average_math_passing
```

```
[202]:
```

Spending Ranges (Per Student)	% Passing Math
<\$585	90.350436
\$585-615	90.788049
\$615-645	73.021426
\$645-675	63.972368

```
[203]: # Calculate the percentage passing rate for reading in each spending range
by_average_reading_passing = scores_by_school[["Spending Ranges (Per Student)",
    ↪"% Passing Reading"]]
by_average_reading_passing = by_average_reading_passing.groupby(["Spending
    ↪Ranges (Per Student)"]).mean()
by_average_reading_passing
```

```
[203]:
```

Spending Ranges (Per Student)	% Passing Reading
<\$585	96.610877
\$585-615	95.900287
\$615-645	86.106569
\$645-675	81.133951

```
[204]: # Calculate the percentage overall passing rate in each spending range
by_average_overall_passing = scores_by_school[["Spending Ranges (Per Student)",
    ↪"% Overall Passing Rate"]]
by_average_overall_passing = by_average_overall_passing.groupby(["Spending
    ↪Ranges (Per Student)"]).mean()
by_average_overall_passing
```

```
[204]:                                     % Overall Passing Rate
Spending Ranges (Per Student)
<$585                                     90.443837
$585-615                                 90.865133
$615-645                                 73.057850
$645-675                                 63.993162
```

1.5.1 Scores by School Size

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

```
[205]: # 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)"]
```

```
[206]: # Create a new column for the bin groups
school_summary_data.head()
scores_by_school["School Size"] = pd.cut(school_summary_data["Total_
↪Students"],size_bins , labels=group_names)
scores_by_school
```

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

```
                                     % Passing Math  % Passing Reading  \
Bailey High School                64.630225                81.933280
Cabrera High School                89.558665                97.039828
Figueroa High School              63.750424                80.739234
Ford High School                  65.753925                79.299014
Griffin High School               89.713896                97.138965
Hernandez High School             64.746494                80.862999
Holden High School                90.632319                96.252927
Huang High School                 63.318478                81.316421
Johnson High School              63.852132                81.222432
```

Pena High School	91.683992	95.945946
Rodriguez High School	64.066017	80.220055
Shelton High School	89.892107	95.854628
Thomas High School	90.214067	97.308869
Wilson High School	90.932983	96.539641
Wright High School	90.277778	96.611111

	% Overall Passing Rate	Spending Ranges (Per Student) \
Bailey High School	64.646691	\$615-645
Cabrera High School	89.610893	<\$585
Figueroa High School	63.777802	\$615-645
Ford High School	65.782877	\$615-645
Griffin High School	89.780067	\$615-645
Hernandez High School	64.763940	\$645-675
Holden High School	90.857735	<\$585
Huang High School	63.346355	\$645-675
Johnson High School	63.869192	\$645-675
Pena High School	91.783728	\$585-615
Rodriguez High School	64.086077	\$615-645
Shelton High School	89.946539	\$585-615
Thomas High School	90.273583	\$615-645
Wilson High School	90.975269	<\$585
Wright High School	90.331451	<\$585

	School Size
Bailey High School	Large (2000-5000)
Cabrera High School	Medium (1000-2000)
Figueroa High School	Large (2000-5000)
Ford High School	Large (2000-5000)
Griffin High School	Medium (1000-2000)
Hernandez High School	Large (2000-5000)
Holden High School	Small (<1000)
Huang High School	Large (2000-5000)
Johnson High School	Large (2000-5000)
Pena High School	Small (<1000)
Rodriguez High School	Large (2000-5000)
Shelton High School	Medium (1000-2000)
Thomas High School	Medium (1000-2000)
Wilson High School	Large (2000-5000)
Wright High School	Medium (1000-2000)

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

```
[207]: math_pass_size = scores_by_school[scores_by_school['% Passing Math']>=70].
        ↳groupby('School Size')['% Passing Math'].count()
        math_pass_size
```

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

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

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

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

```
[209]: % Overall Passing Rate

School Size
Small (<1000)      91.320731
Medium (1000-2000) 89.988507
Large (2000-5000)  67.656025
```

1.5.2 Scores by School Type

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

```
[210]: # Create bins and groups, school type {'Charter', 'District'}
scores_by_type = school_summary_data.drop(columns=["Total Students", "Total_
↳School Budget", "Per Student Budget"])
scores_by_type = scores_by_type.sort_values(["School Type"])
scores_by_type
```

```
[210]: School Type Average Math Score Average Reading Score \
Cabrera High School Charter 83.061895 83.975780
Griffin High School Charter 83.351499 83.816757
Holden High School Charter 83.803279 83.814988
Pena High School Charter 83.839917 84.044699
Shelton High School Charter 83.359455 83.725724
Thomas High School Charter 83.418349 83.848930
Wilson High School Charter 83.274201 83.989488
```


Wright High School	Charter	83.682222	83.955000
Bailey High School	District	77.048432	81.033963
Figueroa High School	District	76.711767	81.158020
Ford High School	District	77.102592	80.746258
Hernandez High School	District	77.289752	80.934412
Huang High School	District	76.629414	81.182722
Johnson High School	District	77.072464	80.966394
Rodriguez High School	District	76.842711	80.744686

	% Passing Math	% Passing Reading \
Cabrera High School	89.558665	97.039828
Griffin High School	89.713896	97.138965
Holden High School	90.632319	96.252927
Pena High School	91.683992	95.945946
Shelton High School	89.892107	95.854628
Thomas High School	90.214067	97.308869
Wilson High School	90.932983	96.539641
Wright High School	90.277778	96.611111
Bailey High School	64.630225	81.933280
Figueroa High School	63.750424	80.739234
Ford High School	65.753925	79.299014
Hernandez High School	64.746494	80.862999
Huang High School	63.318478	81.316421
Johnson High School	63.852132	81.222432
Rodriguez High School	64.066017	80.220055

	% Overall Passing Rate	Spending Ranges (Per Student)
Cabrera High School	89.610893	<\$585
Griffin High School	89.780067	\$615-645
Holden High School	90.857735	<\$585
Pena High School	91.783728	\$585-615
Shelton High School	89.946539	\$585-615
Thomas High School	90.273583	\$615-645
Wilson High School	90.975269	<\$585
Wright High School	90.331451	<\$585
Bailey High School	64.646691	\$615-645
Figueroa High School	63.777802	\$615-645
Ford High School	65.782877	\$615-645
Hernandez High School	64.763940	\$645-675
Huang High School	63.346355	\$645-675
Johnson High School	63.869192	\$645-675
Rodriguez High School	64.086077	\$615-645

```
[211]: by_school_type = scores_by_type.groupby(["School Type"])
by_school_type.mean()
```

```
[211]:
```

	Average Math Score	Average Reading Score	% Passing Math	\
School Type				
Charter	83.473852	83.896421	90.363226	
District	76.956733	80.966636	64.302528	

	% Passing Reading	% Overall Passing Rate
School Type		
Charter	96.586489	90.444908
District	80.799062	64.324705

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

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

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

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

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

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

```
[214]:
```

	% Overall Passing Rate
School Type	
Charter	90.444908
District	64.324705