

4.11.2 Categorize the Spending Bins

Establishing the bin ranges was quite a challenge, but it's done! Now Maria wants you to focus on grouping the schools' spending in the `per_school_summary_df` DataFrame based on the bins you just created.

Using our spending bins and ranges, we can create a new column in the `per_school_summary_df` DataFrame which will be assigned the spending bins from the `per_school_capita` Series.

To do this, we will need to do the following:

- Use the `cut()` function on the `per_school_capita` Series.
- Add the bin data to a new column in the `per_school_summary_df` DataFrame.

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

```
# Categorize spending based on the bins.  
per_school_summary_df["Spending Ranges (Per Student)"] = pd.cut(per_school_c  
  
per_school_summary_df
```

There's a lot going on in this code, so let's break it down.

- **To the left of the equals sign**, we add a new column called `"Spending Ranges (Per Student)"` {style="font-size: 1rem;"} to the `per_school_summary_df` DataFrame. This column will contain the values specified by using the `cut()` function.
- **Inside the `cut()` function**, we add the data (`per_school_capita`) we are going to use for `"Spending Ranges (Per Student)"`. The data must be a one-dimensional array, like a list or the `per_school_capita` Series.
- **Inside the parentheses**, we add the `spending_bins` and the labels for the bins using `labels=group_names`.

When we execute the code, the results should look like this:

per_school_summary_df.head()

	School Type	Total Students	Total School Budget	Per Student Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing	Spending Ranges (Per Student)
Bailey High School	District	4976	\$3,124,928.00	\$628.00	77.048432	81.033963	66.680064	81.933280	54.642283	\$585-629
Cabrera High School	Charter	1858	\$1,081,356.00	\$582.00	83.061895	83.975780	94.133477	97.039828	91.334769	<\$584
Figueroa High School	District	2949	\$1,884,411.00	\$639.00	76.711767	81.158020	65.988471	80.739234	53.204476	\$630-644
Ford High School	District	2739	\$1,763,916.00	\$644.00	77.102592	80.746258	68.309602	79.299014	54.289887	\$630-644
Griffin High School	Charter	1468	\$917,500.00	\$625.00	83.351499	83.816757	93.392371	97.138965	90.599455	\$585-629

The "Spending Ranges (Per Student)" column is added as the last column. Whenever a new column is added to a DataFrame, it becomes the last column. This is because each column in the DataFrame is like an item in a list.

REWIND

When we add a new item to a list, that item will always be added at the end of the list unless we specify a list index where it should be placed.

Great work! Now we need to filter the `per_school_summary_df` DataFrame in order to get the average scores as well as the percentage of students who passed reading and math for each range in the `spending_bins`.

NOTE

For more information, read the [Pandas documentation on the `cut\(\)` function.](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.cut.html) [\(https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.cut.html\)](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.cut.html)