In [173]:
```
#pwd
```

Out[173]:
```
'C:\\Users\\01064913\\desktop\\pandas_homework'
```

In [174]:
```
#cd desktop
```

```
[WinError 2] The system cannot find the file specified: 'desktop'
C:\Users\01064913\desktop\pandas_homework
```

In [110]:
```
#cd pandas_homework/
```

```
[WinError 2] The system cannot find the file specified: 'pandas_homework/'
C:\Users\01064913\desktop\pandas_homework
```

In [111]:
```
import pandas as pd
```

In [112]:
```
school_df = pd.read_csv("school_complete.csv")
school_df.head()
```

Out[112]:

	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 [113]:
```
student_df = pd.read_csv("student_complete.csv")
student_df.head()
```

Out[113]:

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

In [114]:
```
school_data_complete = pd.merge(student_df, school_df, how ="left",on=["school_name", "school_name"])
```

In [115]:
```
school_data_complete.head()
```

Out[115]:

	Student ID	student_name	gender	grade	school_name	reading_score	math_score	School ID	type	size	budget
0	0	Paul Bradley	M	9th	Huang High School	66	79	0	District	2917	1910635
1	1	Victor Smith	M	12th	Huang High School	94	61	0	District	2917	1910635
2	2	Kevin Rodriguez	M	12th	Huang High School	90	60	0	District	2917	1910635
3	3	Dr. Richard Scott	M	12th	Huang High School	67	58	0	District	2917	1910635
4	4	Bonnie Ray	F	9th	Huang High School	97	84	0	District	2917	1910635

In [116]:
```
Math_pass =school_data_complete["math_score"] >=70
#reading_pass =school_data_complete["reading_score"] >=70
#Math_pass.head()
school_data_complete["Math_pass"]= Math_pass
reading_pass =school_data_complete["reading_score"] >=70
school_data_complete["reading_pass"]= reading_pass
school_data_complete.head()
```

Out[116]:

	Student ID	student_name	gender	grade	school_name	reading_score	math_score	School ID	type	size	budget	Math_pass	reading_pass
0	0	Paul Bradley	M	9th	Huang High School	66	79	0	District	2917	1910635	True	False
1	1	Victor Smith	M	12th	Huang High School	94	61	0	District	2917	1910635	False	True
2	2	Kevin Rodriguez	M	12th	Huang High School	90	60	0	District	2917	1910635	False	True
3	3	Dr. Richard Scott	M	12th	Huang High School	67	58	0	District	2917	1910635	False	False
4	4	Bonnie Ray	F	9th	Huang High School	97	84	0	District	2917	1910635	True	True

In [117]:
```
total_school = print(len(school_data_complete["school_name"].unique()))

total_school = 15
```

```
15
```

In [118]:
```
#total_students = print(len(school_data_complete))
total_students = school_data_complete["Student ID"].count()
total_students
```

Out[118]: 39170

In [119]:
```
#school_data_complete.info()
```

In [120]:
```
total_budget = print(sum(school_data_complete["budget"].unique()))
total_budget = 24649424
```

```
24649428
```

In [121]:
```
average_mathscore = print(school_data_complete["math_score"].mean())
average_mathscore =78.98537145774827
```

```
78.98537145774827
```

In [122]:
```
average_readingscore = print(school_data_complete["reading_score"].mean())
average_readingscore = 81.87784018381414
```

```
81.87784018381414
```

In [123]:
```
math_pass = school_data_complete["math_score"] >=70
#df_mathpass = school_data_complete[school_data_complete["math_score"] > 70]
total_math_pass = math_pass.sum()
#type(total_math_pass)
#average_math_passcore =average_math_passscore.mean()
math_pass_pct = (total_math_pass/total_students)*100
math_pass_pct
```

Out[123]: 74.9808526933878

In [124]:
```
reading_passs = school_data_complete["reading_score"] >=70
total_reading_pass = reading_pass.sum()
reading_pass_pct =(total_reading_pass/total_students)*100
reading_pass_pct
```

