Summary Report for Unit 4 Challenge – London Price

* **From the mini project about London price, the final plot is:**

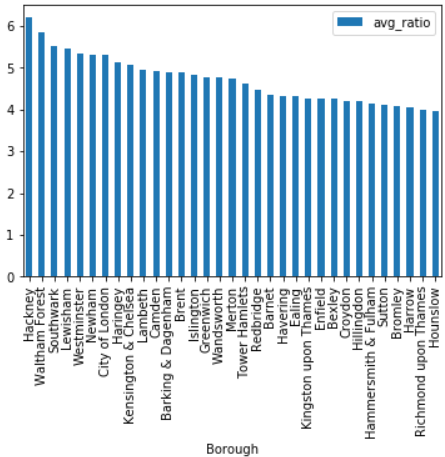
****

Figure: Average ratio of London house price for different boroughs over past two decades

**I found:**

(1) Borough Hackney of London had the greatest house price increase, on average with the ratio around 6.20 over the past two decades.

(2) Except borough Hounslow (ratio is 3.98), all borough's price increase ratios were over 4. This indicated almost house price in London boroughs had quadrupled in last two decades.

* **I used four parts in the whole process to arrive my conclusion.** **These parts are:**

1. Sourcing and loading

2. Cleaning, transforming, and visualizing

3. Modeling

4. Evaluating and concluding

In the first part, I loaded the data by *xls* format online.

In the second part, there were several sub-steps to clean and prepare the data. In sub-step 1, I initially explored the data by *head*. In sub-step 2, I cleaned the data by transpose, reset index. In sub-step 3, I renamed the column names. In sub-step 4, I transformed the data by melt (this is very important since we want less columns in the data in our analysis process). Also changing the data types in the relevant process is important. In sub-step 5, because some items in column *London\_Borough* is not the actual boroughs of London, so I selected part of them to detect and found the 'ID' and 'Average\_price' of these boroughs are NAs, then I dropped all NAs in the columns of 'ID' and 'Average\_price' by function dropna() (another choice is notna()). In sub-step 6, I viewed one the boroughs by plotting. I also used lambda function to transform monthly notation to yearly notation for further average price ratio analysis.

In the third part, I defined a function called "create\_price\_ratio" to calculate the ratio of the price for last 20 decades in London Boroughs first, then applied that function to all boroughs in London and sorted the result to get the bar plot.

In the last part, I made a conclusion about what I found.

* **The main challenges I encountered were data cleaning and transforming.**

The data was not cleaned for me to handle it and I were not sure what were the exactly columns need to be cleaned or transformed, so I checked tier 2’s jupyter notebook and got some hints, I had also reviewed some parts of the DataCamp courses (such as “Python Data Science Toolbox”) to overcome my weakness. The weakness I might not overcome is the work efficiency, it would be better to use less time to finish the work.

* **Something I would like to investigate deeper:**

(1) Why Borough Hackney of London had the greatest house price increase in last two decades? (what factor(s) decide that?)

(2) Almost house price in London boroughs had quadrupled in last two decades by the plot. What were the reasons behind that? Who made the high price? Foreigners or natives? The borough with more schools and more convenient traffic had higher house price, was that true?