**Project Title:-** CU Module 4 Challenge

**Description:-** As a first task, you've been asked to analyze the district-wide standardized test results. You'll be given access to every student's math and reading scores, as well as various information on the schools they attend. Your responsibility is to aggregate the data to and showcase obvious trends in school performance.

**Explanation:-** To complete this assignment, I needed to have access to the dataset and have some familiarity with using pandas. The dataset includes information on student scores and school information, such as school name, type, budget, and size.

**Step 2/3**

Here are two observable trends based on the data:

**Explanation:**

Schools with higher budgets tend to have higher passing rates: The data shows a positive correlation between school budgets and passing rates. Schools with higher budgets tend to have more resources and support, which may contribute to higher student performance.

Charter schools tend to have higher passing rates than district schools: The data shows that charter schools, on average, have higher passing rates than district schools. This may be due to various factors, such as smaller class sizes and more autonomy in decision-making.

**Step 3/3**

Overall, the data analysis provides insights into school performance based on various metrics such as budget, size, and type. This information can be used by the school board and mayor to make informed decisions about future school budgets and priorities.

**Final answer**

The final answer is the data analysis provides insights into school performance based on various metrics such as budget, size, and type. This information can be used by the school board and mayor to make informed decisions about future school budgets and priorities.

**README.md**

**Unit 4 | Assignment - Pandas, Pandas, Pandas**

**Background**

The data dive continues!

Now, it's time to take what you've learned about Python Pandas and apply it to new situations. For this assignment, you'll need to complete **one of two** (not both) Data Challenges. Once again, which challenge you take on is your choice. Just be sure to give it your all -- as the skills you hone will become powerful tools in your data analytics tool belt.

**Academy of Py**

Well done! Having spent years analyzing financial records for big banks, you've finally scratched your idealistic itch and joined the education sector. In your latest role, you've become the Chief Data Scientist for your city's school district. In this capacity, you'll be helping the school board and mayor make strategic decisions regarding future school budgets and priorities.

As a first task, you've been asked to analyze the district-wide standardized test results. You'll be given access to every student's math and reading scores, as well as various information on the schools they attend. Your responsibility is to aggregate the data to and showcase obvious trends in school performance.

Your final report should include each of the following:

**District Summary**

* Create a high level snapshot (in table form) of the district's key metrics, including:
  + Total Schools
  + Total Students
  + Total Budget
  + Average Math Score
  + Average Reading Score
  + % Passing Math
  + % Passing Reading
  + Overall Passing Rate (Average of the above two)

**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 Rate (Average of the above two)

**Top Performing Schools (By Passing Rate)**

* Create a table that highlights the top 5 performing schools based on Overall Passing Rate. Include:
  + 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 (Average of the above two)

**Bottom Performing Schools (By Passing Rate)**

* Create a table that highlights the bottom 5 performing schools based on Overall Passing Rate. Include all of the same metrics as above.

**Math Scores by Grade\*\***

* Create a table that lists the average Math Score for students of each grade level (9th, 10th, 11th, 12th) at each school.

**Reading 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.

**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)

**Scores by School Size**

* Repeat the above breakdown, but this time group schools based on a reasonable approximation of school size (Small, Medium, Large).

**Scores by School Type**

* Repeat the above breakdown, but this time group schools based on school type (Charter vs. District).

As final considerations:

* Use the pandas library and Jupyter Notebook.
* You must submit a link to your Jupyter Notebook with the viewable Data Frames.
* You must include a written description of at least two observable trends based on the data.
* See [Example Solution](https://github.com/shrawantee/Pandas-PyCitySchools/blob/master/PyCitySchools/PyCitySchools_starter.ipynb) for a reference on the expected format.

**Hints and Considerations**

* These are challenging activities for a number of reasons. For one, these activities will require you to analyze thousands of records. Hacking through the data to look for obvious trends in Excel is just not a feasible option. The size of the data may seem daunting, but pandas will allow you to efficiently parse through it.
* Second, these activities will also challenge you by requiring you to learn on your feet. Don't fool yourself into thinking: "I need to study pandas more closely before diving in." Get the basic gist of the library and then *immediately* get to work. When facing a daunting task, it's easy to think: "I'm just not ready to tackle it yet." But that's the surest way to never succeed. Learning to program requires one to constantly tinker, experiment, and learn on the fly. You are doing exactly the *right* thing, if you find yourself constantly practicing Google-Fu and diving into documentation. There is just no way (or reason) to try and memorize it all. Online references are available for you to use when you need them. So use them!
* Take each of these tasks one at a time. Begin your work, answering the basic questions: "How do I import the data?" "How do I convert the data into a DataFrame?" "How do I build the first table?" Don't get intimidated by the number of asks. Many of them are repetitive in nature with just a few tweaks. Be persistent and creative!
* Expect these exercises to take time! Don't get discouraged if you find yourself spending hours initially with little progress. Force yourself to deal with the discomfort of not knowing and forge ahead. Consider these hours an investment in your future!
* As always, feel encouraged to work in groups and get help from your TAs and Instructor. Just remember, true success comes from mastery and *not* a completed homework assignment. So challenge yourself to truly succeed!

## ****PyCitySchools\_starter-checkpoint.ipynb****

**PyCity Schools Analysis**

* As a whole, schools with higher budgets, did not yield better test results. By contrast, schools with higher spending per student actually (645−675)��������������������������ℎ�������ℎ��������������(<585 per student).
* As a whole, smaller and medium sized schools dramatically out-performed large sized schools on passing math performances (89-91% passing vs 67%).
* As a whole, charter schools out-performed the public district schools across all metrics. However, more analysis will be required to glean if the effect is due to school practices or the fact that charter schools tend to serve smaller student populations per school.

