

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=["school_name", "school_name"])
school_data_complete.head()
```

Out[1]:

	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

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]:

```
#variables and caluculations for Distinct summary

total_schools = school_data_complete["school_name"].nunique()
total_students = school_data_complete["Student ID"].nunique()
total_budget= school_data_complete["budget"].unique().sum()
avg_math_score = round(school_data_complete["math_score"].mean(),6)
avg_reading_score = round(school_data_complete["reading_score"].mean(),6)

#Percentages

# % Passing Math
passing_math_score = school_data_complete.loc[(school_data_complete["math_score"] >=70)]
count_students_math_passing = passing_math_score["math_score"].count()
percentage_math =round((count_students_math_passing/ total_students)* 100,6)

# % Passing Reading
passing_reading_score = school_data_complete.loc[(school_data_complete["reading_score"] >=70)]
count_students_reading_passing = passing_reading_score["reading_score"].count()
percentage_reading =round((count_students_reading_passing/ total_students)* 100,6)

# % Overall Passing
#percentage_overall = ((count_students_math_passing+count_students_reading_passing)/total_students)*100
overall_pass = school_data_complete.loc[(school_data_complete["math_score"] >=70) & (school_data_complete["reading_score"] >=70)]
count_students_overall = overall_pass["Student ID"].count()
percentage_overall_pass =round((count_students_overall/total_students)* 100,6)
```

```
In [3]: # District summary dataframe from dictionary
district_summary = pd.DataFrame({
    "Total Schools": [total_schools],
    "Total Students":'{:,.}.'.format(total_students),
    "Total Budget": '${:,.2f}'.format(total_budget),
    "Average Math Score": [avg_math_score],
    "Average Reading Score": [avg_reading_score],
    "% Passing Math": [percentage_math],
    "% Passing Reading": [percentage_reading],
    "% Overall Passing": [percentage_overall_pass]
})
district_summary
```

	Total Schools	Total Students	Total Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing
0	15	39,170	\$24,649,428.00	78.985371	81.87784	74.980853	85.805463	65.172326

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 [4]: # Grouped the complete data frame by school
grouped_school = school_data_complete.groupby(['school_name'])
```

Calculations

```

total_student = grouped_school.size()
school_type = grouped_school['type'].first()
total_budget = grouped_school['budget'].first()
total_budget_per_student = total_budget/total_student
average_math_score = grouped_school['math_score'].mean()
average_reading_score = grouped_school['reading_score'].mean()

# Calculate the percentage
grouped_passing_math = school_data_complete[school_data_complete['math_score']>=70].groupby(['school_name']).size()
percent_passing_math = (grouped_passing_math/total_student)*100

grouped_passing_reading = school_data_complete[school_data_complete['reading_score']>=70].groupby(['school_name']).size()
percent_passing_reading = (grouped_passing_reading/total_student)*100

#percent_overall_passing = (percent_passing_math + percent_passing_reading)/2
overall_passing = (school_data_complete['reading_score'] >= 70) & (school_data_complete['math_score'] >= 70)
percent_overall_passing = (overall_passing/total_student)*100

```

In [5]:

```

# School summary dataframe from dictionary
#TODO CHECK AGAIN--
school={
    'School Type': school_type,
    'Total Students':total_student,
    'Total School Budget': total_budget.map('${:.2f}'.format),
    'Per Student Budget': total_budget_per_student.map('${:.2f}'.format),
    'Average Math Score': average_math_score,
    'Average Reading Score': average_reading_score,
    '% Passing Math': percent_passing_math,
    '% Passing Reading': percent_passing_reading,
    '% Overall Passing Rate': percent_overall_passing,
}

school_summary = pd.DataFrame(school)
school_summary .index.name = None
school_summary

```

Out[5]:

	School Type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing Rate
--	-------------	----------------	---------------------	--------------------	--------------------	-----------------------	----------------	-------------------	------------------------

