

```
In [248]: # Dependencies and Setup
import pandas as pd
import numpy as np

# File to Load
school_data_to_load = "Resources/schools_complete.csv"
student_data_to_load = "Resources/students_complete.csv"

# Read School and Student Data File and store into Pandas Data Frames
school_df= pd.read_csv(school_data_to_load)
student_df = pd.read_csv(student_data_to_load)
```

```
In [249]: #school_df.head()
```

```
In [250]: #student_df.head()
```

```
In [251]: # Combine the data into a single dataset
complete_data = pd.merge(student_df, school_df, how="left", on=["school_name",
"school_name"])
```

```
In [252]: complete_data.head()
```

Out[252]:

	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	Dis
1	1	Victor Smith	M	12th	Huang High School	94	61	0	Dis
2	2	Kevin Rodriguez	M	12th	Huang High School	90	60	0	Dis
3	3	Dr. Richard Scott	M	12th	Huang High School	67	58	0	Dis
4	4	Bonnie Ray	F	9th	Huang High School	97	84	0	Dis

```
In [253]: # Identify incomplete rows
complete_data.count()
```

```
Out[253]: 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 [254]: #Calculate the number of schools in the dataframe
school_count=len(complete_data["school_name"].unique())
school_count
```

```
Out[254]: 15
```

```
In [255]: #Calculate the total number of students in the dataframe
student_count=len(complete_data["Student ID"].value_counts())
student_count
```

```
Out[255]: 39170
```

```
In [256]: # Calculate budget total for all schools
budget_total = complete_data["budget"].unique().sum()
budget_total
```

```
Out[256]: 24649428
```

```
In [257]: #Calculate the average math score in the dataframe
average_math=complete_data["math_score"].mean()
average_math
```

```
Out[257]: 78.98537145774827
```

```
In [258]: #Calculate the average reading score in the dataframe
average_read=complete_data["reading_score"].mean()
average_read
```

```
Out[258]: 81.87784018381414
```

```
In [259]: #Calculate the % passing math rate
passing_math=complete_data[(complete_data["math_score"]>=70)]
#passing_math.head()
```

```
In [260]: #Calculate the total number of students in the dataframe  
student_count_passing_math=len(passing_math["Student ID"].value_counts())  
student_count_passing_math
```

Out[260]: 29370

```
In [261]: math_per=student_count_passing_math / student_count  
math_per
```

Out[261]: 0.749808526933878

```
In [262]: # Use pd.to_numeric() method to convert the datatype of the reading score column  
complete_data["reading_score"] = pd.to_numeric(complete_data["reading_score"])
```

```
In [263]: complete_data["reading_score"].dtype
```

Out[263]: dtype('int64')

```
In [264]: #Calculate the % passing rate reading  
passing_reading=complete_data[(complete_data["reading_score"]>=70)]  
#passing_reading.head()
```

```
In [265]: #Calculate the total number of students in the dataframe  
student_count_passing=len(passing_reading["Student ID"].value_counts())  
student_count_passing
```

Out[265]: 33610

```
In [266]: reading_per=student_count_passing / student_count  
reading_per
```

Out[266]: 0.8580546336482001

```
In [267]: overall = (math_per + reading_per)/2  
overall
```

Out[267]: 0.8039315802910391

```
In [268]: # Place all of the data found into a summary dataframe
summary_df=pd.DataFrame({"Total Schools":[school_count],
                          "Total Students":[student_count],
                          "Total Budget":[budget_total],
                          "Average Math Score":[average_math],
                          "Average Reading Score":[average_read],
                          "% Passing Math":[math_per],
                          "% Passing Reading":[reading_per],
                          "% Overall Passing Rate":[overall],})

summary_df
```

Out[268]:

	Total Schools	Total Students	Total Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing Rate
0	15	39170	24649428	78.985371	81.87784	0.749809	0.858055	0.803932

```
In [269]: summary_df["Total Students"] = summary_df["Total Students"].map("{:,}".format)

summary_df
```

Out[269]:

	Total Schools	Total Students	Total Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing Rate
0	15	39,170	24649428	78.985371	81.87784	0.749809	0.858055	0.803932

```
In [270]: summary_df["Total Budget"] = summary_df["Total Budget"].map("${:,.2f}".format)

summary_df
```

Out[270]:

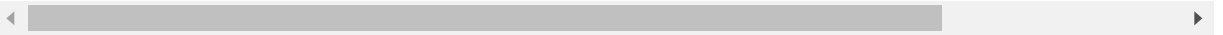
	Total Schools	Total Students	Total Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing Rate
0	15	39,170	\$24,649,428.00	78.985371	81.87784	0.749809	0.858055	0.803932

```
In [271]: # Create a GroupBy object based upon school name  
By_school = complete_data.groupby("school_name")  
By_school.head(2)
```

Out[271]:

	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
2917	2917	Amy Jacobs	F	10th	Figueroa High School	85	87	1
2918	2918	Nathan Campbell	M	12th	Figueroa High School	97	84	1
5866	5866	Jamie Montgomery	F	12th	Shelton High School	70	91	2
5867	5867	Shannon Phillips	F	10th	Shelton High School	84	71	2
7627	7627	Russell Davis	M	10th	Hernandez High School	70	88	3
7628	7628	Timothy Walker	M	12th	Hernandez High School	97	93	3
12262	12262	Heather Wright	F	11th	Griffin High School	79	68	4
12263	12263	Elizabeth Goodwin	F	10th	Griffin High School	91	81	4
13730	13730	Kelli Anderson	F	10th	Wilson High School	84	71	5
13731	13731	Russell Ramirez	M	10th	Wilson High School	72	87	5
16013	16013	Olivia Short	F	11th	Cabrera High School	94	94	6
16014	16014	Kerry Jones	F	9th	Cabrera High School	98	97	6
17871	17871	Blake Martin	M	9th	Bailey High School	75	59	7
17872	17872	Kathryn Kane	F	12th	Bailey High School	84	58	7
22847	22847	Daniel Rodriguez	M	11th	Holden High School	86	92	8
22848	22848	Bryan Perkins	M	9th	Holden High School	91	81	8
23274	23274	Alec Davis	M	9th	Pena High School	91	75	9
23275	23275	Michael Meyer	M	10th	Pena High School	94	76	9
24236	24236	Aaron Johnson	M	10th	Wright High School	89	72	10
24237	24237	Kimberly Hamilton	F	10th	Wright High School	84	93	10
26036	26036	Sherry Jenkins	F	11th	Rodriguez High School	74	81	11

	Student ID	student_name	gender	grade	school_name	reading_score	math_score	School ID
26037	26037	Kimberly Calderon	F	10th	Rodriguez High School	80	86	11
30035	30035	Lisa Casey	F	12th	Johnson High School	87	87	12
30036	30036	Jessica Lopez	F	9th	Johnson High School	98	62	12
34796	34796	Michael Mercado	M	9th	Ford High School	66	94	13
34797	34797	Stephen Wolf	M	11th	Ford High School	68	63	13
37535	37535	Norma Mata	F	10th	Thomas High School	76	76	14
37536	37536	Cody Miller	M	11th	Thomas High School	84	82	14



```
In [272]: # Get the type with the GroupBy object [list]
type = By_school[["type"]].max()
type
```

Out[272]:

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

```
In [273]: # Get the count of students with the GroupBy object [list]
Student_ID = By_school[["Student ID"]].count()
Student_ID
```

Out[273]:

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


```
In [274]: # Get the count of each column with the GroupBy object
School_budget = By_school[["budget"]].mean()
School_budget
```

Out[274]:

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

```
In [275]: # Calculate the per Student budget
per_student_budget = School_budget["budget"] / Student_ID["Student ID"]
per_student_budget
```

Out[275]:

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

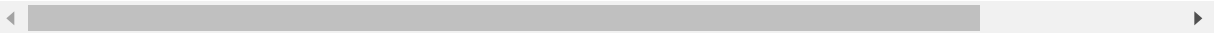
dtype: float64

```
In [276]: # Get the average of each column with the GroupBy object
per_average_math=By_school[["math_score"]].mean()
per_average_reading=By_school[["reading_score"]].mean()
```

```
In [277]: # Filter out passing grades first
passing_math_rate=complete_data[(complete_data["math_score"]>=70)]
passing_math_rate.head()
```

Out[277]:

