```
In [97]:  import pandas as pd
          import numpy as np
```

```
In [98]:  school_data = "Resources/schools_complete.csv"
          student_data = "Resources/students_complete.csv"

          school_data = pd.read_csv(school_data)
          student_data = pd.read_csv(student_data)
          complete_data = pd.merge(student_data, school_data, how="left", on=["school_name"])
```

District Summary

In [99]:
```python
# total school count
school_names = school_data['school_name'].unique()
total_schools = len(school_names)
total_schools

# student count
total_students = school_data['size'].sum()
total_students

# total budget
total_budget = school_data['budget'].sum()
total_budget

# average math score
avg_math_score = student_data['math_score'].mean()
avg_math_score

# average reading score
avg_read_score = student_data['reading_score'].mean()
avg_read_score

# percentage of passing math score
passing_math = student_data.loc[student_data['math_score'] >= 70]['math_score'].count()
pass_math = passing_math/total_students
pass_math

# percentage of passing reading score
passing_read = student_data.loc[student_data['reading_score'] >= 70]['reading_score'].count()
pass_read = passing_read/total_students
pass_read

# convert info to dataframe
district_summary = pd.DataFrame({

    "Total Schools": [total_schools],
    "Total Students": [total_students],
    "Total Budget": [total_budget],
    "Average Math Score": [avg_math_score],
    "Average Reading Score": [avg_read_score],
    "% Passing Math": [pass_math],
    "% Passing Reading":[pass_read],
```

```
})

# re-format numbers
district_summary.style.format({"Total Budget": "${:,.2f}",
                               "Average Math Score": "{:.4f}",
                               "Average Reading Score": "{:.4f}",
                               "% Passing Math": "{:.3%}",
                               "% Passing Reading": "{:.3%}"})
```

Out[99]:

	Total Schools	Total Students	Total Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading
0	15	39170	$24,649,428.00	78.9854	81.8778	74.981%	85.805%

School Summary

In [100]:
```python
# group by school name
school_group = complete_data.set_index('school_name').groupby(['school_name'])

#school types
school_types = school_data.set_index('school_name')['type']

# total students by school
stu_by_school = school_group['Student ID'].count()

# school budget
sch_budget = school_data.set_index('school_name')['budget']

# budget per student
stu_budget = school_data.set_index('school_name')['budget']/school_data.set_index('school_name')['size']

# avg scores by school
avg_math = school_group['math_score'].mean()
avg_read = school_group['reading_score'].mean()

# % passing scores
pass_math_sch = complete_data[complete_data['math_score'] >= 70].groupby('school_name')['Student ID'].count()
/stu_by_school
pass_read_sch = complete_data[complete_data['reading_score'] >= 70].groupby('school_name')['Student ID'].count()/stu_by_school
overall_sch = complete_data[(complete_data['reading_score'] >= 70) & (complete_data['math_score'] >= 70)].groupby('school_name')['Student ID'].count()/stu_by_school

# create dataframe
school_summary = pd.DataFrame({
    "School Type": school_types,
    "Total Students": stu_by_school,
    "Total School Budget": sch_budget,
    "Per Student Budget": stu_budget,
    "Average Math Score": avg_math,
    "Average Reading Score": avg_read,
    '% Passing Math': pass_math_sch,
    '% Passing Reading': pass_read_sch,
    "Overall Passing Rate": overall_sch
})

# format numbers
school_summary.style.format({'Total Students': '{:,}',
```

```
                    "Total School Budget": "${:,}",
                    "Per Student Budget": "${:.2f}",
                    'Average Math Score': "{:.6f}",
                    'Average Reading Score': "{:.6f}",
                    "% Passing Math": "{:.6%}",
                    "% Passing Reading": "{:.6%}",
                    "Overall Passing Rate": "{:.6%}"})
```

Out[100]:

	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	4,976	$3,124,928	$628.00	77.048432	81.033963	66.680064%	81.933280%	54.642283%
Cabrera High School	Charter	1,858	$1,081,356	$582.00	83.061895	83.975780	94.133477%	97.039828%	91.334769%
Figueroa High School	District	2,949	$1,884,411	$639.00	76.711767	81.158020	65.988471%	80.739234%	53.204476%
Ford High School	District	2,739	$1,763,916	$644.00	77.102592	80.746258	68.309602%	79.299014%	54.289887%
Griffin High School	Charter	1,468	$917,500	$625.00	83.351499	83.816757	93.392371%	97.138965%	90.599455%
Hernandez High School	District	4,635	$3,022,020	$652.00	77.289752	80.934412	66.752967%	80.862999%	53.527508%
Holden High School	Charter	427	$248,087	$581.00	83.803279	83.814988	92.505855%	96.252927%	89.227166%
Huang High School	District	2,917	$1,910,635	$655.00	76.629414	81.182722	65.683922%	81.316421%	53.513884%
Johnson High School	District	4,761	$3,094,650	$650.00	77.072464	80.966394	66.057551%	81.222432%	53.539172%
Pena High School	Charter	962	$585,858	$609.00	83.839917	84.044699	94.594595%	95.945946%	90.540541%
Rodriguez High School	District	3,999	$2,547,363	$637.00	76.842711	80.744686	66.366592%	80.220055%	52.988247%
Shelton High School	Charter	1,761	$1,056,600	$600.00	83.359455	83.725724	93.867121%	95.854628%	89.892107%
Thomas High School	Charter	1,635	$1,043,130	$638.00	83.418349	83.848930	93.272171%	97.308869%	90.948012%

	School Type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	Overall Passing Rate
Wilson High School	Charter	2,283	$1,319,574	$578.00	83.274201	83.989488	93.867718%	96.539641%	90.582567%
Wright High School	Charter	1,800	$1,049,400	$583.00	83.682222	83.955000	93.333333%	96.611111%	90.333333%

Top Performing Schools

In [101]:
```python
# top 5 performing schools by passing rate
top_5 = school_summary.sort_values("Overall Passing Rate", ascending = False)
top_5.head().style.format({'Total Students': '{:,}',
                           "Total School Budget": "${:,}",
                           "Per Student Budget": "${:.2f}",
                           "% Passing Math": "{:.6%}",
                           "% Passing Reading": "{:.6%}",
                           "Overall Passing Rate": "{:.6%}"})
```

Out[101]:

	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	1,858	$1,081,356	$582.00	83.0619	83.9758	94.133477%	97.039828%	91.334769%
Thomas High School	Charter	1,635	$1,043,130	$638.00	83.4183	83.8489	93.272171%	97.308869%	90.948012%
Griffin High School	Charter	1,468	$917,500	$625.00	83.3515	83.8168	93.392371%	97.138965%	90.599455%
Wilson High School	Charter	2,283	$1,319,574	$578.00	83.2742	83.9895	93.867718%	96.539641%	90.582567%
Pena High School	Charter	962	$585,858	$609.00	83.8399	84.0447	94.594595%	95.945946%	90.540541%

Bottom Performing Schools

In [102]:
```python
# bottom 5 performing schools by passing rate
bottom_5 = top_5.tail()
bottom_5 = bottom_5.sort_values('Overall Passing Rate')
bottom_5.style.format({'Total Students': '{:,}',
                       "Total School Budget": "${:,}",
                       "Per Student Budget": "${:.2f}",
                       "% Passing Math": "{:.6%}",
                       "% Passing Reading": "{:.6%}",
                       "Overall Passing Rate": "{:.6%}"})
```

Out[102]:

	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	3,999	$2,547,363	$637.00	76.8427	80.7447	66.366592%	80.220055%	52.988247%
Figueroa High School	District	2,949	$1,884,411	$639.00	76.7118	81.158	65.988471%	80.739234%	53.204476%
Huang High School	District	2,917	$1,910,635	$655.00	76.6294	81.1827	65.683922%	81.316421%	53.513884%
Hernandez High School	District	4,635	$3,022,020	$652.00	77.2898	80.9344	66.752967%	80.862999%	53.527508%
Johnson High School	District	4,761	$3,094,650	$650.00	77.0725	80.9664	66.057551%	81.222432%	53.539172%