**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 [2]:

*# 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\_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"])

**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 overall passing rate (overall average score), i.e. (avg. math score + avg. reading score)/2
* 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)
* Create a dataframe to hold the above results
* Optional: give the displayed data cleaner formatting

In [11]:

Out[11]:

|  | **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 | 74.980853 | 85.805463 | 80.431606 |

**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 Rate (Average of the above two)
* Create a dataframe to hold the above results

**Top Performing Schools (By Passing Rate)**

* Sort and display the top five schools in overall passing rate

In [13]:

Out[13]:

|  | | **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 | | $1,08 | |  | |  | |  | |  | |  | |  |
| **Thomas High School** | Charter | | 1635 | | $1,043,130.00 | | $638.00 | | 83.418349 | | 83.848930 | | 93.272171 | | 97.308869 | | 95.290520 | |
| **Pena High School** | Charter | | 962 | | $585,858.00 | | $609.00 | | 83.839917 | | 84.044699 | | 94.594595 | | 95.945946 | | 95.270270 | |
| **Griffin High School** | Charter | | 1468 | | $917,500.00 | | $625.00 | | 83.351499 | | 83.816757 | | 93.392371 | | 97.138965 | | 95.265668 | |
| **Wilson High School** | Charter | | 2283 | | $1,319,574.00 | | $578.00 | | 83.274201 | | 83.989488 | | 93.867718 | | 96.539641 | | 95.203679 | |

**Bottom Performing Schools (By Passing Rate)**

* Sort and display the five worst-performing schools

In [14]:

Out[14]:

|  | **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 | 73.293323 |
| **Figueroa High School** | District | 2949 | $1,884,411.00 | $639.00 | 76.711767 | 81.158020 | 65.988471 | 80.739234 | 73.363852 |
| **Huang High School** | District | 2917 | $1,910,635.00 | $655.00 | 76.629414 | 81.182722 | 65.683922 | 81.316421 | 73.500171 |
| **Johnson High School** | District | 4761 | $3,094,650.00 | $650.00 | 77.072464 | 80.966394 | 66.057551 | 81.222432 | 73.639992 |
| **Ford High School** | District | 2739 | $1,763,916.00 | $644.00 | 77.102592 | 80.746258 | 68.309602 | 79.299014 | 73.804308 |

**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

| **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 [16]:

Out[16]:

|  | **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 [17]:

*# Sample bins. Feel free to create your own bins.*

spending\_bins **=** [0, 585, 615, 645, 675]

group\_names **=** ["<$585", "$585-615", "$615-645", "$645-675"]

In [18]:

Out[18]:

|  | **Average Math Score** | **Average Reading Score** | **% Passing Math** | **% Passing Reading** | **% Overall Passing Rate** |
| --- | --- | --- | --- | --- | --- |
| **Spending Ranges (Per Student)** |  |  |  |  |  |
| **<$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 |

**Scores by School Size**

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

In [ ]:

*# Sample bins. Feel free to create your own bins.*

size\_bins **=** [0, 1000, 2000, 5000]

group\_names **=** ["Small (<1000)", "Medium (1000-2000)", "Large (2000-5000)"]

In [19]:

Out[19]:

|  | **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 |

**Scores by School Type**

* Perform the same operations as above, based on school type.

In [20]:

Out[20]:

|  | **Average Math Score** | **Average Reading Score** | **% Passing Math** | **% Passing Reading** | **% Overall Passing Rate** |
| --- | --- | --- | --- | --- | --- |
| **School Type** |  |  |  |  |  |
| **Charter** | 83.473852 | 83.896421 | 93.620830 | 96.586489 | 95.103660 |
| **District** | 76.956733 | 80.966636 | 66.548453 | 80.799062 | 73.673757 |

In [ ]:

## ****PyCitySchools\_DS\_HW4.ipynb****

**PyCity Schools Analysis**

* As a whole, schools with higher budgets, did not yield better test results. By contrast, schools with higher spending per student actually (645−675)��������������������������ℎ�������ℎ��������������(<585 per student).
* As a whole, smaller and medium sized schools dramatically out-performed large sized schools on passing math performances (89-91% passing vs 67%).
* As a whole, charter schools out-performed the public district schools across all metrics. However, more analysis will be required to glean if the effect is due to school practices or the fact that charter schools tend to serve smaller student populations per school.

In [243]:

*# Note to self --> students\_complete.csv has the following fields: Student\_Id, Student name, gender, grade, school\_name, reading\_score*

*# and math\_score. Number of observations = 39170.*

*#schools\_complete.csv has the following fields: school\_id, school\_name, school\_type, school\_size, school\_budget. Number of observations = 15*

*# The unique identifier in the two datasets is the "school\_name"*

In [244]:

*# Dependencies and Setup*

**import** pandas **as** pd

*# 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\_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"])

