## **4.3 Case Study - London Housing Price Increase**

### **Introduction**

In this unit of the Springboard Data Science course, the rise in housing prices in London was analyzed. Namely, the ratio of average prices in 2020 and 2000 was compared for 32 boroughs of London.

### **Methods**

Data was imported from the [London Datastore](https://data.london.gov.uk/): a free, open-source data-sharing portal for London-oriented datasets. It is in Excel format with the date in the first column, and the other columns containing average housing prices. Column 2 is for the City of London, which is the historic centre, a separate ceremonial county, and *sui generis* local government district that functions quite differently from a London borough [[https://en.wikipedia.org/wiki/London\_boroughs]](https://en.wikipedia.org/wiki/London_boroughs%5D). The next set of 32 columns corresponds to the 32 boroughs of London, sorted alphabetically (including 12 Inner London boroughs and 20 Outer London boroughs). The remaining columns contain the same type of data for larger administrative units of London and the UK.

Data cleaning included the removal of empty columns that are used as delimiters between the groups of boroughs and larger administrative units.

### **Results**

The leader in housing price rise during the last 20 years is the London Borough of Hackney with the ratio of 4.387 (Fig. 1). According to the travel website [<https://theculturetrip.com/europe/united-kingdom/england/london/articles/a-guide-to-hackney-east-london/>], “A former industrial hub, East London’s Hackney has undergone huge regeneration in recent years to become the city’s most liveable borough. Home to young creative types, it’s an area with strong community ties and is known for its great food and unmatched nightlife.”

The next borough in descending order is Waltham Forest with the ratio of 4.368. Across all boroughs, the mean ratio value is 3.421 and the standard deviation is 0.386. The frequency distribution is skewed right, which means that there are few boroughs with the largest ratio values and more boroughs on the smaller ratio side (Fig. 2). Kurtosis is small, which means that the maximal and minimal values are not very far from the mean value.

The most expensive is the Borough of Kensington & Chelsea with the average housing price of £1326363 in 2020. Its ratio of prices is 3.520, which is slightly greater than the mean ratio across boroughs.

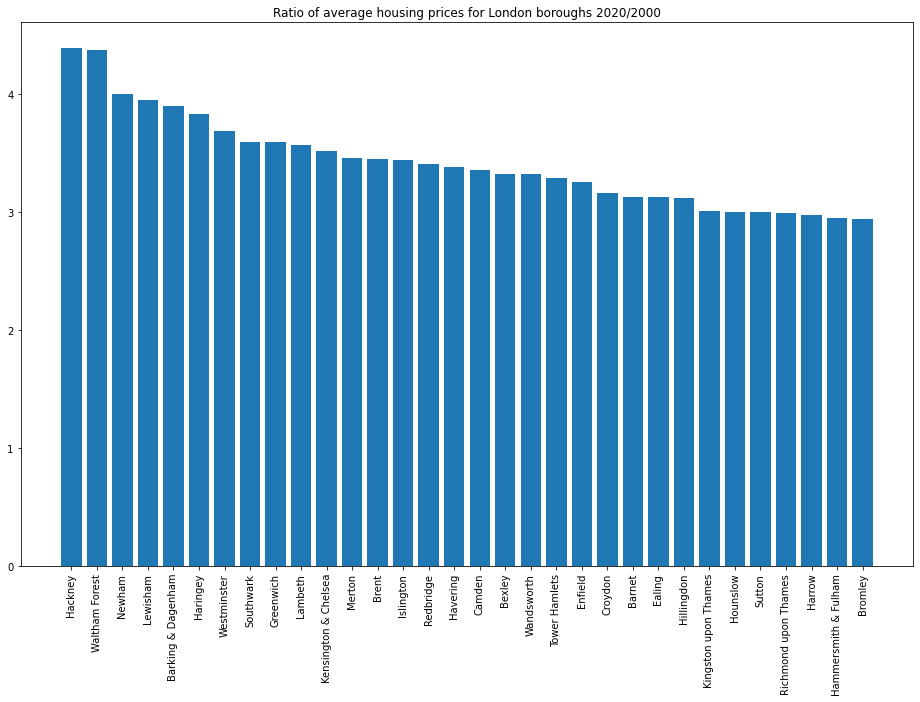


Figure 1. The ratio of average housing prices in 2020 to 2000 for 32 London boroughs.

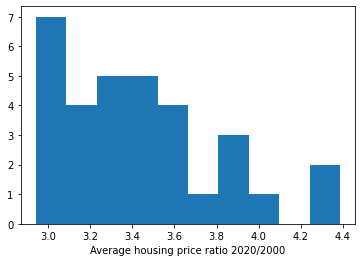


Figure 2. Histogram of the average housing price ratios.

### **Discussion**

#### **Analysis approach**

Data analysis was performed using Jupyter Notebook software in Python 3. The following packages were utilized: pandas, numpy, and matplotlib.pyplot. The data cleaning path can be slightly optimized by first removing the irrelevant columns, which is easier to do by taking advantage of the fact that the 32 columns of interest are contiguous, as opposed to using a set of non-borough column names later after transposition of data. To facilitate the possible time dependence analysis of the price ratio in the future, I enhanced the create\_price\_ratio function by adding the two years as arguments. For easier comparison of price ratio values for different boroughs, the ratio values were sorted and presented as a bar plot rather than a line plot.

#### **Challenges**

The hardest part for me was understanding the concept of an index and working with indexes. Instead of resetting the index in the final dataframe ‘dfg’ and using a dictionary, I preferred to use tuples. The most time-consuming part was trying to understand why it is beneficial to transpose and melt data, given that at first it seemed easier to just filter the rows and columns to extract prices for all months of the reference and recent years, and then average by grouping by year.

#### **Possible follow-up studies**

It is worth analyzing the time dependence of the ratio of average prices to compare the dynamics of price changes for different boroughs. This could serve as a starting point for a detailed study of the reasons for the increase in prices and their differences across boroughs. The approach that would be developed could be applied to other communities, and the results of relevant studies could be compared to gain even more insight into the dynamics of house prices and the dynamics of prices in general.