Out[124]: 85.80546336482001

In [125]:
```
passing_rate = (math_pass_pct+reading_pass_pct)/2
passing_rate
```

Out[125]: 80.39315802910392

In [126]:
```
District_summary =pd.DataFrame( {"Total Schools":[total_school],
                                 "Total Students":[total_students],
                                 "Total Budget":[total_budget],
                                  "Average Math Score":[average_mathscore],
                                 "Average Reading Score":[average_readingscore],
                                   "% Passing Math":[math_pass_pct],
                                    "% Passing Reading":[reading_pass_pct],
                                     "% Overall Passing Rate":[passing_rate]
                                      })

District_summary

#Distric_summary_df = pd.DataFrame
#school_value = [, "39,170", 24649428, 78.985371, 81.87784,74.980852, 85.805463, 80.393158 ]
#District_summary =pd.DataFrame([school_header,school_value])
#District_summary
```

Out[126]:

	Total Schools	Total Students	Total Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing Rate
0	15	39170	24649424	78.985371	81.87784	74.980853	85.805463	80.393158

In [127]:
```
school_Name = school_data_complete["school_name"].unique()
school_Name
```

Out[127]:
```
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 [128]:
```
school_budget = school_data_complete["budget"].unique()
school_budget
```

Out[128]:
```
array([1910635, 1884411, 1056600, 3022020,  917500, 1319574, 1081356,
       3124928,  248087,  585858, 1049400, 2547363, 3094650, 1763916,
       1043130], dtype=int64)
```

In [129]:
```
mathpass = school_data_complete["math_score"] >=70

#type(mathpass)
```

In [130]:
```
grouped_school=school_data_complete.groupby(["school_name"])

grouped_school.count().head(15)
```

Out[130]:

	Student ID	student_name	gender	grade	reading_score	math_score	School ID	type	size	budget	Math_pass	reading_pass
school_name												
Bailey High School	4976	4976	4976	4976	4976	4976	4976	4976	4976	4976	4976	4976
Cabrera High School	1858	1858	1858	1858	1858	1858	1858	1858	1858	1858	1858	1858
Figueroa High School	2949	2949	2949	2949	2949	2949	2949	2949	2949	2949	2949	2949
Ford High School	2739	2739	2739	2739	2739	2739	2739	2739	2739	2739	2739	2739
Griffin High School	1468	1468	1468	1468	1468	1468	1468	1468	1468	1468	1468	1468
Hernandez High School	4635	4635	4635	4635	4635	4635	4635	4635	4635	4635	4635	4635
Holden High School	427	427	427	427	427	427	427	427	427	427	427	427
Huang High School	2917	2917	2917	2917	2917	2917	2917	2917	2917	2917	2917	2917
Johnson High School	4761	4761	4761	4761	4761	4761	4761	4761	4761	4761	4761	4761
Pena High School	962	962	962	962	962	962	962	962	962	962	962	962
Rodriguez High School	3999	3999	3999	3999	3999	3999	3999	3999	3999	3999	3999	3999
Shelton High School	1761	1761	1761	1761	1761	1761	1761	1761	1761	1761	1761	1761
Thomas High School	1635	1635	1635	1635	1635	1635	1635	1635	1635	1635	1635	1635
Wilson High School	2283	2283	2283	2283	2283	2283	2283	2283	2283	2283	2283	2283
Wright High School	1800	1800	1800	1800	1800	1800	1800	1800	1800	1800	1800	1800

In [131]:
```
total_students = grouped_school["student_name"].count()
total_students.head()
```

Out[131]:
```
school_name
Bailey High School      4976
Cabrera High School     1858
Figueroa High School    2949
Ford High School        2739
Griffin High School     1468
Name: student_name, dtype: int64
```

In [132]:
```
Per_student_budget= school_budget/total_students
Per_student_budget
```

Out[132]:
```
school_name
Bailey High School        383.970056
Cabrera High School      1014.214747
Figueroa High School      358.290946
Ford High School         1103.329682
Griffin High School       625.000000
Hernandez High School     284.697735
Holden High School       2532.449649
Huang High School        1071.281454
Johnson High School        52.108171
Pena High School          609.000000
Rodriguez High School     262.415604
Shelton High School      1446.543441
Thomas High School       1892.752294
Wilson High School        772.630749
Wright High School        579.516667
Name: student_name, dtype: float64
```

In [133]:
```
school_type =grouped_school["type"].unique()

