London Housing Case Study:

**Questions for Mentor:**

* Does it matter how many times I run a jupyter notebook?
* properties2 = properties.drop(index=[0, 1, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48]) – there has to be a better way than this to get rid of a range of indexes
* is it poor form to change the name of the dataframe a bunch when I make changes?

**Observations:**

* Need to transpose data – want bouroughs to be the row labels – DONE
* Get rid of first row with the duplicative data to column labels - DONE
* Rename the borough column and the ID column – DONE
* Melt my dataframe – why? – looking at the average over the last 2 decades so create columns with average prices for a specific time resampled time frame (year, 5 years, 10 years?. Could also do pct\_change() - DONE
* Remove null values as I see fit. For the unnamed columns, I want to just delete – they are not providing any value. Also need to delete rows that are not london boroughs (i.e. cit of London and index 33-47) - DONE
* Create a line plot subsetting an individual borough by year - DONE
* Model results – write a function that takes the average price change from 20 years ago against latest year do old price/new price as the ratio – do I want to iterate through boroughs to create ratio for each borough? Currently have a set borough in formula which isn’t very dynamic – created new dataframe with just year, borough and avg price by year, will make modeling section much easier (I hope)
* Conclusion, write a conclusion describing the findings

Properties\_subset contains borough, year, average price

I want to get the 1998 avg price divided by 2018 average price in a ratio

Could iterate through dataframe to get for all boroughs?

Or could just idmax the max value of the ratios