I tried out graphs in spreadsheets to verify for data and check for outliers. By seeing the graphs I didn’t observe any dramatic changes in the lines because the time period for each data was a month which is fairly short. The fact that there was a smooth graph assured that the data did not include outliers and that no data was not misplaced during scrubbing. Regarding making changes to the data I had to remove unnecessary rows, add computed columns, implement a method that selects certain datasets among the whole, merge data from three datasets and sort them according to the dates. Although the code got complicated I used dictionaries to sort my data because it made most sense to have the dates as the key and a list of computed indexes as the value. I added a function that computes the percentage change of two time periods because I wanted to see whether these indexes had a correlation. The three datasets that I merged had different timespans so I also implemented a function that excluded years that seemed to have insufficient number of values. Also another issue I faced was, because I was appending values to the dictionary and trying to get the columns aligned, I had to deal with putting empty values as well as ways to skip over empty values when iterating through to write to the output file.

First of all, the data source “data planet” seemed credible because it had a vast amount of various economic indexes. I selected these data sets because I wanted to see whether there was a correlation between these indexes. I wanted to see an inverse relation between federal interest rates and unemployment rates. If these indexes were a sign of economic welfare, then it would also have a relation to number of houses beginning construction. I think this data can be used to generally anticipate real estate prices because if number of houses beginning construction increase as a result of economic wellbeing the supply in the market will increase thus pulling up market prices under the premise that demand stays fairly steady.

URL to datasets

<https://statisticaldatasets-data-planet-com.proxy.library.nyu.edu/dataplanet/Link?id=cf54995796>

<https://statisticaldatasets-data-planet-com.proxy.library.nyu.edu/dataplanet/Link?id=cf53d4ddc4>

<https://statisticaldatasets-data-planet-com.proxy.library.nyu.edu/dataplanet/Link?id=cf5593dd77>