Building previous Python and Pandas lessons, we learned about how to put together some of the concepts and operations to learn about the capabilities in pandas for manipulating data – in particular, merging and joining dataframes, series, and other objects.

In order to join two dataframes, we first need to know what fields are shared between the two files so we can identify join keys. We can determine what might be shared by calling the columns and printing them as a list.

There are a few different types of joins we can use – inner joins, outer joins, left joins, and right joins. Inner joins combines two dataframes on a join key and only returns the rows matching in both. Outer joins can be full, left, or right. Full outer joins return all the values from both dataframes and matches rows where possible. Left joins keep every row from the left dataframe while a right join keeps only the rows from the right dataframe. Though we could use for loops to obtain such a result, this can be much slower with larger sets of data and require much more code (hello, verbose Rose).

We also discussed binning and grouping, which allows us to cluster values together so we can make better sense of them. For example, binning together numerical values to form categories such as high, medium, or low temperatures.

Finally, we practiced debugging by running code to figure out why it wasn’t executing (e.g., a step needed to be added before we could run it such as defining a dataframe or importing a missing dependency).