|  |
| --- |
|  |
| **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 |
| **5** | 5 | Bryan Miranda | M | 9th | Huang High School | 94 | 94 | 0 | District | 2917 | 1910635 |
| **6** | 6 | Sheena Carter | F | 11th | Huang High School | 82 | 80 | 0 | District | 2917 | 1910635 |
| **7** | 7 | Nicole Baker | F | 12th | Huang High School | 96 | 69 | 0 | District | 2917 | 1910635 |
| **8** | 8 | Michael Roth | M | 10th | Huang High School | 95 | 87 | 0 | District | 2917 | 1910635 |
| **9** | 9 | Matthew Greene | M | 10th | Huang High School | 96 | 84 | 0 | District | 2917 | 1910635 |
| **10** | 10 | Andrew Alexander | M | 10th | Huang High School | 90 | 70 | 0 | District | 2917 | 1910635 |
| **11** | 11 | Daniel Cooper | M | 10th | Huang High School | 78 | 77 | 0 | District | 2917 | 1910635 |
| **12** | 12 | Brittney Walker | F | 9th | Huang High School | 64 | 79 | 0 | District | 2917 | 1910635 |
| **13** | 13 | William Long | M | 9th | Huang High School | 71 | 79 | 0 | District | 2917 | 1910635 |
| **14** | 14 | Tammy Hebert | F | 10th | Huang High School | 85 | 67 | 0 | District | 2917 | 1910635 |
| **15** | 15 | Dr. Jordan Carson | M | 11th | Huang High School | 94 | 88 | 0 | District | 2917 | 1910635 |
| **16** | 16 | Donald Zamora | M | 9th | Huang High School | 88 | 55 | 0 | District | 2917 | 1910635 |
| **17** | 17 | Kimberly Santiago | F | 9th | Huang High School | 74 | 75 | 0 | District | 2917 | 1910635 |
| **18** | 18 | Kevin Stevens | M | 9th | Huang High School | 64 | 69 | 0 | District | 2917 | 1910635 |
| **19** | 19 | Brandi Lyons | F | 9th | Huang High School | 89 | 80 | 0 | District | 2917 | 1910635 |
| **20** | 20 | Lisa Davis | F | 10th | Huang High School | 91 | 89 | 0 | District | 2917 | 1910635 |
| **21** | 21 | Kristen Lopez | F | 10th | Huang High School | 90 | 77 | 0 | District | 2917 | 1910635 |
| **22** | 22 | Kimberly Stewart | F | 11th | Huang High School | 99 | 84 | 0 | District | 2917 | 1910635 |
| **23** | 23 | Christopher Parker | M | 9th | Huang High School | 81 | 68 | 0 | District | 2917 | 1910635 |
| **24** | 24 | Chelsea Griffith | F | 11th | Huang High School | 85 | 73 | 0 | District | 2917 | 1910635 |
| **25** | 25 | Cesar Morris | M | 9th | Huang High School | 92 | 70 | 0 | District | 2917 | 1910635 |
| **26** | 26 | Melanie Decker | F | 9th | Huang High School | 63 | 85 | 0 | District | 2917 | 1910635 |
| **27** | 27 | Tracey Oconnor | F | 10th | Huang High School | 80 | 58 | 0 | District | 2917 | 1910635 |
| **28** | 28 | Kelly James | F | 11th | Huang High School | 73 | 55 | 0 | District | 2917 | 1910635 |
| **29** | 29 | Nicole Brown | F | 12th | Huang High School | 90 | 88 | 0 | District | 2917 | 1910635 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **39140** | 39140 | Cheyenne Hernandez | F | 9th | Thomas High School | 76 | 99 | 14 | Charter | 1635 | 1043130 |
| **39141** | 39141 | Jonathan Sullivan | M | 10th | Thomas High School | 86 | 93 | 14 | Charter |  |  |
| **39142** | 39142 | Deborah Higgins DDS | F | 10th | Thomas High School | 96 | 83 | 14 | Charter | 1635 | 1043130 |
| **39143** | 39143 | Steven Jackson | M | 11th | Thomas High School | 71 | 93 | 14 | Charter | 1635 | 1043130 |
| **39144** | 39144 | Cristian Webster | M | 12th | Thomas High School | 77 | 87 | 14 | Charter | 1635 | 1043130 |
| **39145** | 39145 | Audrey Fry | F | 10th | Thomas High School | 94 | 74 | 14 | Charter | 1635 | 1043130 |
| **39146** | 39146 | Michael Carroll | M | 9th | Thomas High School | 69 | 95 | 14 | Charter | 1635 | 1043130 |
| **39147** | 39147 | Raymond Hawkins | M | 10th | Thomas High School | 88 | 81 | 14 | Charter | 1635 | 1043130 |
| **39148** | 39148 | Christopher Wilson | M | 10th | Thomas High School | 89 | 89 | 14 | Charter | 1635 | 1043130 |
| **39149** | 39149 | Glenda Fletcher | F | 11th | Thomas High School | 82 | 93 | 14 | Charter | 1635 | 1043130 |
| **39150** | 39150 | Jennifer Hamilton | F | 11th | Thomas High School | 80 | 75 | 14 | Charter | 1635 | 1043130 |
| **39151** | 39151 | Shannon Williams | F | 10th | Thomas High School | 84 | 73 | 14 | Charter | 1635 | 1043130 |
| **39152** | 39152 | Lori Moore | F | 9th | Thomas High School | 98 | 84 | 14 | Charter | 1635 | 1043130 |
| **39153** | 39153 | William Hubbard | M | 9th | Thomas High School | 80 | 75 | 14 | Charter | 1635 | 1043130 |
| **39154** | 39154 | Bradley Johnson | M | 12th | Thomas High School | 91 | 71 | 14 | Charter | 1635 | 1043130 |
| **39155** | 39155 | John Brooks | M | 10th | Thomas High School | 92 | 98 | 14 | Charter | 1635 | 1043130 |
| **39156** | 39156 | Stephanie Contreras | F | 11th | Thomas High School | 79 | 95 | 14 | Charter | 1635 | 1043130 |
| **39157** | 39157 | Kristen Gonzalez | F | 9th | Thomas High School | 79 | 94 | 14 | Charter | 1635 | 1043130 |
| **39158** | 39158 | Kari Holloway | F | 10th | Thomas High School | 87 | 90 | 14 | Charter | 1635 | 1043130 |
| **39159** | 39159 | Kimberly Cabrera | F | 11th | Thomas High School | 85 | 72 | 14 | Charter | 1635 | 1043130 |
| **39160** | 39160 | Katie Weaver | F | 11th | Thomas High School | 89 | 86 | 14 | Charter | 1635 | 1043130 |
| **39161** | 39161 | April Reyes | F | 10th | Thomas High School | 70 | 84 | 14 | Charter | 1635 | 1043130 |
| **39162** | 39162 | Derek Weeks | M | 12th | Thomas High School | 94 | 77 | 14 | Charter | 1635 | 1043130 |
| **39163** | 39163 | John Reese | M | 11th | Thomas High School | 90 | 75 | 14 | Charter | 1635 | 1043130 |
| **39164** | 39164 | Joseph Anthony | M | 9th | Thomas High School | 97 | 76 | 14 | Charter | 1635 | 1043130 |
| **39165** | 39165 | Donna Howard | F | 12th | Thomas High School | 99 | 90 | 14 | Charter | 1635 | 1043130 |
| **39166** | 39166 | Dawn Bell | F | 10th | Thomas High School | 95 | 70 | 14 | Charter | 1635 | 1043130 |
| **39167** | 39167 | Rebecca Tanner | F | 9th | Thomas High School | 73 | 84 | 14 | Charter | 1635 | 1043130 |
| **39168** | 39168 | Desiree Kidd | F | 10th | Thomas High School | 99 | 90 | 14 | Charter | 1635 | 1043130 |
| **39169** | 39169 | Carolyn Jackson | F | 11th | Thomas High School | 95 | 75 | 14 | Charter | 1635 | 1043130 |

39170 rows × 11 columns

**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 overall passing rate (overall average score), i.e. (avg. math score + avg. reading score)/2
* 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)
* Create a dataframe to hold the above results
* Optional: give the displayed data cleaner formatting

In [245]:

*#Calculate the total number of schools*

total\_number\_schools **=** len(school\_data\_complete["School ID"]**.**unique())

total\_number\_schools

Out[245]:

15

In [246]:

*#Calculate the total number of students*

total\_number\_students **=** len(school\_data\_complete["Student ID"]**.**unique())

total\_number\_students

Out[246]:

39170

In [247]:

*#Calculate the total budget*

total\_budget **=** school\_data["budget"]**.**sum()

total\_budget

Out[247]:

24649428

*#Calculate the average math score*

average\_math\_score **=** student\_data["math\_score"]**.**mean()

average\_math\_score

Out[248]:

78.98537145774827

In [249]:

*#Calculate the average reading score*

average\_reading\_score **=** student\_data["reading\_score"]**.**mean()

average\_reading\_score

Out[249]:

81.87784018381414

In [250]:

*#Calculate the overall passing rate (overall average score), i.e. (avg. math score + avg. reading score)/2*

overall\_average\_score **=** (average\_math\_score **+** average\_reading\_score)**/**2

overall\_average\_score

Out[250]:

80.43160582078121

In [251]:

*#Calculate the percentage of students with a passing math score (70 or greater)*

*#Create 1s for the passing students by math or reading. 0 otherwise. And take the mean to get the average. It will give the percent of the passing students.*

student\_data["#passing\_math"] **=** student\_data["math\_score"] **>=** 70

student\_data["#passing\_reading"] **=** student\_data["reading\_score"] **>=** 70

percent\_passing\_math **=** ((student\_data["#passing\_math"])**.**mean())**\***100

percent\_passing\_math

Out[251]:

74.9808526933878

In [252]:

*#Calculate the percentage of students with a passing reading score (70 or greater)*

percent\_passing\_reading **=** ((student\_data["#passing\_reading"])**.**mean())**\***100

percent\_passing\_reading

Out[252]:

85.80546336482001

In [253]:

*#Calculate overall percentage*

overall\_passing\_rate **=** (percent\_passing\_math **+** percent\_passing\_reading)**/**2

overall\_passing\_rate

Out[253]:

80.39315802910392

In [254]:

*#Create a dataframe to hold the above results*

*#Optional: give the displayed data cleaner formatting*

district\_results **=** [{"Total Schools": total\_number\_schools,

"Total Students": total\_number\_students,

"Total Budget": total\_budget,

"Average Math Score": round(average\_math\_score,2),

"Average Reading Score": round(average\_reading\_score,2),

"% Passing Math": round(percent\_passing\_math,2),

"% Passing Reading": round(percent\_passing\_reading,2),

"% Overall Passing Rate": round(overall\_passing\_rate,2)}]

district\_summary\_table **=** pd**.**DataFrame(district\_results)

*#Formatting*

district\_summary\_table["% Passing Math"] **=** district\_summary\_table["% Passing Math"]**.**map("{:,.2f}%"**.**format)

district\_summary\_table["% Passing Reading"] **=** district\_summary\_table["% Passing Reading"]**.**map("{:,.2f}%"**.**format)

district\_summary\_table["% Overall Passing Rate"] **=** district\_summary\_table["% Overall Passing Rate"]**.**map("{:,.2f}%"**.**format)

district\_summary\_table["Total Budget"] **=** district\_summary\_table["Total Budget"]**.**map("${:,.2f}"**.**format)

district\_summary\_table["Total Students"] **=** district\_summary\_table["Total Students"]**.**map("{:,}"**.**format)

*#Display*

district\_summary\_table

Out[254]:

|  | **% Overall Passing Rate** | **% Passing Math** | **% Passing Reading** | **Average Math Score** | **Average Reading Score** | **Total Budget** | **Total Schools** | **Total Students** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 80.39% | 74.98% | 85.81% | 78.99 | 81.88 | $24,649,428.00 | 15 | 39,170 |

**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 Rate (Average of the above two)
* Create a dataframe to hold the above results

In [255]:

*#For this part, school\_data\_complete*

school\_data\_complete["passing\_math"] **=** school\_data\_complete["math\_score"] **>=** 70

school\_data\_complete["passing\_reading"] **=** school\_data\_complete["reading\_score"] **>=** 70

