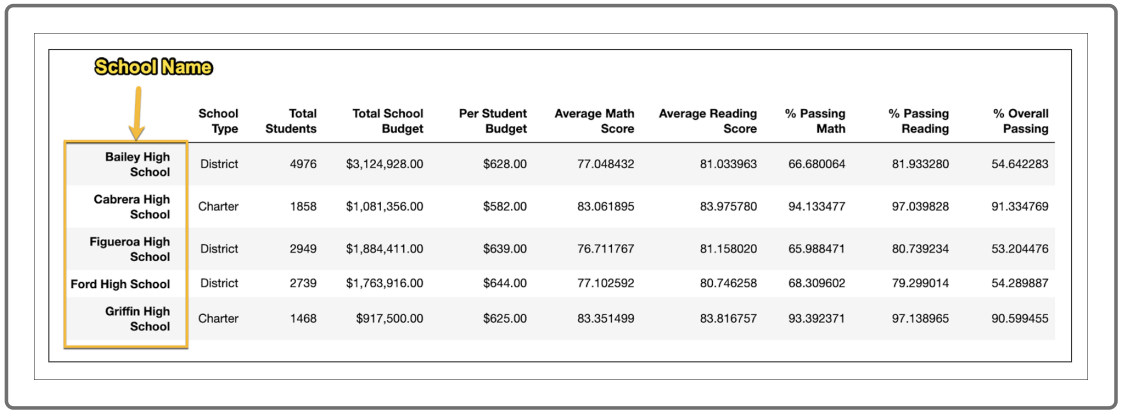
4.8.1 Set the Index to the School Name

Maria is impressed with the work you have done generating the district summary. Now she would like to generate a similar summary for each school in the district.

This next project requires you to get the following key metrics for each school and place them in a school summary DataFrame. As a reminder, here are the key metrics you're working with:

- 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 school summary DataFrame should look like the image below. You'll see that the `school_name` is now the index for this DataFrame. This is important, because as we perform calculations and gather data, we will be creating either Series or DataFrames that need to have the `school_name` as the index.



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

Now we'll create a new DataFrame. As mentioned, the index needs to be the `school_name` with "School Type" as the first column.

We can use the data from the `school_data_df` DataFrame because the "school_name" and "type" columns are both Series:

	School ID	school_name	type	size	budget
0	0	Huang High School	District	2917	1910635
1	1	Figueroa High School	District	2949	1884411
2	2	Shelton High School	Charter	1761	1056600
3	3	Hernandez High School	District	4635	3022020
4	4	Griffin High School	Charter	1468	917500
5	5	Wilson High School	Charter	2283	1319574
6	6	Cabrera High School	Charter	1858	1081356
7	7	Bailey High School	District	4976	3124928
8	8	Holden High School	Charter	427	248087
9	9	Pena High School	Charter	962	585858
10	10	Wright High School	Charter	1800	1049400
11	11	Rodriguez High School	District	3999	2547363
12	12	Johnson High School	District	4761	3094650
13	13	Ford High School	District	2739	1763916
14	14	Thomas High School	Charter	1635	1043130

We'll take these Series and use one as the index and the other as the first column. Add the following code to a new cell and run the cell.

```
# Determine the school type.
per_school_types = school_data_df.set_index(["school_name"])["type"]
per_school_types
```

In this code, we are setting the index to the `school_name` column with the `set_index` method. This method will return a Series with the index as the `school_name` and a column with the type of school, like this:

```
school_name
Huang High School         District
Figueroa High School      District
Shelton High School        Charter
Hernandez High School     District
Griffin High School        Charter
Wilson High School         Charter
Cabrera High School        Charter
Bailey High School        District
Holden High School         Charter
Pena High School           Charter
Wright High School         Charter
Rodriguez High School     District
Johnson High School       District
Ford High School          District
Thomas High School         Charter
Name: type, dtype: object
```

Now we'll create a new DataFrame by converting this Series to a DataFrame as follows:

```
# Add the per_school_types into a DataFrame for testing.
df = pd.DataFrame(per_school_types)
df
```

When we run this cell, we get a DataFrame with "school_name" as the index and "School Types" as a column.

school_name	School Types
Huang High School	District
Figueroa High School	District
Shelton High School	Charter
Hernandez High School	District
Griffin High School	Charter
Wilson High School	Charter
Cabrera High School	Charter
Bailey High School	District
Holden High School	Charter
Pena High School	Charter
Wright High School	Charter
Rodriguez High School	District
Johnson High School	District
Ford High School	District
Thomas High School	Charter

You created the school summary DataFrame—nice work! Now we can start adding the other columns to this DataFrame.

IMPORTANT

> As we get the rest of the data for this DataFrame, we need to make sure the index for the data being added as columns is always "school_name."

NOTE

> For more information, see the **Pandas documentation on the set_index() method (https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.set_index.html)**.