

Conclusion/Analysis:

- What did you find? Which borough is the most expensive? Any other interesting trends?

An individual who has dreams of visiting London one day, I was delighted to come across the topic of Boroughs and examine the average housing prices of each district. By looking at multiple line plots I can tell that as the years progressed, the housing prices increased (expected). Kensington & Chelsea, Westminster, and City of London (not a borough) are the top 3 most expensive districts. This makes sense because they are all located at the heart of the city. Barking & Dagenham, Bexley, and Newham are the least expensive boroughs. According to Wikipedia, these are in the municipal and county boroughs making it much cheaper. Looking at the years, 2018 was the most expensive and 1995 was the least expensive. Like New York, the closer you live to downtown (heart) the more expensive your living is.

- How did you arrive at your conclusion?

After getting a basic understanding of the python toolbox, I used aggregated functions from the manipulating data frames course to get insight. Prior to even analyzing the information, I needed to tidy the data.

- What were the main challenges you encountered? How did you overcome them? What could you not overcome?

Data never comes clean and simple. After reading the prompt and peeking at the data frame, I had immediately realized that there needed a lot of tidying to be done.

- Challenge 1: Organizing the data frame
 - Action: Firstly, I needed to organize the table in a way that is more readable by transposing the frame. Having the Boroughs as rows and dates as columns was much better than its initial look. Instead of having 306 columns of every month of every year, it was going to be much easier to examine the information after melting the table into separate columns for date (month/year) and average price while the rows remained as borough names.
- Challenge 2: Dealing with null values
 - Action: After glancing at the indexes, I realized there were unnamed Boroughs and each of them has null values for both ID and average price. To resolve this issue, I decided to store all the values that were not null into another data frame.
- Challenge 3: Identifying the proper Boroughs of interest
 - Action: Both the problem statement and Wikipedia page shared state that there only exist 32 boroughs within greater London. Looking at the current data I have, there were 45 boroughs listed. To remove all the nonboroughs, I created a list and negated the list to store it into a new data frame. Now, we have our finalized data table

- Is there anything you would like to investigate deeper?

I also wanted to observe which month, if not the same, has the most expensive price. With more features, I can check if there exists a correlation between the price and location. Additionally, I want to do more data visualization if possible.