school\_data\_complete

| **Student ID** | | **student\_name** | | | **gender** | | | **grade** | | | **school\_name** | | | **reading\_score** | | | **math\_score** | | | | **School ID** | | | | **type** | | | **size** | **budget** | | | **passing\_math** | | | **passing\_reading** | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | | 0 | | | Paul Bradley | | | M | | | 9th | | | Huang High School | | | 66 | | | | 79 | | | | 0 | | | District | 2917 | | | 1910635 | | | True | | | | | False |
| **1** | | 1 | | | Victor Smith | | | M | | | 12th | | | Huang High School | | | 94 | | | | 61 | | | | 0 | | | District | 2917 | | | 1910635 | | | False | | | | | True |
| **2** | | 2 | | | Kevin Rodriguez | | | M | | | 12th | | | Huang High School | | | 90 | | | | 60 | | | | 0 | | | District | 2917 | | | 1910635 | | | False | | | | | True |
| **3** | | 3 | | | Dr. Richard Scott | | | M | | | 12th | | | Huang High School | | | 67 | | | | 58 | | | | 0 | | | District | 2917 | | | 1910635 | | | False | | | | | False |
| **4** | | 4 | | | Bonnie Ray | | | F | | | 9th | | | Huang High School | | | 97 | | | | 84 | | | | 0 | | | District | 2917 | | | 1910635 | | | True | | | | | True |
| **5** | | 5 | | | Bryan Miranda | | | M | | | 9th | | | Huang High School | | | 94 | | | | 94 | | | | 0 | | | District | 2917 | | | 1910635 | | | True | | | | | True |
| **6** | | 6 | | | Sheena Carter | | | F | | | 11th | | | Huang High School | | | 82 | | | | 80 | | | | 0 | | | District | 2917 | | | 1910635 | | | True | | | | | True |
| **7** | | 7 | | | Nicole Baker | | | F | | | 12th | | | Huang High School | | | 96 | | | | 69 | | | | 0 | | | District | 2917 | | | 1910635 | | | False | | | | | True |
| **8** | | 8 | | | Michael Roth | | | M | | | 10th | | | Huang High School | | | 95 | | | | 87 | | | | 0 | | | District | 2917 | | | 1910635 | | | True | | | | | True |
| **9** | | 9 | | | Matthew Greene | | | M | | | 10th | | | Huang High School | | | 96 | | | | 84 | | | | 0 | | | District | 2917 | | | 1910635 | | | True | | | | | True |
| **10** | | 10 | | | Andrew Alexander | | | M | | | 10th | | | Huang High School | | | 90 | | | | 70 | | | | 0 | | | District | 2917 | | | 1910635 | | | True | | | | | True |
| **11** | | 11 | | | Daniel Cooper | | | M | | | 10th | | | Huang High School | | | 78 | | | | 77 | | | | 0 | | | District | 2917 | | | 1910635 | | | True | | | | | True |
| **12** | | 12 | | | Brittney Walker | | | F | | | 9th | | | Huang High School | | | 64 | | | | 79 | | | | 0 | | | District | 2917 | | | 1910635 | | | True | | | | | False |
| **13** | | 13 | | | William Long | | | M | | | 9th | | | Huang High School | | | 71 | | | | 79 | | | | 0 | | | District | 2917 | | | 1910635 | | | True | | | | | True |
| **14** | | 14 | | | Tammy Hebert | | | F | | | 10th | | | Huang High School | | | 85 | | | | 67 | | | | 0 | | | District | 2917 | | | 1910635 | | | False | | | | | True |
| **15** | | 15 | | | Dr. Jordan Carson | | | M | | | 11th | | | Huang High School | | | 94 | | | | 88 | | | | 0 | | | District | 2917 | | | 1910635 | | | True | | | | | True |
| **16** | | 16 | | | Donald Zamora | | | M | | | 9th | | | Huang High School | | | 88 | | | | 55 | | | | 0 | | | District | 2917 | | | 1910635 | | | False | | | | | True |
| **17** | | 17 | | | Kimberly Santiago | | | F | | | 9th | | | Huang High School | | | 74 | | | | 75 | | | | 0 | | | District | 2917 | | | 1910635 | | | True | | | | | True |
| **18** | | 18 | | | Kevin Stevens | | | M | | | 9th | | | Huang High School | | | 64 | | | | 69 | | | | 0 | | | District | 2917 | | | 1910635 | | | False | | | | | False |
| **19** | | 19 | | | Brandi Lyons | | | F | | | 9th | | | Huang High School | | | 89 | | | | 80 | | | | 0 | | | District | 2917 | | | 1910635 | | | True | | | | | True |
| **20** | | 20 | | | Lisa Davis | | | F | | | 10th | | | Huang High School | | | 91 | | | | 89 | | | | 0 | | | District | 2917 | | | 1910635 | | | True | | | | | True |
| **21** | | 21 | | | Kristen Lopez | | | F | | | 10th | | | Huang High School | | | 90 | | | | 77 | | | | 0 | | | District | 2917 | | | 1910635 | | | True | | | | | True |
| **22** | | 22 | | | Kimberly Stewart | | | F | | | 11th | | | Huang High School | | | 99 | | | | 84 | | | | 0 | | | District | 2917 | | | 1910635 | | | True | | | | | True |
| **23** | | 23 | | | Christopher Parker | | | M | | | 9th | | | Huang High School | | | 81 | | | | 68 | | | | 0 | | | District | 2917 | | | 1910635 | | | False | | | | | True |
| **24** | | 24 | | | Chelsea Griffith | | | F | | | 11th | | | Huang High School | | | 85 | | | | 73 | | | | 0 | | | District | 2917 | | | 1910635 | | | True | | | | | True |
| **25** | | 25 | | | Cesar Morris | | | M | | | 9th | | | Huang High School | | | 92 | | | | 70 | | | | 0 | | | District | 2917 | | | 1910635 | | | True | | | | | True |
| **26** | | 26 | | | Melanie Decker | | | F | | | 9th | | | Huang High School | | | 63 | | | | 85 | | | | 0 | | | District | 2917 | | | 1910635 | | | True | | | | | False |
| **27** | | 27 | | | Tracey Oconnor | | | F | | | 10th | | | Huang High School | | | 80 | | | | 58 | | | | 0 | | | District | 2917 | | | 1910635 | | | False | | | | | True |
| **28** | | 28 | | | Kelly James | | | F | | | 11th | | | Huang High School | | | 73 | | | | 55 | | | | 0 | | | District | 2917 | | | 1910635 | | | False | | | | | True |
| **29** | | 29 | | | Nicole Brown | | | F | | | 12th | | | Huang High School | | | 90 | | | | 88 | | | | 0 | | | District | 2917 | | | 1910635 | | | True | | | | | True |
| **...** | | ... | | | ... | | | ... | | | ... | | | ... | | | ... | | | | ... | | | | ... | | | ... | ... | | | ... | | | ... | | | | | ... |
| **39140** | | 39140 | | | Cheyenne Hernandez | | | F | | | 9th | | | Thomas High School | | | 76 | | | | 99 | | | | 14 | | | Charter | 1635 | | | 1043130 | | | True | | | | | True |
| **39141** | | 39141 | | | Jonathan Sullivan | | | M | | | 10th | | | Thomas High School | | | 86 | | | | 93 | | | | 14 | | | Charter | 1635 | | | 1043130 | | | True | | | | | True |
| **39142** | | 39142 | | | Deborah Higgins DDS | | | F | | | 10th | | | Thomas High School | | | 96 | | | | 83 | | | | 14 | | | Charter | 1635 | | | 1 | | |  | | | | |  |
|  | | |
| **39143** | | | 39143 | | | Steven Jackson | | M | | | | 11th | | | Thomas High School | | 71 | | 93 | | | 14 | | | Charter | | | | | 1635 | | | 1043130 | | True | | True | | |
| **39144** | | | 39144 | | | Cristian Webster | | M | | | | 12th | | | Thomas High School | | 77 | | 87 | | | 14 | | | Charter | | | | | 1635 | | | 1043130 | | True | | True | | |
| **39145** | | | 39145 | | | Audrey Fry | | F | | | | 10th | | | Thomas High School | | 94 | | 74 | | | 14 | | | Charter | | | | | 1635 | | | 1043130 | | True | | True | | |
| **39146** | | | 39146 | | | Michael Carroll | | M | | | | 9th | | | Thomas High School | | 69 | | 95 | | | 14 | | | Charter | | | | | 1635 | | | 1043130 | | True | | False | | |
| **39147** | | | 39147 | | | Raymond Hawkins | | M | | | | 10th | | | Thomas High School | | 88 | | 81 | | | 14 | | | Charter | | | | | 1635 | | | 1043130 | | True | | True | | |
| **39148** | | | 39148 | | | Christopher Wilson | | M | | | | 10th | | | Thomas High School | | 89 | | 89 | | | 14 | | | Charter | | | | | 1635 | | | 1043130 | | True | | True | | |
| **39149** | | | 39149 | | | Glenda Fletcher | | F | | | | 11th | | | Thomas High School | | 82 | | 93 | | | 14 | | | Charter | | | | | 1635 | | | 1043130 | | True | | True | | |
| **39150** | | | 39150 | | | Jennifer Hamilton | | F | | | | 11th | | | Thomas High School | | 80 | | 75 | | | 14 | | | Charter | | | | | 1635 | | | 1043130 | | True | | True | | |
| **39151** | | | 39151 | | | Shannon Williams | | F | | | | 10th | | | Thomas High School | | 84 | | 73 | | | 14 | | | Charter | | | | | 1635 | | | 1043130 | | True | | True | | |
| **39152** | | | 39152 | | | Lori Moore | | F | | | | 9th | | | Thomas High School | | 98 | | 84 | | | 14 | | | Charter | | | | | 1635 | | | 1043130 | | True | | True | | |
| **39153** | | | 39153 | | | William Hubbard | | M | | | | 9th | | | Thomas High School | | 80 | | 75 | | | 14 | | | Charter | | | | | 1635 | | | 1043130 | | True | | True | | |
| **39154** | | | 39154 | | | Bradley Johnson | | M | | | | 12th | | | Thomas High School | | 91 | | 71 | | | 14 | | | Charter | | | | | 1635 | | | 1043130 | | True | | True | | |
| **39155** | | | 39155 | | | John Brooks | | M | | | | 10th | | | Thomas High School | | 92 | | 98 | | | 14 | | | Charter | | | | | 1635 | | | 1043130 | | True | | True | | |
| **39156** | | | 39156 | | | Stephanie Contreras | | F | | | | 11th | | | Thomas High School | | 79 | | 95 | | | 14 | | | Charter | | | | | 1635 | | | 1043130 | | True | | True | | |
| **39157** | | | 39157 | | | Kristen Gonzalez | | F | | | | 9th | | | Thomas High School | | 79 | | 94 | | | 14 | | | Charter | | | | | 1635 | | | 1043130 | | True | | True | | |
| **39158** | | | 39158 | | | Kari Holloway | | F | | | | 10th | | | Thomas High School | | 87 | | 90 | | | 14 | | | Charter | | | | | 1635 | | | 1043130 | | True | | True | | |
| **39159** | | | 39159 | | | Kimberly Cabrera | | F | | | | 11th | | | Thomas High School | |  | |  | | |  | | |  | | | | |  | | |  | |  | |  | | |
| **39160** | | | 39160 | | | Katie Weaver | | | | F | | 11th | | | Thomas High School | | | | 89 | | | 86 | | 14 | | | Charter | | | 1635 | | | 1043130 | | | | True | | True | |
| **39161** | | | 39161 | | | April Reyes | | | | F | | 10th | | | Thomas High School | | | | 70 | | | 84 | | 14 | | | Charter | | | 1635 | | | 1043130 | | | | True | | True | |
| **39162** | | | 39162 | | | Derek Weeks | | | | M | | 12th | | | Thomas High School | | | | 94 | | | 77 | | 14 | | | Charter | | | 1635 | | | 1043130 | | | | True | | True | |
| **39163** | | | 39163 | | | John Reese | | | | M | | 11th | | | Thomas High School | | | | 90 | | | 75 | | 14 | | | Charter | | | 1635 | | | 1043130 | | | | True | | True | |
| **39164** | | | 39164 | | | Joseph Anthony | | | | M | | 9th | | | Thomas High School | | | | 97 | | | 76 | | 14 | | | Charter | | | 1635 | | | 1043130 | | | | True | | True | |
| **39165** | | | 39165 | | | Donna Howard | | | | F | | 12th | | | Thomas High School | | | | 99 | | | 90 | | 14 | | | Charter | | | 1635 | | | 1043130 | | | | True | | True | |
| **39166** | | | 39166 | | | Dawn Bell | | | | F | | 10th | | | Thomas High School | | | | 95 | | | 70 | | 14 | | | Charter | | | 1635 | | | 1043130 | | | | True | | True | |
| **39167** | | | 39167 | | | Rebecca Tanner | | | | F | | 9th | | | Thomas High School | | | | 73 | | | 84 | | 14 | | | Charter | | | 1635 | | | 1043130 | | | | True | | True | |
| **39168** | | | 39168 | | | Desiree Kidd | | | | F | | 10th | | | Thomas High School | | | | 99 | | | 90 | | 14 | | | Charter | | | 1635 | | | 1043130 | | | | True | | True | |
| **39169** | | | 39169 | | | Carolyn Jackson | | | | F | | 11th | | | Thomas High School | | | | 95 | | | 75 | | 14 | | | Charter | | | 1635 | | | 1043130 | | | | True | | True | |

39170 rows × 13 columns

In [256]:

*# Use groupby by school\_name*

school\_group **=** school\_data\_complete**.**groupby(["school\_name"])**.**mean()

school\_group["Per Student Budget"] **=** school\_group["budget"]**/**school\_group["size"]

school\_group["% Passing Math"] **=** round(school\_group["passing\_math"]**\***100,2)

school\_group["% Passing Reading"] **=** round(school\_group["passing\_reading"]**\***100,2)

school\_group["% Overall Passing Rate"] **=** round(((school\_group["passing\_math"] **+** school\_group["passing\_reading"])**/**2)**\***100,3)

*#Merge with school\_data to collect information about the type, size and budget*

school\_data\_summary **=** pd**.**merge(school\_group, school\_data, how**=**"left", on**=**["school\_name", "school\_name"])

**del** school\_data\_summary['size\_y']

**del** school\_data\_summary['budget\_y']

**del** school\_data\_summary['Student ID']

**del** school\_data\_summary['School ID\_x']

*#Create a dataframe to store the results*

school\_summary\_dataframe **=** pd**.**DataFrame({"School Name": school\_data\_summary["school\_name"],

"School Type": school\_data\_summary["type"],

"Total Students":school\_data\_summary["size\_x"],

"Total School Budget": school\_data\_summary["budget\_x"],

"Per Student Budget":school\_data\_summary["Per Student Budget"],

"Average Math Score":round(school\_data\_summary["math\_score"],2),

"Average Reading Score":round(school\_data\_summary["reading\_score"],2),

"% Passing Math": school\_data\_summary["% Passing Math"],

"% Passing Reading": school\_data\_summary["% Passing Reading"],

"% Overall Passing Rate": school\_data\_summary["% Overall Passing Rate"]})

*#Formatting*

school\_summary\_dataframe["Total Students"] **=** school\_summary\_dataframe["Total Students"]**.**map("{:,.0f}"**.**format)

school\_summary\_dataframe["Total School Budget"] **=** school\_summary\_dataframe["Total School Budget"]**.**map("${:,.2f}"**.**format)

school\_summary\_dataframe["Per Student Budget"] **=** school\_summary\_dataframe["Per Student Budget"]**.**map("${:,.2f}"**.**format)