	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	Dis
4	4	Bonnie Ray	F	9th	Huang High School	97	84	0	Dis
5	5	Bryan Miranda	M	9th	Huang High School	94	94	0	Dis
6	6	Sheena Carter	F	11th	Huang High School	82	80	0	Dis
8	8	Michael Roth	M	10th	Huang High School	95	87	0	Dis



```
In [278]: math_per=passing_math_rate.groupby(["school_name"])["student_name"].count() /
Student_ID["Student ID"]*100
math_per
```

```
Out[278]: school_name
Bailey High School      66.680064
Cabrera High School     94.133477
Figueroa High School    65.988471
Ford High School        68.309602
Griffin High School     93.392371
Hernandez High School   66.752967
Holden High School      92.505855
Huang High School       65.683922
Johnson High School     66.057551
Pena High School        94.594595
Rodriguez High School    66.366592
Shelton High School     93.867121
Thomas High School      93.272171
Wilson High School      93.867718
Wright High School      93.333333
dtype: float64
```

```
In [279]: # Filter out passing reading grades first - then groupby
passing_reading_rate=complete_data[(complete_data["reading_score"]>=70)]
passing_reading_rate.head()
```

Out[279]:

	Student ID	student_name	gender	grade	school_name	reading_score	math_score	School ID	
1	1	Victor Smith	M	12th	Huang High School	94	61	0	Dis
2	2	Kevin Rodriguez	M	12th	Huang High School	90	60	0	Dis
4	4	Bonnie Ray	F	9th	Huang High School	97	84	0	Dis
5	5	Bryan Miranda	M	9th	Huang High School	94	94	0	Dis
6	6	Sheena Carter	F	11th	Huang High School	82	80	0	Dis

```
In [280]: #Calculate the % passing reading rate
reading_per=passing_reading_rate.groupby(["school_name"])["student_name"].count() / Student_ID["Student ID"]*100
reading_per
```

Out[280]:

school_name	
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

dtype: float64

```
In [281]: # Calculate overall passing rate by school
per_overall_passing=(math_per + reading_per)/2
per_overall_passing
```

```
Out[281]: 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
Hernandez High School   73.807983
Holden High School      94.379391
Huang High School       73.500171
Johnson High School     73.639992
Pena High School        95.270270
Rodriguez High School   73.293323
Shelton High School     94.860875
Thomas High School      95.290520
Wilson High School      95.203679
Wright High School      94.972222
dtype: float64
```

```
In [282]: # Place all of the data found into a summary dataframe
school_summary_df=pd.DataFrame({"type":type["type"],
                                "Total Students":Student_ID["Student ID"],
                                "Total School Budget":School_budget["budget"],
                                "Per Student Budget" : per_student_budget,
                                "Average Math Score": per_average_math["math_score"],
                                "Average Reading Score": per_average_reading["reading_
score"],
                                "% Passing Math": math_per,
                                "% Passing Reading": reading_per,
                                "% Overall Passing Rate": per_overall_passing,})
school_summary_df
```

Out[282]:

	type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading
school_name								
Bailey High School	District	4976	3124928	628.0	77.048432	81.033963	66.680064	81.933280
Cabrera High School	Charter	1858	1081356	582.0	83.061895	83.975780	94.133477	97.039828
Figueroa High School	District	2949	1884411	639.0	76.711767	81.158020	65.988471	80.739234
Ford High School	District	2739	1763916	644.0	77.102592	80.746258	68.309602	79.299014
Griffin High School	Charter	1468	917500	625.0	83.351499	83.816757	93.392371	97.138965
Hernandez High School	District	4635	3022020	652.0	77.289752	80.934412	66.752967	80.862999
Holden High School	Charter	427	248087	581.0	83.803279	83.814988	92.505855	96.252927
Huang High School	District	2917	1910635	655.0	76.629414	81.182722	65.683922	81.316421
Johnson High School	District	4761	3094650	650.0	77.072464	80.966394	66.057551	81.222432
Pena High School	Charter	962	585858	609.0	83.839917	84.044699	94.594595	95.945946
Rodriguez High School	District	3999	2547363	637.0	76.842711	80.744686	66.366592	80.220055
Shelton High School	Charter	1761	1056600	600.0	83.359455	83.725724	93.867121	95.854628
Thomas High School	Charter	1635	1043130	638.0	83.418349	83.848930	93.272171	97.308869
Wilson High School	Charter	2283	1319574	578.0	83.274201	83.989488	93.867718	96.539641
Wright High School	Charter	1800	1049400	583.0	83.682222	83.955000	93.333333	96.611111

```
In [283]: #Formatting
school_summary_df["Total School Budget"] = school_summary_df["Total School Budget"].map("${:, .2f}".format)

school_summary_df.head()
```