	School Type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing Rate
Bailey High School	District	4976	\$3,124,928.00	\$628.00	77.048432	81.033963	66.680064	81.933280	54.642283
Cabrera High School	Charter	1858	\$1,081,356.00	\$582.00	83.061895	83.975780	94.133477	97.039828	91.334769
Figueroa High School	District	2949	\$1,884,411.00	\$639.00	76.711767	81.158020	65.988471	80.739234	53.204476
Ford High School	District	2739	\$1,763,916.00	\$644.00	77.102592	80.746258	68.309602	79.299014	54.289887
Griffin High School	Charter	1468	\$917,500.00	\$625.00	83.351499	83.816757	93.392371	97.138965	90.599455
Hernandez High School	District	4635	\$3,022,020.00	\$652.00	77.289752	80.934412	66.752967	80.862999	53.527508
Holden High School	Charter	427	\$248,087.00	\$581.00	83.803279	83.814988	92.505855	96.252927	89.227166
Huang High School	District	2917	\$1,910,635.00	\$655.00	76.629414	81.182722	65.683922	81.316421	53.513884
Johnson High School	District	4761	\$3,094,650.00	\$650.00	77.072464	80.966394	66.057551	81.222432	53.539172
Pena High School	Charter	962	\$585,858.00	\$609.00	83.839917	84.044699	94.594595	95.945946	90.540541
Rodriguez High School	District	3999	\$2,547,363.00	\$637.00	76.842711	80.744686	66.366592	80.220055	52.988247
Shelton High School	Charter	1761	\$1,056,600.00	\$600.00	83.359455	83.725724	93.867121	95.854628	89.892107
Thomas High School	Charter	1635	\$1,043,130.00	\$638.00	83.418349	83.848930	93.272171	97.308869	90.948012
Wilson High School	Charter	2283	\$1,319,574.00	\$578.00	83.274201	83.989488	93.867718	96.539641	90.582567
Wright High School	Charter	1800	\$1,049,400.00	\$583.00	83.682222	83.955000	93.333333	96.611111	90.333333

Top Performing Schools (By % Overall Passing)

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

In [6]:

```
top_performing_schools = school_summary.sort_values(by='% Overall Passing Rate', ascending=False).head()
top_performing_schools.index.name = None
top_performing_schools
```

Out[6]:

	School Type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing Rate
Cabrera High School	Charter	1858	\$1,081,356.00	\$582.00	83.061895	83.975780	94.133477	97.039828	91.334769
Thomas High School	Charter	1635	\$1,043,130.00	\$638.00	83.418349	83.848930	93.272171	97.308869	90.948012
Griffin High School	Charter	1468	\$917,500.00	\$625.00	83.351499	83.816757	93.392371	97.138965	90.599455
Wilson High School	Charter	2283	\$1,319,574.00	\$578.00	83.274201	83.989488	93.867718	96.539641	90.582567
Pena High School	Charter	962	\$585,858.00	\$609.00	83.839917	84.044699	94.594595	95.945946	90.540541

Bottom Performing Schools (By % Overall Passing)

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

In [7]:

```
bottom_performing_schools = school_summary.sort_values(by='% Overall Passing Rate').head()
bottom_performing_schools.index.name = None
bottom_performing_schools
```

Out[7]:

	School Type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing Rate
Rodriguez High School	District	3999	\$2,547,363.00	\$637.00	76.842711	80.744686	66.366592	80.220055	52.988247

	School Type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing Rate
Figueroa High School	District	2949	\$1,884,411.00	\$639.00	76.711767	81.158020	65.988471	80.739234	53.204476
Huang High School	District	2917	\$1,910,635.00	\$655.00	76.629414	81.182722	65.683922	81.316421	53.513884
Hernandez High School	District	4635	\$3,022,020.00	\$652.00	77.289752	80.934412	66.752967	80.862999	53.527508
Johnson High School	District	4761	\$3,094,650.00	\$650.00	77.072464	80.966394	66.057551	81.222432	53.539172

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 [8]:

```
school_data_complete.head()
```

Out[8]:

