4.8.2 Get the Student Count Per School

Next to the "type" of school column, Maria wants you to add the total number of students in each school. However, you have the number of students in two DataFrames: `school_data_df` and `school_data_complete_df`. Which one do you use? You'll need to get the number of students from both DataFrames and find which one has "school_name" as the index.

There are two DataFrames that have the total number of students per school: `school_data_df` and `school_data_complete_df`. First, we will use the `school_data_df` DataFrame.

REWIND

Recall that the `school_data_df` DataFrame looks like this:

Looking at the `school_data_df` DataFrame, we can see the student count is in the "size" column.

Now let's get the student count in the `school_data_df` DataFrame. Add the following code to a new cell and run the cell.

```
# Calculate the total student count.
per_school_counts = school_data_df["size"]
per_school_counts
```

When we run this cell, we get the following output: a Series showing the number of students in each school with a numerical index (0–14), which is the same as the index of the `school_data_df` DataFrame.

Unfortunately, this Series doesn't have an index with "school_name." Therefore, we can't use the "size" column from `school_data_df` to get the count of the student population.

To fix this, we can use the `set_index()` method on the "school_name" column, and then select the "size" column to display the student count for each school.

Add the following code to a new cell and run the cell.

```
# Calculate the total student count.
per_school_counts = school_data_df.set_index(["school_name"])["size"]
per_school_counts
```

When we run this cell, the output shows the `school_name` as the index and has the number of students as the data.

Next, we'll use the `school_data_complete_df` DataFrame.

REWIND

Recall that the `school_data_complete_df` DataFrame looks like this:

To get the number of students from the `school_data_complete_df` DataFrame, we can count the number of times a high school appears using `value_counts()` on the "school_name" column. The `value_counts()` method will return a Series of data with the number of times each `school_name` appears in a row.

To get the total student count per high school, add the following code to a new cell:

```
# Calculate the total student count.
per_school_counts = school_data_complete_df["school_name"].value_counts()
per_school_counts
```

This code returns a Series with the number of times the school appeared in the "school_name" column.

Now we have two methods to get the student count. As long as we make sure the index is the "school_name" column, we can use either method.

NOTE

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