Out[283]:

	type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	Pass Read
school_name								
Bailey High School	District	4976	\$3,124,928.00	628.0	77.048432	81.033963	66.680064	81.933%
Cabrera High School	Charter	1858	\$1,081,356.00	582.0	83.061895	83.975780	94.133477	97.039%
Figueroa High School	District	2949	\$1,884,411.00	639.0	76.711767	81.158020	65.988471	80.739%
Ford High School	District	2739	\$1,763,916.00	644.0	77.102592	80.746258	68.309602	79.299%
Griffin High School	Charter	1468	\$917,500.00	625.0	83.351499	83.816757	93.392371	97.138%

```
In [284]: school_summary_df["Per Student Budget"] = school_summary_df["Per Student Budget"].map("${:, .2f}".format)

school_summary_df.head()
```

Out[284]:

	type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	Pass Read
school_name								
Bailey High School	District	4976	\$3,124,928.00	\$628.00	77.048432	81.033963	66.680064	81.933%
Cabrera High School	Charter	1858	\$1,081,356.00	\$582.00	83.061895	83.975780	94.133477	97.039%
Figueroa High School	District	2949	\$1,884,411.00	\$639.00	76.711767	81.158020	65.988471	80.739%
Ford High School	District	2739	\$1,763,916.00	\$644.00	77.102592	80.746258	68.309602	79.299%
Griffin High School	Charter	1468	\$917,500.00	\$625.00	83.351499	83.816757	93.392371	97.138%

```
In [285]: # Create table that highlights the top 5 performing schools based on Overall Passing Rate
Top5_Overall_df=school_summary_df.sort_values("% Overall Passing Rate", ascending=False)
Top5_Overall_df.head(5)
```

Out[285]:

	type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	Pass Read
school_name								
Cabrera High School	Charter	1858	\$1,081,356.00	\$582.00	83.061895	83.975780	94.133477	97.0398
Thomas High School	Charter	1635	\$1,043,130.00	\$638.00	83.418349	83.848930	93.272171	97.3088
Pena High School	Charter	962	\$585,858.00	\$609.00	83.839917	84.044699	94.594595	95.9458
Griffin High School	Charter	1468	\$917,500.00	\$625.00	83.351499	83.816757	93.392371	97.1388
Wilson High School	Charter	2283	\$1,319,574.00	\$578.00	83.274201	83.989488	93.867718	96.5398

```
In [286]: # Create table that highlights the bottom 5 performing schools based on Overall Passing Rate
Bottom5_Overall_df=school_summary_df.sort_values("% Overall Passing Rate", ascending=True)
Bottom5_Overall_df.head(5)
```

Out[286]:

	type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	Passing Reading
school_name								
Rodriguez High School	District	3999	\$2,547,363.00	\$637.00	76.842711	80.744686	66.366592	80.2200
Figueroa High School	District	2949	\$1,884,411.00	\$639.00	76.711767	81.158020	65.988471	80.7392
Huang High School	District	2917	\$1,910,635.00	\$655.00	76.629414	81.182722	65.683922	81.3164
Johnson High School	District	4761	\$3,094,650.00	\$650.00	77.072464	80.966394	66.057551	81.2224
Ford High School	District	2739	\$1,763,916.00	\$644.00	77.102592	80.746258	68.309602	79.2990

```
In [287]: #Create a table that lists the average Math 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.

By_school_9th = complete_data.loc[(complete_data['grade'] == '9th')]
By_school_9th.head()
```

Out[287]:

	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	0
4	4	Bonnie Ray	F	9th	Huang High School	97	84	0	0
5	5	Bryan Miranda	M	9th	Huang High School	94	94	0	0
12	12	Brittney Walker	F	9th	Huang High School	64	79	0	0
13	13	William Long	M	9th	Huang High School	71	79	0	0


```
In [288]: # Create a GroupBy object based upon school name to determine averages by grade
By_school_9thm = By_school_9th.groupby("school_name")
By_school_9thm.head()
```

Out[288]:

	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
4	4	Bonnie Ray	F	9th	Huang High School	97	84	0
5	5	Bryan Miranda	M	9th	Huang High School	94	94	0
12	12	Brittney Walker	F	9th	Huang High School	64	79	0
13	13	William Long	M	9th	Huang High School	71	79	0
2920	2920	Jennifer Brown	F	9th	Figueroa High School	97	64	1
2922	2922	Amanda Hamilton DDS	F	9th	Figueroa High School	72	93	1
2925	2925	Natasha Maxwell	F	9th	Figueroa High School	76	77	1
2926	2926	Christine Mann	F	9th	Figueroa High School	71	73	1
2928	2928	Nicole Atkinson	F	9th	Figueroa High School	65	64	1
5871	5871	Earl Martinez	M	9th	Shelton High School	77	86	2
5873	5873	Andrew Shea	M	9th	Shelton High School	71	94	2
5875	5875	Matthew Fowler	M	9th	Shelton High School	74	75	2
5878	5878	Jessica Herrera	F	9th	Shelton High School	93	77	2
5879	5879	Blake Salinas	M	9th	Shelton High School	85	77	2
7635	7635	Natalie Rivera MD	F	9th	Hernandez High School	99	97	3
7636	7636	John Hunt	M	9th	Hernandez High School	97	60	3
7638	7638	Kathryn Horn	F	9th	Hernandez High School	92	81	3
7641	7641	Robert Lambert	M	9th	Hernandez High School	70	95	3
7643	7643	Paul Owens	M	9th	Hernandez High School	67	55	3
12264	12264	Michelle Wong	F	9th	Griffin High School	78	89	4
12270	12270	Thomas Smith	M	9th	Griffin High School	87	79	4
12274	12274	Gregory Long	M	9th	Griffin High School	91	89	4

	Student ID	student_name	gender	grade	school_name	reading_score	math_score	School ID
12279	12279	Evelyn Kirby MD	F	9th	Griffin High School	70	92	4
12281	12281	Mark Yang	M	9th	Griffin High School	76	80	4
13734	13734	Gail Hall	F	9th	Wilson High School	79	72	5
13735	13735	Angela Walters	F	9th	Wilson High School	80	91	5
13739	13739	Tamara Torres	F	9th	Wilson High School	84	76	5
13740	13740	Emma Day	F	9th	Wilson High School	77	80	5
13742	13742	Madison Gregory	F	9th	Wilson High School	91	90	5
...
23274	23274	Alec Davis	M	9th	Pena High School	91	75	9
23279	23279	Jenna Weaver	F	9th	Pena High School	76	84	9
23285	23285	Joshua Johnston	M	9th	Pena High School	94	93	9
23288	23288	Justin Taylor	M	9th	Pena High School	69	92	9
23294	23294	Pamela Riley	F	9th	Pena High School	77	88	9
24245	24245	Eric Espinoza	M	9th	Wright High School	86	74	10
24250	24250	Christopher Newton	M	9th	Wright High School	82	76	10
24256	24256	Sierra Martinez	F	9th	Wright High School	86	86	10
24261	24261	Judy King	F	9th	Wright High School	78	99	10
24278	24278	Jonathan Wilson	M	9th	Wright High School	77	86	10
26044	26044	Brian Smith	M	9th	Rodriguez High School	67	97	11
26046	26046	Gary Orr	M	9th	Rodriguez High School	89	60	11
26047	26047	Melissa Lowery	F	9th	Rodriguez High School	85	92	11
26050	26050	Tiffany Smith	F	9th	Rodriguez High School	91	87	11
26059	26059	Michelle Wall	F	9th	Rodriguez High School	91	70	11

	Student ID	student_name	gender	grade	school_name	reading_score	math_score	School ID
30036	30036	Jessica Lopez	F	9th	Johnson High School	98	62	12
30038	30038	Andrew Smith	M	9th	Johnson High School	66	85	12
30041	30041	Kevin Odonnell	M	9th	Johnson High School	77	73	12
30043	30043	Walter Turner	M	9th	Johnson High School	79	80	12
30051	30051	Daniel Ochoa	M	9th	Johnson High School	67	73	12
34796	34796	Michael Mercado	M	9th	Ford High School	66	94	13
34801	34801	Mr. William Scott	M	9th	Ford High School	73	75	13
34803	34803	Steven Lee	M	9th	Ford High School	77	84	13
34805	34805	Veronica English	F	9th	Ford High School	75	93	13
34807	34807	Christopher Garcia	M	9th	Ford High School	96	93	13
37537	37537	Erik Snyder	M	9th	Thomas High School	80	90	14
37538	37538	Tanya Martinez	F	9th	Thomas High School	71	69	14
37539	37539	Noah Erickson	M	9th	Thomas High School	86	76	14
37540	37540	Austin Meyer	M	9th	Thomas High School	73	96	14
37543	37543	Madison Hampton	F	9th	Thomas High School	82	73	14