Math Score by Grade

In [103]:
```python
# average math score for school by grade level
ninth_math = student_data.loc[student_data['grade'] == '9th'].groupby('school_name')["math_score"].mean()
tenth_math = student_data.loc[student_data['grade'] == '10th'].groupby('school_name')["math_score"].mean()
eleventh_math = student_data.loc[student_data['grade'] == '11th'].groupby('school_name')["math_score"].mean()
twelfth_math = student_data.loc[student_data['grade'] == '12th'].groupby('school_name')["math_score"].mean()

# create dataframe
math_scores = pd.DataFrame({
        "9th": ninth_math,
        "10th": tenth_math,
        "11th": eleventh_math,
        "12th": twelfth_math
})
math_scores = math_scores[['9th', '10th', '11th', '12th']]
math_scores.index.name = "School"

# format
math_scores.style.format({'9th': '{:.6f}',
                          "10th": '{:.6f}',
                          "11th": "{:.6f}",
                          "12th": "{:.6f}"})
```

Out[103]:

	9th	10th	11th	12th
School				
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

```
In [104]:  # average reading score for school by grade level
           ninth_reading = student_data.loc[student_data['grade'] == '9th'].groupby('school_name')["reading_score"].mean
           ()
           tenth_reading = student_data.loc[student_data['grade'] == '10th'].groupby('school_name')["reading_score"].mea
           n()
           eleventh_reading = student_data.loc[student_data['grade'] == '11th'].groupby('school_name')["reading_score"].
           mean()
           twelfth_reading = student_data.loc[student_data['grade'] == '12th'].groupby('school_name')["reading_score"].m
           ean()

           # create dataframe
           reading_scores = pd.DataFrame({
                   "9th": ninth_reading,
                   "10th": tenth_reading,
                   "11th": eleventh_reading,
                   "12th": twelfth_reading
           })
           reading_scores = reading_scores[['9th', '10th', '11th', '12th']]
           reading_scores.index.name = "School"

           # format
           reading_scores.style.format({'9th': '{:.6f}',
                                        "10th": '{:.6f}',
                                        "11th": "{:.6f}",
                                        "12th": "{:.6f}"})
```

Out[104]:

School	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

In [105]:
```python
# create bins for spending
spending_bins = [0, 585, 615, 645, 675]
group_names = ["<$585", "$585-615", "$615-645", "$645-675"]
complete_data['spending_bins'] = pd.cut(complete_data['budget']/complete_data['size'], spending_bins, labels
= group_names)

# group by spending
by_spending = complete_data.groupby('spending_bins')

# get averages
avg_math = by_spending['math_score'].mean()
avg_read = by_spending['reading_score'].mean()
pass_math = complete_data[complete_data['math_score'] >= 70].groupby('spending_bins')['Student ID'].count()/b
y_spending['Student ID'].count()
pass_read = complete_data[complete_data['reading_score'] >= 70].groupby('spending_bins')['Student ID'].count
()/by_spending['Student ID'].count()
overall = complete_data[(complete_data['reading_score'] >= 70) & (complete_data['math_score'] >= 70)].groupby
('spending_bins')['Student ID'].count()/by_spending['Student ID'].count()


# create dataframe
scores_by_spend = pd.DataFrame({
    "Average Math Score": avg_math,
    "Average Reading Score": avg_read,
    '% Passing Math': pass_math,
    '% Passing Reading': pass_read,
    "Overall Passing Rate": overall

})

scores_by_spend.index.name = "Spending Ranges (Per Student)"
scores_by_spend = scores_by_spend.reindex(group_names)

# re-format
scores_by_spend.style.format({'Average Math Score': '{:.6f}',
                              'Average Reading Score': '{:.6f}',
                              '% Passing Math': '{:.6%}',
                              '% Passing Reading':'{:.6%}',
                              'Overall Passing Rate': '{:.6%}'})
```

