The school district summary will be a high-level snapshot of the district's key metrics:

* Total number of students = 39170 (Out [26])
  + # Get the total number of students in 1 step
  + student\_count = school\_data\_complete\_df["Student ID"].count()
  + student\_count
* Total number of schools = 15 (Out [29])
  + # Calculate the number of high schools from the array school\_count\_2
  + len(school\_data\_complete\_df["school\_name"].unique())
* Total budget = 24649428 (Out [30])
  + # Calculate the total budget. (note this is on school\_data\_df, not school\_data\_complete\_df)
  + total\_budget = school\_data\_df["budget"].sum()
  + total\_budget
* Average math score = 78.98537145774827 (Out [32]
  + # Calculate the average math score.
  + average\_math\_score = school\_data\_complete\_df["math\_score"].mean()
  + average\_math\_score
* Average reading score = 81.87784018381414 (Out [31]
  + # Calculate the average reading score.
  + average\_reading\_score = school\_data\_complete\_df["reading\_score"].mean()
  + average\_reading\_score
* Percentage of students who passed math
* Percentage of students who passed reading
* Overall passing percentage

To get the percentage of students who passed math and reading, we will write code to:

1. Determine the passing grade. (In [35] and In [36])
   * # Get all the students that are passing reading in a new DataFrame.
   * passing\_reading = school\_data\_complete\_df[school\_data\_complete\_df["reading\_score"] >= 70]
2. Get the number of students who passed math and reading in separate DataFrames. (In [40])
   * # Calculate the number of students passing math.
   * passing\_math\_count = passing\_math["student\_name"].count()
   * print(passing\_math\_count)
   * # Calculate the number of students passing reading.
   * passing\_reading\_count = passing\_reading["student\_name"].count()
   * print(passing\_reading\_count)
3. Calculate the number of students who passed math and reading. (Out 40) = 29370 math and 33610 reading
4. Calculate the percentage of students who passed math and reading.

To get the overall passing percentage, we will write code to:

1. Get the number of students who passed both math and reading in a DataFrame.
2. Calculate the number of students who passed both math and reading.
3. Calculate the percentage of students who passed both math and reading.

We'll find this information and visualize the data with a table like the following:

In order to get all of this data organized in one table, we'll need to merge the two DataFrames and perform analysis on the single, merged DataFrame. Although we'll be merging the DataFrames, it may be more efficient to use either school\_data\_df or student\_data\_df when performing certain calculations.

To merge two DataFrames, there must be a column in each of the DataFrames with the same name. So before we merge, let's review the column names in each DataFrame.

The columns in school\_data\_df are:

* School ID
* school\_name
* type
* size
* budget

The columns in the student\_data\_df are:

* Student ID
* student\_name
* gender
* grade
* school\_name
* reading\_score
* math\_score

We'll merge school\_data\_df and student\_data\_dfon a shared column using the merge() method. The column that these DataFrames have in common is school\_name. Inside the parentheses of the merge() method, we'll do the following:

* Add the DataFrames to be merged.
* Add the shared column to each DataFrame so that the merge can occur.
* Define how the DataFrames should be merged: left, right, inner, or outer. The default is inner. (You will learn more about merging later in this course.)