75 rows × 11 columns



```
In [289]: # Get the mean of 9th grade math with the GroupBy object
By_school_9th_aver = By_school_9thm[["math_score"]].mean()

By_school_9th_aver
```

Out[289]:

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

```
In [290]: # Get the mean of 10th grade math with the GroupBy object
By_school_10th = complete_data.loc[(complete_data['grade'] == '10th')]
By_school_10thm = By_school_10th.groupby("school_name")
By_school_10th_aver = By_school_10thm[["math_score"]].mean()
By_school_10th_aver
```

Out[290]:

	math_score
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
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

```
In [291]: # Get the mean of 11th grade math with the GroupBy object
By_school_11th = complete_data.loc[(complete_data['grade'] == '11th')]
By_school_11thm = By_school_11th.groupby("school_name")
By_school_11th_aver = By_school_11thm[["math_score"]].mean()
By_school_11th_aver
```

Out[291]:

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

```
In [292]: # Get the mean of 12th grade math with the GroupBy object
By_school_12th = complete_data.loc[(complete_data['grade'] == '12th')]
By_school_12thm = By_school_12th.groupby("school_name")
By_school_12th_aver = By_school_12thm[["math_score"]].mean()
By_school_12th_aver
```

Out[292]:

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


```
In [293]: # Place all of the data for math averages, by grade, by school, into a summary
          # dataframe
          math_school_summary_df=pd.DataFrame({"9th":By_school_9th_aver["math_score"],
          "10th":By_school_10th_aver["math_score"],
          "11th":By_school_11th_aver["math_score"],
          "12th":By_school_12th_aver["math_score"]})
          math_school_summary_df
```

Out[293]:

	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

```
In [294]: # Get the mean of 9th grade reading with the GroupBy object
By_school_9th_aver_r = By_school_9thm[["reading_score"]].mean()
By_school_9th_aver_r
```

Out[294]:

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

```
In [295]: # Get the mean of 10th grade math with the GroupBy object
By_school_10th_aver_r = By_school_10thm[["reading_score"]].mean()
By_school_10th_aver_r
```

Out[295]:

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

```
In [296]: # Get the mean of 11th grade math with the GroupBy object
By_school_11th_aver_r = By_school_11thm[["reading_score"]].mean()
By_school_11th_aver_r
```

Out[296]:

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

```
In [297]: # Get the mean of 12th grade reading with the GroupBy object
By_school_12th_aver_r = By_school_12thm[["reading_score"]].mean()
By_school_12th_aver_r
```

Out[297]:

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

```
In [298]: # Place all of the data for reading averages, by grade, by school, into a summary dataframe
reading_school_summary_df=pd.DataFrame({"9th":By_school_9th_aver_r["reading_score"],
                                         "10th":By_school_10th_aver_r["reading_score"],
                                         "11th":By_school_11th_aver_r["reading_score"],
                                         "12th":By_school_12th_aver_r["reading_score"]})
reading_school_summary_df
```

Out[298]:

	9th	10th	11th	12th
school_name				
Bailey High School	81.303155	80.907183	80.945643	80.912451
Cabrera High School	83.676136	84.253219	83.788382	84.287958
Figueroa High School	81.198598	81.408912	80.640339	81.384863
Ford High School	80.632653	81.262712	80.403642	80.662338
Griffin High School	83.369193	83.706897	84.288089	84.013699
Hernandez High School	80.866860	80.660147	81.396140	80.857143
Holden High School	83.677165	83.324561	83.815534	84.698795
Huang High School	81.290284	81.512386	81.417476	80.305983
Johnson High School	81.260714	80.773431	80.616027	81.227564
Pena High School	83.807273	83.612000	84.335938	84.591160
Rodriguez High School	80.993127	80.629808	80.864811	80.376426
Shelton High School	84.122642	83.441964	84.373786	82.781671
Thomas High School	83.728850	84.254157	83.585542	83.831361
Wilson High School	83.939778	84.021452	83.764608	84.317673
Wright High School	83.833333	83.812757	84.156322	84.073171

```
In [299]: # Place all of the data found into a summary dataframe
school_summary_df_spend=pd.DataFrame({"type":type["type"],
                                     "Total Students":Student_ID["Student ID"],
                                     "Total School Budget":School_budget["budget"],
                                     "Per Student Budget": per_student_budget,
                                     "Average Math Score": per_average_math["math_score"],
                                     "Average Reading Score": per_average_reading["reading_
score"],
                                     "% Passing Math": math_per,
                                     "% Passing Reading": reading_per,
                                     "% Overall Passing Rate": per_overall_passing,})
school_summary_df_spend.head()
```