*#Display*

school\_summary\_dataframe

| **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** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | Bailey High School | District | 4,976 | $3,124,928.00 | $628.00 | 77.05 | 81.03 | 66.68 | 81.93 | 74.307 |
| **1** | Cabrera High School | Charter | 1,858 | $1,081,356.00 | $582.00 | 83.06 | 83.98 | 94.13 | 97.04 | 95.587 |
| **2** | Figueroa High School | District | 2,949 | $1,884,411.00 | $639.00 | 76.71 | 81.16 | 65.99 | 80.74 | 73.364 |
| **3** | Ford High School | District | 2,739 | $1,763,916.00 | $644.00 | 77.10 | 80.75 | 68.31 | 79.30 | 73.804 |
| **4** | Griffin High School | Charter | 1,468 | $917,500.00 | $625.00 | 83.35 | 83.82 | 93.39 | 97.14 | 95.266 |
| **5** | Hernandez High School | District | 4,635 | $3,022,020.00 | $652.00 | 77.29 | 80.93 | 66.75 | 80.86 | 73.808 |
| **6** | Holden High School | Charter | 427 | $248,087.00 | $581.00 | 83.80 | 83.81 | 92.51 | 96.25 | 94.379 |
| **7** | Huang High School | District | 2,917 | $1,910,635.00 | $655.00 | 76.63 | 81.18 | 65.68 | 81.32 | 73.500 |
| **8** | Johnson High School | District | 4,761 | $3,094,650.00 | $650.00 | 77.07 | 80.97 | 66.06 | 81.22 | 73.640 |
| **9** | Pena High School | Charter | 962 | $585,858.00 | $609.00 | 83.84 | 84.04 | 94.59 | 95.95 | 95.270 |
| **10** | Rodriguez High School | District | 3,999 | $2,547,363.00 | $637.00 | 76.84 | 80.74 | 66.37 | 80.22 | 73.293 |
| **11** | Shelton High School | Charter | 1,761 | $1,056,600.00 | $600.00 | 83.36 | 83.73 | 93.87 | 95.85 | 94.861 |
| **12** | Thomas High School | Charter | 1,635 | $1,043,130.00 | $638.00 | 83.42 | 83.85 | 93.27 | 97.31 | 95.291 |
| **13** | Wilson High School | Charter | 2,283 | $1,319,574.00 | $578.00 | 83.27 | 83.99 | 93.87 | 96.54 | 95.204 |
| **14** | Wright High School | Charter | 1,800 | $1,049,400.00 | $583.00 | 83.68 | 83.96 | 93.33 | 96.61 | 94.972 |

**Top Performing Schools (By Passing Rate)**

* Sort and display the top five schools in overall passing rate

In [257]:

*#Sort and display the top five schools in overall passing rate*

top\_five\_schools **=** school\_summary\_dataframe**.**sort\_values(["% Overall Passing Rate"], ascending**=False**)

top\_five\_schools**.**head()

| **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** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | Cabrera High School | Charter | 1,858 | $1,081,356.00 | $582.00 | 83.06 | 83.98 | 94.13 | 97.04 | 95.587 |
| **12** | Thomas High School | Charter | 1,635 | $1,043,130.00 | $638.00 | 83.42 | 83.85 | 93.27 | 97.31 | 95.291 |
| **9** | Pena High School | Charter | 962 | $585,858.00 | $609.00 | 83.84 | 84.04 | 94.59 | 95.95 | 95.270 |
| **4** | Griffin High School | Charter | 1,468 | $917,500.00 | $625.00 | 83.35 | 83.82 | 93.39 | 97.14 | 95.266 |
| **13** | Wilson High School | Charter | 2,283 | $1,319,574.00 | $578.00 | 83.27 | 83.99 | 93.87 | 96.54 | 95.204 |

**Bottom Performing Schools (By Passing Rate)**

* Sort and display the five worst-performing schools

In [258]:

*#Sort and display the five worst-performing schools*

bottom\_five\_schools **=** school\_summary\_dataframe**.**sort\_values(["% Overall Passing Rate"], ascending**=True**)

bottom\_five\_schools**.**head()

| **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** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **10** | Rodriguez High School | District | 3,999 | $2,547,363.00 | $637.00 | 76.84 | 80.74 | 66.37 | 80.22 | 73.293 |
| **2** | Figueroa High School | District | 2,949 | $1,884,411.00 | $639.00 | 76.71 | 81.16 | 65.99 | 80.74 | 73.364 |
| **7** | Huang High School | District | 2,917 | $1,910,635.00 | $655.00 | 76.63 | 81.18 | 65.68 | 81.32 | 73.500 |
| **8** | Johnson High School | District | 4,761 | $3,094,650.00 | $650.00 | 77.07 | 80.97 | 66.06 | 81.22 | 73.640 |
| **3** | Ford High School | District | 2,739 | $1,763,916.00 | $644.00 | 77.10 | 80.75 | 68.31 | 79.30 | 73.804 |

**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
* *#Create a pandas series for each grade. Group each series by school.*
* nineth\_grade**=** school\_data\_complete[school\_data\_complete["grade"] **==** "9th"]**.**groupby("school\_name")**.**mean()["math\_score"]
* tenth\_grade **=** school\_data\_complete[school\_data\_complete["grade"] **==** "10th"]**.**groupby("school\_name")**.**mean()["math\_score"]
* eleventh\_grade **=** school\_data\_complete[school\_data\_complete["grade"] **==** "11th"]**.**groupby("school\_name")**.**mean()["math\_score"]
* twelveth\_grade**=** school\_data\_complete[school\_data\_complete["grade"] **==** "12th"]**.**groupby("school\_name")**.**mean()["math\_score"]
* *#Combine the series into a dataframe*
* math\_grade\_dataframe **=** pd**.**DataFrame({"Ninth Grade":nineth\_grade, "Tenth Grade":tenth\_grade,
* "Eleventh Grade":eleventh\_grade, "Twelveth Grade":twelveth\_grade})
* *#Optional formatting: Give the displayed data cleaner formatting*
* math\_grade\_dataframe[["Ninth Grade","Tenth Grade","Eleventh Grade","Twelveth Grade"]] **=** math\_grade\_dataframe[["Ninth Grade","Tenth Grade","Eleventh Grade","Twelveth Grade"]]**.**applymap("{:.2f}"**.**format)
* *#Display*
* math\_grade\_dataframe

Ninth GradeTenth GradeEleventh GradeTwelveth Gradeschool\_nameBailey High School77.0877.0077.5276.49Cabrera High School83.0983.1582.7783.28Figueroa High School76.4076.5476.8877.15Ford High School77.3677.6776.9276.18Griffin High School82.0484.2383.8483.36Hernandez High School77.4477.3477.1477.19Holden High School83.7983.4385.0082.86Huang High School77.0375.9176.4577.23Johnson High School77.1976.6977.4976.86Pena High School83.6383.3784.3384.12Rodriguez High School76.8676.6176.4077.69Shelton High School83.4282.9283.3883.78Thomas High School83.5983.0983.5083.50Wilson High School83.0983.7283.2083.04Wright High School83.2684.0183.8483.64

**Reading Score by Grade**

* Perform the same operations as above for reading scores

In [260]:

*#Perform the same operations as above for reading scores*

*#Create a pandas series for each grade. Group each series by school.*

nineth\_grade**=** school\_data\_complete[school\_data\_complete["grade"] **==** "9th"]**.**groupby("school\_name")**.**mean()["reading\_score"]

tenth\_grade **=** school\_data\_complete[school\_data\_complete["grade"] **==** "10th"]**.**groupby("school\_name")**.**mean()["reading\_score"]

eleventh\_grade **=** school\_data\_complete[school\_data\_complete["grade"] **==** "11th"]**.**groupby("school\_name")**.**mean()["reading\_score"]

twelveth\_grade**=** school\_data\_complete[school\_data\_complete["grade"] **==** "12th"]**.**groupby("school\_name")**.**mean()["reading\_score"]

*#Combine the series into a dataframe*

reading\_grade\_dataframe **=** pd**.**DataFrame({"Ninth Grade":nineth\_grade, "Tenth Grade":tenth\_grade,

"Eleventh Grade":eleventh\_grade, "Twelveth Grade":twelveth\_grade})

*#Optional formatting: Give the displayed data cleaner formatting*

reading\_grade\_dataframe[["Ninth Grade","Tenth Grade","Eleventh Grade","Twelveth Grade"]] **=** reading\_grade\_dataframe[["Ninth Grade","Tenth Grade","Eleventh Grade","Twelveth Grade"]]**.**applymap("{:.2f}"**.**format)

*#Display*

reading\_grade\_dataframe

Ninth GradeTenth GradeEleventh GradeTwelveth Gradeschool\_nameBailey High School81.3080.9180.9580.91Cabrera High School83.6884.2583.7984.29Figueroa High School81.2081.4180.6481.38Ford High School80.6381.2680.4080.66Griffin High School83.3783.7184.2984.01Hernandez High School80.8780.6681.4080.86Holden High School83.6883.3283.8284.70Huang High School81.2981.5181.4280.31Johnson High School81.2680.7780.6281.23Pena High School83.8183.6184.3484.59Rodriguez High School80.9980.6380.8680.38Shelton High School84.1283.4484.3782.78Thomas High School83.7384.2583.5983.83Wilson High School83.9484.0283.7684.32Wright High School83.8383.8184.1684.07

**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 [261]:

*# Sample bins. Feel free to create your own bins.*

spending\_bins **=** [0, 585, 615, 645, 675]

group\_names **=** ["<$585", "$585-615", "$615-645", "$645-675"]

*# Use 4 reasonable bins to group school spending.*

school\_data\_summary["Spending Ranges (Per Student)"] **=** pd**.**cut(school\_data\_summary["Per Student Budget"], spending\_bins, labels**=**group\_names)

school\_spending\_grouped **=** school\_data\_summary**.**groupby("Spending Ranges (Per Student)")**.**mean()

*#Remove the unwanted columns as per the sample provided*

**del** school\_spending\_grouped['size\_x']

**del** school\_spending\_grouped['budget\_x']

**del** school\_spending\_grouped['Per Student Budget']

**del** school\_spending\_grouped['School ID\_y']

**del** school\_spending\_grouped['passing\_math']

**del** school\_spending\_grouped['passing\_reading']

school\_spending\_grouped

Out[261]:

|  | **reading\_score** | **math\_score** | **% Passing Math** | **% Passing Reading** | **% Overall Passing Rate** |
| --- | --- | --- | --- | --- | --- |
| **Spending Ranges (Per Student)** |  |  |  |  |  |
| **<$585** | 83.933814 | 83.455399 | 93.460000 | 96.610000 | 95.035500 |
| **$585-615** | 83.885211 | 83.599686 | 94.230000 | 95.900000 | 95.065500 |
| **$615-645** | 81.891436 | 79.079225 | 75.668333 | 86.106667 | 80.887500 |
| **$645-675** | 81.027843 | 76.997210 | 66.163333 | 81.133333 | 73.649333 |
| **Scores by School Size**   * Perform the same operations as above, based on school size.   In [262]:  *# Sample bins. Feel free to create your own bins.*  size\_bins **=** [0, 1000, 2000, 5000]  group\_names **=** ["Small (<1000)", "Medium (1000-2000)", "Large (2000-5000)"]  *# Use 4 reasonable bins to group school size.*  school\_data\_summary["School Size"] **=** pd**.**cut(school\_data\_summary["size\_x"], size\_bins, labels**=**group\_names)  school\_data\_summary  *#group by size\_x*  school\_size\_grouped **=** school\_data\_summary**.**groupby("School Size")**.**mean()  school\_size\_grouped  *#Remove the unwanted columns as per the sample provided*  *#del school\_size\_grouped['size\_x']*  **del** school\_size\_grouped['budget\_x']  **del** school\_size\_grouped['Per Student Budget']  **del** school\_size\_grouped['School ID\_y']  **del** school\_size\_grouped['passing\_math']  **del** school\_size\_grouped['passing\_reading']  *#Display*  school\_size\_grouped  Out[262]:   |  | **reading\_score** | **math\_score** | **size\_x** | **% Passing Math** | **% Passing Reading** | **% Overall Passing Rate** | | --- | --- | --- | --- | --- | --- | --- | | **School Size** |  |  |  |  |  |  | | **Small (<1000)** | 83.929843 | 83.821598 | 694.500 | 93.55000 | 96.10000 | 94.8245 | | **Medium (1000-2000)** | 83.864438 | 83.374684 | 1704.400 | 93.59800 | 96.79000 | 95.1954 | | **Large (2000-5000)** | 81.344493 | 77.746417 | 3657.375 | 69.96375 | 82.76625 | 76.3650 |   **Scores by School Type**   * Perform the same operations as above, based on school type.   In [263]:  school\_type\_grouped **=** school\_data\_summary**.**groupby("type")**.**mean()  *#Remove the unwanted columns as per the sample provided*  **del** school\_type\_grouped['size\_x']  **del** school\_type\_grouped['budget\_x']  **del** school\_type\_grouped['Per Student Budget']  **del** school\_type\_grouped['School ID\_y']  **del** school\_type\_grouped['passing\_math']  **del** school\_type\_grouped['passing\_reading']  school\_type\_grouped  Out[263]:   |  | **reading\_score** | **math\_score** | **% Passing Math** | **% Passing Reading** | **% Overall Passing Rate** | | --- | --- | --- | --- | --- | --- | | **type** |  |  |  |  |  | | **Charter** | 83.896421 | 83.473852 | 93.620000 | 96.586250 | 95.103750 | | **District** | 80.966636 | 76.956733 | 66.548571 | 80.798571 | 73.673714 | |  |  |  |  |  |