	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 [9]:

```
#calculate average math score for each grade
avg_math_9th = school_data_complete.loc[school_data_complete["grade"]=="9th"].groupby("school_name")["math_score"].mean()
avg_math_10th = school_data_complete.loc[school_data_complete["grade"]=="10th"].groupby("school_name")["math_score"].mean()
avg_math_11th = school_data_complete.loc[school_data_complete["grade"]=="11th"].groupby("school_name")["math_score"].mean()
avg_math_12th = school_data_complete.loc[school_data_complete["grade"]=="12th"].groupby("school_name")["math_score"].mean()

#create new dataframe using dictionary
avg_math_score =pd.DataFrame({"9th":avg_math_9th,
                               "10th":avg_math_10th,
                               "11th":avg_math_11th,
                               "12th":avg_math_12th})
avg_math_score.index.name = None
avg_math_score
```

Out[9]:

	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 [10]:

```
#calculate average reading score for each grade
avg_reading_9th = school_data_complete.loc[school_data_complete["grade"]=="9th"].groupby("school_name")["reading_score"].mean()
avg_reading_10th = school_data_complete.loc[school_data_complete["grade"]=="10th"].groupby("school_name")["reading_score"].mean()
avg_reading_11th = school_data_complete.loc[school_data_complete["grade"]=="11th"].groupby("school_name")["reading_score"].mean()
avg_reading_12th = school_data_complete.loc[school_data_complete["grade"]=="12th"].groupby("school_name")["reading_score"].mean()

#create new dataframe using dictionary
avg_reading_score = pd.DataFrame({"9th":avg_reading_9th,
                                    "10th":avg_reading_10th,
                                    "11th":avg_reading_11th,
                                    "12th":avg_reading_12th})
avg_reading_score.index.name = None
avg_reading_score
```

Out[10]:

	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
Hemandez 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

	9th	10th	11th	12th
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 [11]:

```
#create spending range bins
spending_bins = [0, 585, 630, 645, 680]
spending_bins_labels = ["<$585", "$585-630", "$630-645", "$645-680"]

#school_summary.columns

school_summary["Per Student Spending Range"] = pd.cut(total_budget_per_student, spending_bins, labels = spending_bins_labels)
#school_summary["Per Student Spending Range"] = pd.cut(school_summary["Per Student Budget"], spending_bins, labels = spending_bins

spending_math_scores = school_summary.groupby(["Per Student Spending Range"]).mean()["Average Math Score"]
spending_reading_scores = school_summary.groupby(["Per Student Spending Range"]).mean()["Average Reading Score"]
spending_passing_math = school_summary.groupby(["Per Student Spending Range"]).mean()["% Passing Math"]
spending_passing_reading = school_summary.groupby(["Per Student Spending Range"]).mean()["% Passing Reading"]
overall_passing_rate = school_summary.groupby(["Per Student Spending Range"]).mean()["% Overall Passing Rate"]

spending_summary = pd.DataFrame({"Average Math Score" : spending_math_scores,
                                 "Average Reading Score": spending_reading_scores,
                                 "% Passing Math": spending_passing_math,
                                 "% Passing Reading": spending_passing_reading,
                                 "% Overall Passing Rate": overall_passing_rate})

spending_summary
```

Out[11]:

Per Student Spending Range	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing Rate
<\$585	83.455399	83.933814	93.460096	96.610877	90.369459
\$585-630	81.899826	83.155286	87.133538	92.718205	81.418596
\$630-645	78.518855	81.624473	73.484209	84.391793	62.857656
\$645-680	76.997210	81.027843	66.164813	81.133951	53.526855

In [12]:

school_summary

Out[12]:

School Name	School Type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing Rate	Per Student Spending Range
Bailey High School	District	4976	\$3,124,928.00	\$628.00	77.048432	81.033963	66.680064	81.933280	54.642283	\$585-630
Cabrera High School	Charter	1858	\$1,081,356.00	\$582.00	83.061895	83.975780	94.133477	97.039828	91.334769	<\$585
Figueroa High School	District	2949	\$1,884,411.00	\$639.00	76.711767	81.158020	65.988471	80.739234	53.204476	\$630-645
Ford High School	District	2739	\$1,763,916.00	\$644.00	77.102592	80.746258	68.309602	79.299014	54.289887	\$630-645
Griffin High School	Charter	1468	\$917,500.00	\$625.00	83.351499	83.816757	93.392371	97.138965	90.599455	\$585-630
Hernandez High School	District	4635	\$3,022,020.00	\$652.00	77.289752	80.934412	66.752967	80.862999	53.527508	\$645-680
Holden High School	Charter	427	\$248,087.00	\$581.00	83.803279	83.814988	92.505855	96.252927	89.227166	<\$585
Huang High School	District	2917	\$1,910,635.00	\$655.00	76.629414	81.182722	65.683922	81.316421	53.513884	\$645-680
Johnson High School	District	4761	\$3,094,650.00	\$650.00	77.072464	80.966394	66.057551	81.222432	53.539172	\$645-680