Out[299]:

	type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading
school_name								
Bailey High School	District	4976	3124928	628.0	77.048432	81.033963	66.680064	81.933280
Cabrera High School	Charter	1858	1081356	582.0	83.061895	83.975780	94.133477	97.039828
Figueroa High School	District	2949	1884411	639.0	76.711767	81.158020	65.988471	80.739234
Ford High School	District	2739	1763916	644.0	77.102592	80.746258	68.309602	79.299014
Griffin High School	Charter	1468	917500	625.0	83.351499	83.816757	93.392371	97.138965

```
In [300]: # Create a table that breaks down school performance 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)

# Spending bins.
spending_bins = [0, 585, 615, 645, 675]
group_names = ["<$585", "$585-615", "$615-645", "$645-675"]
```

```
In [301]: # Slice the data and place it into bins
pd.cut(school_summary_df_spend["Per Student Budget"], spending_bins, labels=group_names).head()
```

```
Out[301]: school_name
Bailey High School    $615-645
Cabrera High School   <$585
Figueroa High School  $615-645
Ford High School      $615-645
Griffin High School   $615-645
Name: Per Student Budget, dtype: category
Categories (4, object): [<$585 < $585-615 < $615-645 < $645-675]
```

```
In [302]: # Place the data series into a new column inside of the DataFrame

school_summary_df_spend["Spending Range"] = pd.cut(school_summary_df_spend["Per Student Budget"], spending_bins, labels=group_names)
school_summary_df_spend.head()
```

Out[302]:

	type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading
school_name								
Bailey High School	District	4976	3124928	628.0	77.048432	81.033963	66.680064	81.933280
Cabrera High School	Charter	1858	1081356	582.0	83.061895	83.975780	94.133477	97.039828
Figueroa High School	District	2949	1884411	639.0	76.711767	81.158020	65.988471	80.739234
Ford High School	District	2739	1763916	644.0	77.102592	80.746258	68.309602	79.299014
Griffin High School	Charter	1468	917500	625.0	83.351499	83.816757	93.392371	97.138965

```
In [303]: # Create a GroupBy object based upon "Size Group"

Spending_group = school_summary_df_spend.groupby("Spending Range")
Spending_group
```

Out[303]: <pandas.core.groupby.generic.DataFrameGroupBy object at 0x000002BAC43F8FD0>

```
In [304]: # Get the average of each column within the GroupBy object

Spending_group[["Average Math Score", "Average Reading Score", "% Passing Math", "% Passing Reading", "% Overall Passing Rate"]].mean()
```

Out[304]:

	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing Rate
Spending Range					
<\$585	83.455399	83.933814	93.460096	96.610877	95.035486
\$585-615	83.599686	83.885211	94.230858	95.900287	95.065572
\$615-645	79.079225	81.891436	75.668212	86.106569	80.887391
\$645-675	76.997210	81.027843	66.164813	81.133951	73.649382


```
In [305]: # Get the size with the GroupBy object [list]
size = By_school[["size"]].max()
size
```

Out[305]:

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

```
In [306]: # Place all of the data found into a summary dataframe
school_summary_df_size=pd.DataFrame({"size":size["size"],
                                     "type":type["type"],
                                     "Total Students":Student_ID["Student ID"],
                                     "Total School Budget":School_budget["budget"],
                                     "Per Student Budget" : per_student_budget,
                                     "Average Math Score": per_average_math["math_score"],
                                     "Average Reading Score": per_average_reading["reading_
score"],
                                     "% Passing Math": math_per,
                                     "% Passing Reading": reading_per,
                                     "% Overall Passing Rate": per_overall_passing,})
school_summary_df_size.head()
```