Out[105]:

	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	Overall Passing Rate
Spending Ranges (Per Student)					
<$585	83.363065	83.964039	93.702889%	96.686558%	90.640704%
$585-615	83.529196	83.838414	94.124128%	95.886889%	90.121190%
$615-645	78.061635	81.434088	71.400428%	83.614770%	60.289317%
$645-675	77.049297	81.005604	66.230813%	81.109397%	53.528791%

Scores by School Size

In [106]:

```python
# create bins for school size
size_bins = [0, 1000, 2000, 5000]
group_names = ["Small (<1000)", "Medium (1000-2000)", "Large (2000-5000)"]
complete_data['size_bins'] = pd.cut(complete_data['size'], size_bins, labels = group_names)

# group by size
by_size = complete_data.groupby('size_bins')

# get averages
avg_math = by_size['math_score'].mean()
avg_read = by_size['math_score'].mean()
pass_math = complete_data[complete_data['math_score'] >= 70].groupby('size_bins')['Student ID'].count()/by_si
ze['Student ID'].count()
pass_read = complete_data[complete_data['reading_score'] >= 70].groupby('size_bins')['Student ID'].count()/by
_size['Student ID'].count()
overall = complete_data[(complete_data['reading_score'] >= 70) & (complete_data['math_score'] >= 70)].groupby
('size_bins')['Student ID'].count()/by_size['Student ID'].count()


# create dataframe
scores_by_size = pd.DataFrame({
    "Average Math Score": avg_math,
    "Average Reading Score": avg_read,
    '% Passing Math': pass_math,
    '% Passing Reading': pass_read,
    "Overall Passing Rate": overall

})

scores_by_size.index.name = "School Size"
scores_by_size = scores_by_size.reindex(group_names)

# re-format
scores_by_size.style.format({'Average Math Score': '{:.6f}',
                             'Average Reading Score': '{:.6f}',
                             '% Passing Math': '{:.6%}',
                             '% Passing Reading':'{:.6%}',
                             'Overall Passing Rate': '{:.6%}'})
```

Out[106]:

	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	Overall Passing Rate
School Size					
Small (<1000)	83.828654	83.828654	93.952484%	96.040317%	90.136789%
Medium (1000-2000)	83.372682	83.372682	93.616522%	96.773058%	90.624267%
Large (2000-5000)	77.477597	77.477597	68.652380%	82.125158%	56.574046%

Scores by School Type

In [107]:
```python
# group by school type
by_type = complete_data.groupby("type")

# get averages
avg_math = by_type['math_score'].mean()
avg_read = by_type['math_score'].mean()
pass_math = complete_data[complete_data['math_score'] >= 70].groupby('type')['Student ID'].count()/by_type['Student ID'].count()
pass_read = complete_data[complete_data['reading_score'] >= 70].groupby('type')['Student ID'].count()/by_type['Student ID'].count()
overall = complete_data[(complete_data['reading_score'] >= 70) & (complete_data['math_score'] >= 70)].groupby('type')['Student ID'].count()/by_type['Student ID'].count()

# create dataframe
scores_by_type = pd.DataFrame({
    "Average Math Score": avg_math,
    "Average Reading Score": avg_read,
    '% Passing Math': pass_math,
    '% Passing Reading': pass_read,
    "Overall Passing Rate": overall})
scores_by_type.index.name = "School Type"


# re-format
scores_by_type.style.format({'Average Math Score': '{:.6f}',
                             'Average Reading Score': '{:.6f}',
                             '% Passing Math': '{:.6%}',
                             '% Passing Reading':'{:.6%}',
                             'Overall Passing Rate': '{:.6%}'})
```

Out[107]:

School Type	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	Overall Passing Rate
Charter	83.406183	83.406183	93.701821%	96.645891%	90.560932%
District	76.987026	76.987026	66.518387%	80.905249%	53.695878%