	School Type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing Rate	Per Student Spending Range
Pena High School	Charter	962	\$585,858.00	\$609.00	83.839917	84.044699	94.594595	95.945946	90.540541	\$585-630
Rodriguez High School	District	3999	\$2,547,363.00	\$637.00	76.842711	80.744686	66.366592	80.220055	52.988247	\$630-645
Shelton High School	Charter	1761	\$1,056,600.00	\$600.00	83.359455	83.725724	93.867121	95.854628	89.892107	\$585-630
Thomas High School	Charter	1635	\$1,043,130.00	\$638.00	83.418349	83.848930	93.272171	97.308869	90.948012	\$630-645
Wilson High School	Charter	2283	\$1,319,574.00	\$578.00	83.274201	83.989488	93.867718	96.539641	90.582567	<\$585
Wright High School	Charter	1800	\$1,049,400.00	\$583.00	83.682222	83.955000	93.333333	96.611111	90.333333	<\$585

Scores by School Size

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

In [13]:

```
#create school size bins
school_size_bins = [0, 1000, 2000, 5000]
school_size_bin_labels = ["Small (<1000)", "Medium (1000-2000)", "Large (2000-5000)"]
```

In [14]:

```
# Categorize the spending based on the bins
school_summary["School Size"] = pd.cut(school_summary["Total Students"], school_size_bins, labels=school_size_bin_labels)

# Calculations
size_math_scores = school_summary.groupby(["School Size"]).mean()["Average Math Score"]
size_reading_scores = school_summary.groupby(["School Size"]).mean()["Average Reading Score"]
size_passing_math = school_summary.groupby(["School Size"]).mean()["% Passing Math"]
size_passing_reading = school_summary.groupby(["School Size"]).mean()["% Passing Reading"]
overall_passing_rate = school_summary.groupby(["School Size"]).mean()["% Overall Passing Rate"]

#data frame
```

```
school_size_summary = pd.DataFrame({"Average Math Score" : size_math_scores,
                                    "Average Reading Score": size_reading_scores,
                                    "% Passing Math": size_passing_math,
                                    "% Passing Reading": size_passing_reading,
                                    "% Overall Passing Rate": overall_passing_rate})
```

```
school_size_summary
```

Out[14]:

School Size	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing Rate
Small (<1000)	83.821598	83.929843	93.550225	96.099437	89.883853
Medium (1000-2000)	83.374684	83.864438	93.599695	96.790680	90.621535
Large (2000-5000)	77.746417	81.344493	69.963361	82.766634	58.286003

Scores by School Type

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

In [15]:

```
school_summary.columns
```

Out[15]:

```
Index(['School Type', 'Total Students', 'Total School Budget',
       'Per Student Budget', 'Average Math Score', 'Average Reading Score',
       '% Passing Math', '% Passing Reading', '% Overall Passing Rate',
       'Per Student Spending Range', 'School Size'],
      dtype='object')
```

In [16]:

```
#New summary dataframe
school_summary_type = school_summary[['School Type',
                                      'Per Student Budget', 'Average Math Score', 'Average Reading Score',
                                      '% Passing Math', '% Passing Reading', '% Overall Passing Rate',
                                      'Per Student Spending Range', 'School Size']].copy()
```

```
#group by school type and average columns
school_summary_type = school_summary_type.groupby('School Type').mean()
```

```
#display dataframe  
school_summary_type
```

Out[16]:

School Type	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing Rate
Charter	83.473852	83.896421	93.620830	96.586489	90.432244
District	76.956733	80.966636	66.548453	80.799062	53.672208