Out[306]:

	size	type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	Pas Rea
school_name									
Bailey High School	4976	District	4976	3124928	628.0	77.048432	81.033963	66.680064	81.93
Cabrera High School	1858	Charter	1858	1081356	582.0	83.061895	83.975780	94.133477	97.03
Figueroa High School	2949	District	2949	1884411	639.0	76.711767	81.158020	65.988471	80.73
Ford High School	2739	District	2739	1763916	644.0	77.102592	80.746258	68.309602	79.29
Griffin High School	1468	Charter	1468	917500	625.0	83.351499	83.816757	93.392371	97.13

```
In [307]: # Create bins in which to place values based upon size
size_bins = [0, 1000, 2000, 5000]

# Create labels for these bins
group_labels = ["Small (<1000)", "Medium (1000-2000)", "Large (2000-5000)"]

In [308]: # Slice the data and place it into bins
pd.cut(school_summary_df_size["size"], size_bins, labels=group_labels).head()
```

```
Out[308]: school_name
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)
Name: size, dtype: category
Categories (3, object): [Small (<1000) < Medium (1000-2000) < Large (2000-5000)]
```

```
In [309]: # Place the data series into a new column inside of the DataFrame

school_summary_df_size["School Size"] = pd.cut(school_summary_df_size["size"],
size_bins, labels=group_labels)
school_summary_df_size.head()
```

Out[309]:

	size	type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	Pas Rea
school_name									
Bailey High School	4976	District	4976	3124928	628.0	77.048432	81.033963	66.680064	81.93
Cabrera High School	1858	Charter	1858	1081356	582.0	83.061895	83.975780	94.133477	97.03
Figueroa High School	2949	District	2949	1884411	639.0	76.711767	81.158020	65.988471	80.73
Ford High School	2739	District	2739	1763916	644.0	77.102592	80.746258	68.309602	79.29
Griffin High School	1468	Charter	1468	917500	625.0	83.351499	83.816757	93.392371	97.13

```
In [310]: # Create a GroupBy object based upon "Size Group"
Size_group = school_summary_df_size.groupby("School Size")
Size_group
```

Out[310]: <pandas.core.groupby.generic.DataFrameGroupBy object at 0x000002BAC7138E48>

```
In [311]: # Get the average of each column within the GroupBy object
Size_group[["Average Math Score", "Average Reading Score", "% Passing Math",
"% Passing Reading", "% Overall Passing Rate"]].mean()
```

Out[311]:

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

In []:

```
In [312]: # Place all of the data found into a summary dataframe
school_summary_df_type=pd.DataFrame({"size":size["size"],
                                     "type":type["type"],
                                     "Total Students":Student_ID["Student ID"],
                                     "Total School Budget":School_budget["budget"],
                                     "Per Student Budget" : per_student_budget,
                                     "Average Math Score": per_average_math["math_score"],
                                     "Average Reading Score": per_average_reading["reading_
score"],
                                     "% Passing Math": math_per,
                                     "% Passing Reading": reading_per,
                                     "% Overall Passing Rate": per_overall_passing,})
school_summary_df_type.head()
```

Out[312]:

	size	type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	Pas Rea
school_name									
Bailey High School	4976	District	4976	3124928	628.0	77.048432	81.033963	66.680064	81.93
Cabrera High School	1858	Charter	1858	1081356	582.0	83.061895	83.975780	94.133477	97.03
Figueroa High School	2949	District	2949	1884411	639.0	76.711767	81.158020	65.988471	80.73
Ford High School	2739	District	2739	1763916	644.0	77.102592	80.746258	68.309602	79.29
Griffin High School	1468	Charter	1468	917500	625.0	83.351499	83.816757	93.392371	97.13

```
In [313]: # Create a GroupBy object based upon "Size Group"
Type_group = school_summary_df_type.groupby("type")
Type_group
```

Out[313]: <pandas.core.groupby.generic.DataFrameGroupBy object at 0x000002BAC711A5F8>

```
In [314]: # Get the average of each column within the GroupBy object
Type_group[["Average Math Score", "Average Reading Score", "% Passing Math",
"% Passing Reading", "% Overall Passing Rate"]].mean()
```

Out[314]:

	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing Rate
type					
Charter	83.473852	83.896421	93.620830	96.586489	95.103660
District	76.956733	80.966636	66.548453	80.799062	73.673757

In []: