Note

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

In [1]:
```
# Dependencies and Setup
import pandas as pd

# File to Load (Remember to Change These)
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 DataFrames
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=["sc
hool_name", "school_name"])
school_data_complete.head()
```

Out[1]:

	Student ID	student_name	gender	grade	school_name	reading_score	math_score	School ID	t
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

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 percentage of students with a passing math score (70 or greater)
- Calculate the percentage of students with a passing reading score (70 or greater)
- Calculate the percentage of students who passed math **and** reading (% Overall Passing)
- Create a dataframe to hold the above results
- Optional: give the displayed data cleaner formatting

In [2]:
```
# calculate unique school name
total_schools = len(school_data_complete.school_name.unique())
total_students = len(school_data_complete['Student ID'].unique())

#Calculate Total Budget
total_budget = school_data_complete['budget'].unique().sum()

# Calculate avergage math score
avg_math_score = school_data_complete['math_score'].mean()

# Calculate the average reading score
avg_reading_score = school_data_complete['reading_score'].mean()

# set new index as "math score"
new_school_data_complete = school_data_complete.set_index("math_score")
new_school_data_complete.head()
```

Out[2]:

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

In [3]:
```
# Calculate the percentage of students with a passing math score (70 or greate
r)
# 1. No. of students with passing math score over 70
students_math_with_above_70 = school_data_complete.loc[(school_data_complete[
'math_score']) >= 70].count()
print(numberOfStudent_passing_math)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-3-1943d01e4578> in <module>
      2 # 1. No. of students with passing math score over 70
      3 students_math_with_above_70 = school_data_complete.loc[(school_data_c
omplete['math_score']) >= 70].count()
----> 4 print(numberOfStudent_passing_math)

NameError: name 'numberOfStudent_passing_math' is not defined
```

In [ ]:
```
# 2. Calculate percentage of students with a passing math score (70 or greate
r)
percentage_students_math_with_above_70=students_math_with_above_70['math_scor
e']/total_students * 100
percentage_students_math_with_above_70
```

In [ ]:
```
# Calculate the numbers of students with a passing reading score (70 or greate
r)
students_reading_with_above_70 = school_data_complete.loc[(school_data_complet
e['reading_score']) >= 70].count()
students_reading_with_above_70
```

In [ ]:
```
# 2. Calculate percentage of students with a passing reading score (70 or grea
ter)
percentage_students_reading_with_above_70 = students_reading_with_above_70['re
ading_score']/total_students * 100
percentage_students_reading_with_above_70
```

In [ ]:
```
# Calculate the percentage of students who passed math and reading (% Overall
 Passing)

# 1. No. of students with passing math AND reading score over 70
students_overall_with_above_70 = school_data_complete.loc[((school_data_comple
te['math_score']) >= 70)|((school_data_complete['reading_score']) >= 70)].coun
t()
print(students_overall_with_above_70)
```

In [ ]:
```
# 2. Calculate percentage of students with a passing math AND reading score (7
0 or greater)
percentage_students_overall_with_above_70 = students_overall_with_above_70['St
udent ID']/total_students * 100
percentage_students_overall_with_above_70
```

In [ ]:
```
# create a dictionary summary

dictionary_sumary = pd.DataFrame([{
    "Total Schools": total_schools,
    "Total Students": total_students,
    "Total Budget": total_budget,
    "Average Math Score": avg_math_score,
    "Average Reading Score": avg_reading_score,
    "% Passing Math": percentage_students_math_with_above_70,
    "% Passing Reading": percentage_students_reading_with_above_70,
    "% Overall Passing": percentage_students_overall_with_above_70
}])

dictionary_sumary
```

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 (The percentage of students that passed math **and** reading.)
- Create a dataframe to hold the above results

In [ ]:
```
school_data_complete.describe()
```

In [6]:
```
#group by school name
group_school = school_data_complete.groupby(['school_name'])
print(group_school)
group_school.count()
```

```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x00000213774F8AC8>
```

Out[6]:

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

In [7]:
```
# student by school
#total_students_by_school = school_data_complete.groupby('school_name')['stude
nt_name'].count()
total_students_by_school = group_school['student_name'].count()

total_students_by_school

#total_students = len(school_data_complete['Student ID'].unique()
```

Out[7]:
```
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
Name: student_name, dtype: int64
```

In [8]:
```
# total budget  by school name
total_budget_by_school = group_school['budget'].sum()
total_budget_by_school
```

Out[8]:
```
school_name
Bailey High School       15549641728
Cabrera High School       2009159448
Figueroa High School      5557128039
Ford High School          4831365924
Griffin High School       1346890000
Hernandez High School    14007062700
Holden High School         105933149
Huang High School         5573322295
Johnson High School      14733628650
Pena High School           563595396
Rodriguez High School    10186904637
Shelton High School       1860672600
Thomas High School        1705517550
Wilson High School        3012587442
Wright High School        1888920000
Name: budget, dtype: int64
```

In [24]:
```
# Per Student Budget
Per_Student_Budget = total_budget_by_school/total_students_by_school
Per_Student_Budget
```

Out[24]:
```
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
Hernandez High School    3022020.0
Holden High School        248087.0
Huang High School        1910635.0
Johnson High School      3094650.0
Pena High School          585858.0
Rodriguez High School    2547363.0
Shelton High School      1056600.0
Thomas High School       1043130.0
Wilson High School       1319574.0
Wright High School       1049400.0
dtype: float64
```

In [10]:
```
# Average math score by school
average_math_score_by_school = group_school['math_score'].mean()
average_math_score_by_school
```

Out[10]:
```
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
Hernandez High School    77.289752
Holden High School       83.803279
Huang High School        76.629414
Johnson High School      77.072464
Pena High School         83.839917
Rodriguez High School    76.842711
Shelton High School      83.359455
Thomas High School       83.418349
Wilson High School       83.274201
Wright High School       83.682222
Name: math_score, dtype: float64
```

In [11]:
```
# Average reading score by school
average_reading_score_by_school = group_school['reading_score'].mean()
average_reading_score_by_school
```

Out[11]:
```
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
Hernandez High School    80.934412
Holden High School       83.814988
Huang High School        81.182722
Johnson High School      80.966394
Pena High School         84.044699
Rodriguez High School    80.744686
Shelton High School      83.725724
Thomas High School       83.848930
Wilson High School       83.989488
Wright High School       83.955000
Name: reading_score, dtype: float64
```

In [15]:
```
# % Passing Math
student_math_score_by_school = school_data_complete.loc[(school_data_complete['math_score']) >= 70].groupby('school_name').count()
Average_Math_Percentage_Score_by_school = student_math_score_by_school['math_score']/total_students_by_school * 100

# % Passing Reading
student_reading_score_by_school = school_data_complete.loc[(school_data_complete['reading_score']) >= 70].groupby('school_name').count()
Average_Reading_Percentage_Score_by_school = student_reading_score_by_school['reading_score']/total_students_by_school * 100

# % Overall Passing (The percentage of students that passed math and reading.)
total_pass_percentage_by_school = student_data.loc[(student_data['math_score'] >= 70) &
                                                   (student_data['reading_score'] >= 70)
                                                   ].groupby('school_name').count()

overall_pass_percentage_by_school = total_pass_percentage_by_school['Student ID']/total_students_by_school * 100
```

In [23]:
```
# create a summary data frame

school_summary = pd.DataFrame({'School Type': group_school['type'].first(),
                               'Total Students': total_students_by_school,
                               'Total School Budget': total_budget_by_school,
                               'Per Student Budget': Per_Student_Budget,
                               'Average Math Score': average_math_score_by_sch
ool,
                               'Average Reading Score': average_reading_score_
by_school,
                               '% Passing Math': Average_Reading_Percentage_Sc
ore_by_school,
                               '% Passing Reading': Average_Reading_Percentage
_Score_by_school,
                               '% Overall Passing': overall_pass_percentage_by
_school
                              })
#formatting
school_summary['Total School Budget']=school_summary['Total School Budget'].as
type(float).map("${:,.2f}".format)
#school_summary['Per Student Budget']=school_summary['Per Student Budget'].map
("${:,.2f}".format)
school_summary.style.format({'Per Student Budget':"${:,.2}",})
school_summary.index.name=None
school_summary
```

Out[23]:

	School Type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	
Bailey High School	District	4976	$15,549,641,728.00	3124928.0	77.048432	81.033963	81.933280	81
Cabrera High School	Charter	1858	$2,009,159,448.00	1081356.0	83.061895	83.975780	97.039828	97
Figueroa High School	District	2949	$5,557,128,039.00	1884411.0	76.711767	81.158020	80.739234	80
Ford High School	District	2739	$4,831,365,924.00	1763916.0	77.102592	80.746258	79.299014	79
Griffin High School	Charter	1468	$1,346,890,000.00	917500.0	83.351499	83.816757	97.138965	97
Hernandez High School	District	4635	$14,007,062,700.00	3022020.0	77.289752	80.934412	80.862999	80
Holden High School	Charter	427	$105,933,149.00	248087.0	83.803279	83.814988	96.252927	96
Huang High School	District	2917	$5,573,322,295.00	1910635.0	76.629414	81.182722	81.316421	81
Johnson High School	District	4761	$14,733,628,650.00	3094650.0	77.072464	80.966394	81.222432	81
Pena High School	Charter	962	$563,595,396.00	585858.0	83.839917	84.044699	95.945946	95
Rodriguez High School	District	3999	$10,186,904,637.00	2547363.0	76.842711	80.744686	80.220055	80
Shelton High School	Charter	1761	$1,860,672,600.00	1056600.0	83.359455	83.725724	95.854628	95
Thomas High School	Charter	1635	$1,705,517,550.00	1043130.0	83.418349	83.848930	97.308869	97
Wilson High School	Charter	2283	$3,012,587,442.00	1319574.0	83.274201	83.989488	96.539641	96
Wright High School	Charter	1800	$1,888,920,000.00	1049400.0	83.682222	83.955000	96.611111	9(

Top Performing Schools (By % Overall Passing)

- Sort and display the top five performing schools by % overall passing.

In [25]:
```
school_summary.sort_values('% Overall Passing',ascending=False).head()
```

Out[25]:

	School Type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	Pass Rea
Cabrera High School	Charter	1858	$2,009,159,448.00	1081356.0	83.061895	83.975780	97.039828	97.039
Thomas High School	Charter	1635	$1,705,517,550.00	1043130.0	83.418349	83.848930	97.308869	97.308
Griffin High School	Charter	1468	$1,346,890,000.00	917500.0	83.351499	83.816757	97.138965	97.138
Wilson High School	Charter	2283	$3,012,587,442.00	1319574.0	83.274201	83.989488	96.539641	96.539
Pena High School	Charter	962	$563,595,396.00	585858.0	83.839917	84.044699	95.945946	95.945

Bottom Performing Schools (By % Overall Passing)

- Sort and display the five worst-performing schools by % overall passing.

In [26]:
```
school_summary.sort_values('% Overall Passing',ascending=True).head()
```

Out[26]:

	School Type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	I
Rodriguez High School	District	3999	$10,186,904,637.00	2547363.0	76.842711	80.744686	80.220055	80
Figueroa High School	District	2949	$5,557,128,039.00	1884411.0	76.711767	81.158020	80.739234	80
Huang High School	District	2917	$5,573,322,295.00	1910635.0	76.629414	81.182722	81.316421	81
Hernandez High School	District	4635	$14,007,062,700.00	3022020.0	77.289752	80.934412	80.862999	80
Johnson High School	District	4761	$14,733,628,650.00	3094650.0	77.072464	80.966394	81.222432	81

Math Scores by Grade

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

In [27]:
```
avg_math_score_by_9th_grade = school_data_complete.loc[school_data_complete['grade']=='9th'].groupby('school_name')['math_score'].mean()
avg_math_score_by_10th_grade = school_data_complete.loc[school_data_complete['grade']=='10th'].groupby('school_name')['math_score'].mean()
avg_math_score_by_11th_grade = school_data_complete.loc[school_data_complete['grade']=='11th'].groupby('school_name')['math_score'].mean()
avg_math_score_by_12th_grade = school_data_complete.loc[school_data_complete['grade']=='12th'].groupby('school_name')['math_score'].mean()

Mathscores_summary = pd.DataFrame({
    '9th': avg_math_score_by_9th_grade,
    '10th': avg_math_score_by_10th_grade,
    '11th': avg_math_score_by_11th_grade,
    '12th': avg_math_score_by_12th_grade
})

Mathscores_summary
```

Out[27]:

school_name	9th	10th	11th	12th
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

Reading Score by Grade

- Perform the same operations as above for reading scores

In [29]:
```
avg_reading_score_by_9th_grade = school_data_complete.loc[school_data_complete
['grade']=='9th'].groupby('school_name')['reading_score'].mean()
avg_reading_score_by_10th_grade = school_data_complete.loc[school_data_complet
e['grade']=='10th'].groupby('school_name')['reading_score'].mean()
avg_reading_score_by_11th_grade = school_data_complete.loc[school_data_complet
e['grade']=='11th'].groupby('school_name')['reading_score'].mean()
avg_reading_score_by_12th_grade = school_data_complete.loc[school_data_complet
e['grade']=='12th'].groupby('school_name')['reading_score'].mean()

Readingscores_summary = pd.DataFrame({
    '9th': avg_reading_score_by_9th_grade,
    '10th': avg_reading_score_by_10th_grade,
    '11th': avg_reading_score_by_11th_grade,
    '12th': avg_reading_score_by_12th_grade
})
Readingscores_summary
```

Out[29]:

school_name	9th	10th	11th	12th
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

Scores by School Spending

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

In [30]:
```
bins = [0,584,629,644,674]
group_names=['<$584',"$585-629",'$630-644','$645-675']

Avg_Spending = school_summary.loc[:,['Average Math Score','Average Reading Sco
re','% Passing Math','% Passing Reading','% Overall Passing']]

Avg_Spending['Spending Ranges (Per Student)'] = pd.cut(school_summary['Per Stu
dent Budget'],bins,labels=group_names,include_lowest=True)

Avg_Spending = Avg_Spending.groupby('Spending Ranges (Per Student)').mean()

Avg_Spending.style.format({
    'Average Math Score':"{:,.2f}",
    'Average Reading Score':"{:,.2f}",
    '% Passing Math':"{:,.2f}",'% Passing Reading':"{:,.2f}",
    '% Overall Passing':"{:,.2f}"})
```

Out[30]:

Spending Ranges (Per Student)	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing
<$584	nan	nan	nan	nan	nan
$585-629	nan	nan	nan	nan	nan
$630-644	nan	nan	nan	nan	nan
$645-675	nan	nan	nan	nan	nan

Scores by School Size

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

In [31]:
```
bins = [0,999.99,1999.99,9999]
group_names=['Small (<1000)',"Medium (1000-2000)",'Large (2000-5000)']

School_size = school_summary.loc[:,['Average Math Score','Average Reading Score','% Passing Math','% Passing Reading','% Overall Passing']]

School_size['School Size'] = pd.cut(school_summary['Total Students'],bins,labels=group_names,include_lowest=True)

School_size = School_size.groupby('School Size').mean()

School_size
```

Out[31]:

School Size	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing
Small (<1000)	83.821598	83.929843	96.099437	96.099437	89.883853
Medium (1000-2000)	83.374684	83.864438	96.790680	96.790680	90.621535
Large (2000-5000)	77.746417	81.344493	82.766634	82.766634	58.286003

Scores by School Type

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

In [32]:
```
School_Type = school_summary.loc[:,['School Type','Average Math Score','Average Reading Score','% Passing Math','% Passing Reading','% Overall Passing']]
School_Type = School_Type.groupby('School Type').mean()
School_Type
```

Out[32]:

School Type	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing
Charter	83.473852	83.896421	96.586489	96.586489	90.432244
District	76.956733	80.966636	80.799062	80.799062	53.672208