school_type.head()
```

Out[133]:
```
school_name
Bailey High School      [District]
Cabrera High School      [Charter]
Figueroa High School    [District]
Ford High School        [District]
Griffin High School      [Charter]
Name: type, dtype: object
```

In [134]:
```
average_math_score = grouped_school["math_score"].mean()
average_math_score.head()
```

Out[134]:
```
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
```

In [135]:
```
average_reading_score = grouped_school["reading_score"].mean()
average_reading_score.head()
```

Out[135]:
```
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
```

In [136]:
```
Math_pass =  grouped_school["Math_pass"].sum()
Math_pass_pct = (Math_pass/total_students)*100
#Math_pass_pct.head()

reading_pass =  grouped_school["reading_pass"].sum()
reading_pass_pct = (reading_pass/total_students)*100
#reading_pass_pct.head()
```

In [137]:
```
overall_pass_rate = (Math_pass_pct+reading_pass_pct)/2
overall_pass_rate.head()
```

Out[137]:
```
school_name
Bailey High School      74.306672
Cabrera High School     95.586652
Figueroa High School    73.363852
Ford High School        73.804308
Griffin High School     95.265668
dtype: float64
```

In [138]:
```
school_summary1=pd.DataFrame({"School Type":school_type,"Total Students": total_students, "Total Budget":school_budget,
                              "Per Student Budget":Per_student_budget, "Average Math Score":average_math_score,
                             "Average Reading Score": average_reading_score,"% Passing Math":Math_pass_pct, "% Passing Reading":
                              reading_pass_pct,"Overall Passing Rate":overall_pass_rate

                             })

school_summary1.head(15)
```

Out[138]:

	School Type	Total Students	Total Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	Overall Passing Rate
school_name									
Bailey High School	[District]	4976	1910635	383.970056	77.048432	81.033963	66.680064	81.933280	74.306672
Cabrera High School	[Charter]	1858	1884411	1014.214747	83.061895	83.975780	94.133477	97.039828	95.586652
Figueroa High School	[District]	2949	1056600	358.290946	76.711767	81.158020	65.988471	80.739234	73.363852
Ford High School	[District]	2739	3022020	1103.329682	77.102592	80.746258	68.309602	79.299014	73.804308
Griffin High School	[Charter]	1468	917500	625.000000	83.351499	83.816757	93.392371	97.138965	95.265668
Hernandez High School	[District]	4635	1319574	284.697735	77.289752	80.934412	66.752967	80.862999	73.807983
Holden High School	[Charter]	427	1081356	2532.449649	83.803279	83.814988	92.505855	96.252927	94.379391
Huang High School	[District]	2917	3124928	1071.281454	76.629414	81.182722	65.683922	81.316421	73.500171
Johnson High School	[District]	4761	248087	52.108171	77.072464	80.966394	66.057551	81.222432	73.639992
Pena High School	[Charter]	962	585858	609.000000	83.839917	84.044699	94.594595	95.945946	95.270270
Rodriguez High School	[District]	3999	1049400	262.415604	76.842711	80.744686	66.366592	80.220055	73.293323
Shelton High School	[Charter]	1761	2547363	1446.543441	83.359455	83.725724	93.867121	95.854628	94.860875
Thomas High School	[Charter]	1635	3094650	1892.752294	83.418349	83.848930	93.272171	97.308869	95.290520
Wilson High School	[Charter]	2283	1763916	772.630749	83.274201	83.989488	93.867718	96.539641	95.203679
Wright High School	[Charter]	1800	1043130	579.516667	83.682222	83.955000	93.333333	96.611111	94.972222

In [139]:
```
school_summary1.sort_values(by= ["Overall Passing Rate"], inplace = True, ascending = False)
school_summary1.head()
```

Out[139]:

	School Type	Total Students	Total Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	Overall Passing Rate
school_name									
Cabrera High School	[Charter]	1858	1884411	1014.214747	83.061895	83.975780	94.133477	97.039828	95.586652
Thomas High School	[Charter]	1635	3094650	1892.752294	83.418349	83.848930	93.272171	97.308869	95.290520
Pena High School	[Charter]	962	585858	609.000000	83.839917	84.044699	94.594595	95.945946	95.270270
Griffin High School	[Charter]	1468	917500	625.000000	83.351499	83.816757	93.392371	97.138965	95.265668
Wilson High School	[Charter]	2283	1763916	772.630749	83.274201	83.989488	93.867718	96.539641	95.203679

In [140]:
```
school_summary1.sort_values(by= ["Overall Passing Rate"], inplace = True, )
school_summary1.head()
```

Out[140]:

	School Type	Total Students	Total Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	Overall Passing Rate
school_name									
Rodriguez High School	[District]	3999	1049400	262.415604	76.842711	80.744686	66.366592	80.220055	73.293323
Figueroa High School	[District]	2949	1056600	358.290946	76.711767	81.158020	65.988471	80.739234	73.363852
Huang High School	[District]	2917	3124928	1071.281454	76.629414	81.182722	65.683922	81.316421	73.500171
Johnson High School	[District]	4761	248087	52.108171	77.072464	80.966394	66.057551	81.222432	73.639992
Ford High School	[District]	2739	3022020	1103.329682	77.102592	80.746258	68.309602	79.299014	73.804308

In [141]:
```
Ninth_grade = school_data_complete.loc[school_data_complete["grade"]=="9th",:]

Ninth_grade.head()
```

Out[141]:

	Student ID	student_name	gender	grade	school_name	reading_score	math_score	School ID	type	size	budget	Math_pass	reading_pass
0	0	Paul Bradley	M	9th	Huang High School	66	79	0	District	2917	1910635	True	False
4	4	Bonnie Ray	F	9th	Huang High School	97	84	0	District	2917	1910635	True	True
5	5	Bryan Miranda	M	9th	Huang High School	94	94	0	District	2917	1910635	True	True
12	12	Brittney Walker	F	9th	Huang High School	64	79	0	District	2917	1910635	True	False
13	13	William Long	M	9th	Huang High School	71	79	0	District	2917	1910635	True	True

In [142]:
```
tenth_grade = school_data_complete.loc[school_data_complete["grade"]=="10th",:]
```

In [143]:
```
Eleventh_grade = school_data_complete.loc[school_data_complete["grade"]=="11th",:]
```

In [144]:
```
twelve_grade =school_data_complete.loc[school_data_complete["grade"]=="12th",:]
```

In [145]:
```
g9 = Ninth_grade.groupby(["school_name"])
g9_math_average = g9["math_score"].mean()
g9_math_average.head()
```

Out[145]:
```
school_name
Bailey High School      77.083676
Cabrera High School     83.094697
Figueroa High School    76.403037
Ford High School        77.361345
Griffin High School     82.044010
Name: math_score, dtype: float64
```

In [146]:
```
g10 = tenth_grade.groupby(["school_name"])
g10_math_average = g10["math_score"].mean()
g10_math_average.head()
```

Out[146]:
```
school_name
Bailey High School      76.996772
Cabrera High School     83.154506
Figueroa High School    76.539974
Ford High School        77.672316
Griffin High School     84.229064
Name: math_score, dtype: float64
```

In [147]:
```
g11 = tenth_grade.groupby(["school_name"])
g11_math_average = g11["math_score"].mean()
g11_math_average.head()
```

Out[147]:
```
school_name
Bailey High School      76.996772
Cabrera High School     83.154506
Figueroa High School    76.539974
Ford High School        77.672316
Griffin High School     84.229064
Name: math_score, dtype: float64
```

In [148]:
```
g12 = tenth_grade.groupby(["school_name"])
g12_math_average = g12["math_score"].mean()
g12_math_average.head()
```

Out[148]:
```
school_name
Bailey High School      76.996772
Cabrera High School     83.154506
Figueroa High School    76.539974
Ford High School        77.672316
Griffin High School     84.229064
Name: math_score, dtype: float64
```

In [149]:
```
math_score_by_grade = pd.DataFrame({"9th": g9_math_average, "10th":g10_math_average,
                                    "11th":g11_math_average, "12th":g12_math_average })

math_score_by_grade
```

Out[149]:

	9th	10th	11th	12th
school_name				
Bailey High School	77.083676	76.996772	76.996772	76.996772
Cabrera High School	83.094697	83.154506	83.154506	83.154506
Figueroa High School	76.403037	76.539974	76.539974	76.539974
Ford High School	77.361345	77.672316	77.672316	77.672316
Griffin High School	82.044010	84.229064	84.229064	84.229064
Hernandez High School	77.438495	77.337408	77.337408	77.337408
Holden High School	83.787402	83.429825	83.429825	83.429825
Huang High School	77.027251	75.908735	75.908735	75.908735
Johnson High School	77.187857	76.691117	76.691117	76.691117
Pena High School	83.625455	83.372000	83.372000	83.372000
Rodriguez High School	76.859966	76.612500	76.612500	76.612500
Shelton High School	83.420755	82.917411	82.917411	82.917411
Thomas High School	83.590022	83.087886	83.087886	83.087886
Wilson High School	83.085578	83.724422	83.724422	83.724422
Wright High School	83.264706	84.010288	84.010288	84.010288

In [150]:
```
g9_reading_average = g9["reading_score"].mean()
g10_reading_average = g10["reading_score"].mean()
g11_reading_average = g11["reading_score"].mean()
g12_reading_average = g12["reading_score"].mean()
```

In [151]:
```
reading_score_by_grade = pd.DataFrame({"9th": g9_reading_average, "10th":g10_reading_average,
                                    "11th":g11_reading_average, "12th":g12_reading_average })

reading_score_by_grade
```

Out[151]:

	9th	10th	11th	12th
school_name				
Bailey High School	81.303155	80.907183	80.907183	80.907183
Cabrera High School	83.676136	84.253219	84.253219	84.253219
Figueroa High School	81.198598	81.408912	81.408912	81.408912
Ford High School	80.632653	81.262712	81.262712	81.262712
Griffin High School	83.369193	83.706897	83.706897	83.706897
Hernandez High School	80.866860	80.660147	80.660147	80.660147
Holden High School	83.677165	83.324561	83.324561	83.324561
Huang High School	81.290284	81.512386	81.512386	81.512386
Johnson High School	81.260714	80.773431	80.773431	80.773431
Pena High School	83.807273	83.612000	83.612000	83.612000
Rodriguez High School	80.993127	80.629808	80.629808	80.629808
Shelton High School	84.122642	83.441964	83.441964	83.441964
Thomas High School	83.728850	84.254157	84.254157	84.254157
Wilson High School	83.939778	84.021452	84.021452	84.021452
Wright High School	83.833333	83.812757	83.812757	83.812757

In [152]:
```
spending_bins = [0, 300, 600, 900, 1500, 2600,]
group_names = ["<$300", "$300-600", "$600-900", "$900-1500", "$1500-2600"]
```

In [153]:
```
#school_spending = pd.DataFrame({"Average Math Score": average_math_score, "Average Reading Score":average_reading_score,
#                                 "Overall Score":overall_pass_rate, "Per Student Budget":Per_student_budget }).reset_index()
#school_spending.head()
```

In [154]:
```
req_df =school_summary1.loc[:,["Average Math Score","Average Reading Score","% Passing Math","% Passing Reading",
                               "Overall Passing Rate","Per Student Budget" ]].reset_index()

req_df_withoutschool=req_df.drop(columns="school_name")
req_df_withoutschool.head()
```

Out[154]:

	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	Overall Passing Rate	Per Student Budget
0	76.842711	80.744686	66.366592	80.220055	73.293323	262.415604
1	76.711767	81.158020	65.988471	80.739234	73.363852	358.290946
2	76.629414	81.182722	65.683922	81.316421	73.500171	1071.281454
3	77.072464	80.966394	66.057551	81.222432	73.639992	52.108171
4	77.102592	80.746258	68.309602	79.299014	73.804308	1103.329682

In [155]:
```
req_df_withoutschool["Spending Range"] = pd.cut(req_df_withoutschool["Per Student Budget"], spending_bins, labels=group_names)
#spending_range= req_df_withoutschool.groupby("Spending Range")
#spending_range.head()
req_df_withoutschool
```

Out[155]:

	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	Overall Passing Rate	Per Student Budget	Spending Range
0	76.842711	80.744686	66.366592	80.220055	73.293323	262.415604	<$300
1	76.711767	81.158020	65.988471	80.739234	73.363852	358.290946	$300-600
2	76.629414	81.182722	65.683922	81.316421	73.500171	1071.281454	$900-1500
3	77.072464	80.966394	66.057551	81.222432	73.639992	52.108171	<$300
4	77.102592	80.746258	68.309602	79.299014	73.804308	1103.329682	$900-1500
5	77.289752	80.934412	66.752967	80.862999	73.807983	284.697735	<$300
6	77.048432	81.033963	66.680064	81.933280	74.306672	383.970056	$300-600
7	83.803279	83.814988	92.505855	96.252927	94.379391	2532.449649	$1500-2600
8	83.359455	83.725724	93.867121	95.854628	94.860875	1446.543441	$900-1500
9	83.682222	83.955000	93.333333	96.611111	94.972222	579.516667	$300-600
10	83.274201	83.989488	93.867718	96.539641	95.203679	772.630749	$600-900
11	83.351499	83.816757	93.392371	97.138965	95.265668	625.000000	$600-900
12	83.839917	84.044699	94.594595	95.945946	95.270270	609.000000	$600-900
13	83.418349	83.848930	93.272171	97.308869	95.290520	1892.752294	$1500-2600
14	83.061895	83.975780	94.133477	97.039828	95.586652	1014.214747	$900-1500

In [156]:
```
grouped_spending =req_df_withoutschool.groupby("Spending Range")
```

In [157]:
```
grouped_spending.mean()
```

Out[157]:

	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	Overall Passing Rate	Per Student Budget
Spending Range						
<$300	77.068309	80.881831	66.392370	80.768495	73.580433	199.740503
$300-600	79.147474	82.048994	75.333956	86.427875	80.880915	440.592556
$600-900	83.488539	83.950315	93.951561	96.541517	95.246539	668.876916
$900-1500	80.038339	82.407621	80.498530	88.377473	84.438002	1158.842331
$1500-2600	83.610814	83.831959	92.889013	96.780898	94.834955	2212.600971

In [158]:
```
school_size_bins =[0,1000,2000,5000]
school_group_names = ["Small(<1000)", "Medium(1000-2000)", "Large(2000-5000)"]
```

In [159]:
```
#req_df1 =school_summary1.loc[:,["Average Math Score","Average Reading Score","% Passing Math","% Passing Reading",
                                 # "Overall Passing Rate","Total Students" ]].reset_index()

#school_size = school_summary1.drop("School Type","Total Budget","Per Student Budget")
school_size =school_summary1.loc[:,["Average Math Score","Average Reading Score","% Passing Math","% Passing Reading",
                                  "Overall Passing Rate","Total Students" ]].reset_index()
school_size=school_size.drop(columns="school_name")
school_size.head()
```

Out[159]:

	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	Overall Passing Rate	Total Students
0	76.842711	80.744686	66.366592	80.220055	73.293323	3999
1	76.711767	81.158020	65.988471	80.739234	73.363852	2949
2	76.629414	81.182722	65.683922	81.316421	73.500171	2917
3	77.072464	80.966394	66.057551	81.222432	73.639992	4761
4	77.102592	80.746258	68.309602	79.299014	73.804308	2739

In [160]:
```
school_size["Size Range"]=pd.cut(school_size["Total Students"], school_size_bins, labels= school_group_names)
school_size
```

Out[160]:

	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	Overall Passing Rate	Total Students	Size Range
0	76.842711	80.744686	66.366592	80.220055	73.293323	3999	Large(2000-5000)
1	76.711767	81.158020	65.988471	80.739234	73.363852	2949	Large(2000-5000)
2	76.629414	81.182722	65.683922	81.316421	73.500171	2917	Large(2000-5000)
3	77.072464	80.966394	66.057551	81.222432	73.639992	4761	Large(2000-5000)
4	77.102592	80.746258	68.309602	79.299014	73.804308	2739	Large(2000-5000)
5	77.289752	80.934412	66.752967	80.862999	73.807983	4635	Large(2000-5000)
6	77.048432	81.033963	66.680064	81.933280	74.306672	4976	Large(2000-5000)
7	83.803279	83.814988	92.505855	96.252927	94.379391	427	Small(<1000)
8	83.359455	83.725724	93.867121	95.854628	94.860875	1761	Medium(1000-2000)
9	83.682222	83.955000	93.333333	96.611111	94.972222	1800	Medium(1000-2000)
10	83.274201	83.989488	93.867718	96.539641	95.203679	2283	Large(2000-5000)
11	83.351499	83.816757	93.392371	97.138965	95.265668	1468	Medium(1000-2000)
12	83.839917	84.044699	94.594595	95.945946	95.270270	962	Small(<1000)
13	83.418349	83.848930	93.272171	97.308869	95.290520	1635	Medium(1000-2000)
14	83.061895	83.975780	94.133477	97.039828	95.586652	1858	Medium(1000-2000)

In [161]:
```
grouped_school_size = school_size.groupby("Size Range")
```

In [162]:
```
grouped_school_size.mean()
```

Out[162]:

	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	Overall Passing Rate	Total Students
Size Range						
Small(<1000)	83.821598	83.929843	93.550225	96.099437	94.824831	694.500
Medium(1000-2000)	83.374684	83.864438	93.599695	96.790680	95.195187	1704.400
Large(2000-5000)	77.746417	81.344493	69.963361	82.766634	76.364998	3657.375

In [163]:
```
school_type =school_summary1.loc[:,["Average Math Score","Average Reading Score","% Passing Math","% Passing Reading",
                                  "Overall Passing Rate","School Type" ]].reset_index()
school_type=school_type.drop(columns="school_name")
school_type.head()
```

Out[163]:

	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	Overall Passing Rate	School Type
0	76.842711	80.744686	66.366592	80.220055	73.293323	[District]
1	76.711767	81.158020	65.988471	80.739234	73.363852	[District]
2	76.629414	81.182722	65.683922	81.316421	73.500171	[District]
3	77.072464	80.966394	66.057551	81.222432	73.639992	[District]
4	77.102592	80.746258	68.309602	79.299014	73.804308	[District]

In [164]:
```
grouped_schooltype = school_type.groupby("School Type")
```

In [165]:
```
grouped = school_data_complete.groupby("type")
dfNew = pd.DataFrame({"Average Math Score":grouped["math_score"].mean()})
```

In [166]:
```
dfNew["Average Reading Score"] = grouped["reading_score"].mean()
dfNew
```

Out[166]:

	Average Math Score	Average Reading Score
type		
Charter	83.406183	83.902821
District	76.987026	80.962485

In [167]:
```
df_mathpass = school_data_complete[school_data_complete["math_score"] > 70]
dfNew["% Math pass"]=df_mathpass.groupby("type")["Student ID"].count()/school_data_complete.groupby("type")["Student ID"].count() * 10
0
```

In [168]:
```
dfNew
```

Out[168]:

	Average Math Score	Average Reading Score	% Math pass
type			
Charter	83.406183	83.902821	90.282106
District	76.987026	80.962485	64.305308

In [169]:
```
df_readpass = school_data_complete[school_data_complete["reading_score"] > 70]
dfNew["% Reading pass"]=df_readpass.groupby("type")["Student ID"].count()/school_data_complete.groupby("type")["Student ID"].count() *
100
```

In [170]:
```
dfNew
```

Out[170]:

	Average Math Score	Average Reading Score	% Math pass	% Reading pass
type				
Charter	83.406183	83.902821	90.282106	93.152370
District	76.987026	80.962485	64.305308	78.369662

In [171]:
```
dfNew["%overall Pass"]=(dfNew["% Math pass"]+dfNew["% Reading pass"])/2
```

In [172]:
```
dfNew
```

Out[172]:

	Average Math Score	Average Reading Score	% Math pass	% Reading pass	%overall Pass
type					
Charter	83.406183	83.902821	90.282106	93.152370	91.717238
District	76.987026	80.962485	64.305308	78.369662	71.337485