Case Study - London Housing

Libi Voshin

Springboard

This project will focus on answering the following question:

Which boroughs of London have seen the greatest increase in housing prices, on average, over the last two decades?

Data Source:

The London Datastore: a free, open-source data-sharing portal for London-oriented datasets.

CAREER TRACK UNIT 4 CHALLENGE

Introduction

The main goal of this challenge is to apply the skills that have been learned, bridging the gap

between the controlled environment of DataCamp and the work that data scientists do with actual

datasets.

In this challenge, we're going to solve: ***which boroughs of London have seen the greatest***

increase in housing prices, on average, over the last two decades?

There are 32 boroughs within Greater London. Some of them are more desirable areas to live in,

and the data will reflect that with a greater rise in housing prices.

Method

1. Sourcing and loading
2. Cleaning:
   - Checking for NA values across the data
   - Deleted columns "unnamed: 33", "unnamed: 36", "unnamed: 46" containing NA
   - Resetting index to Range index instead of "Borough"s column
   - Making sure we have an equal number of observations to final columns
   - Filtering and getting rid of all data that is not concerning the 32 boroughs of London
   - Filtering and getting rid of all observation rows that are not between 1998 and 2018

   Transforming:
   - Transposing the DataFrame so that each borough will be presented as a row instead of a column
   - Renaming meaningful columns names ("London Borough" and "ID")
   - Melting the data by 2 main columns: London Borough and ID.
   - Renaming columns: Month and Avg_price for the melted data frame.

   Additional data manipulations:
   - Limiting the number of data points, by extracting the year from every month value
   - Grouping-by Borough and year to better present wanted data

3. Visualizing
   Modeling:
   - Creating a function that calculates a ratio of house prices for each borough, comparing the average price of a house in a particular year to the price in a different year. In our case, we were comparing the 2018 price to 1998.
   - Creating a function that creates top X boroughs that had the greatest increase in housing prices over a chosen period of time, and plots them in a combined plot-line.

4. Evaluating and concluding

Visualizing the Data

Fig.1:

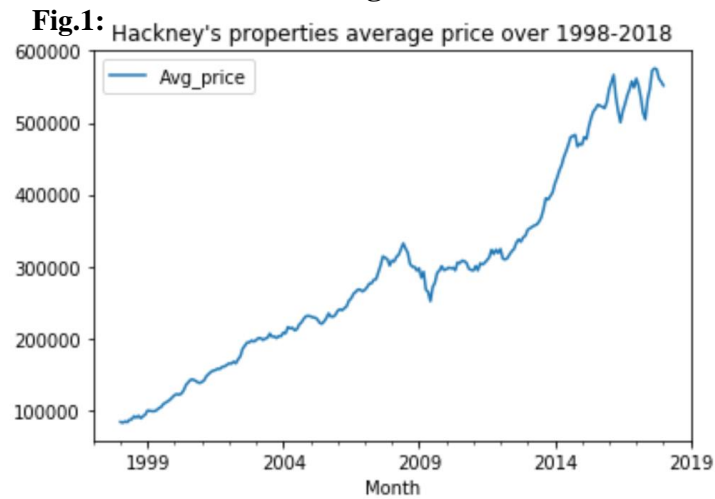Hackney's properties average price over 1998-2018

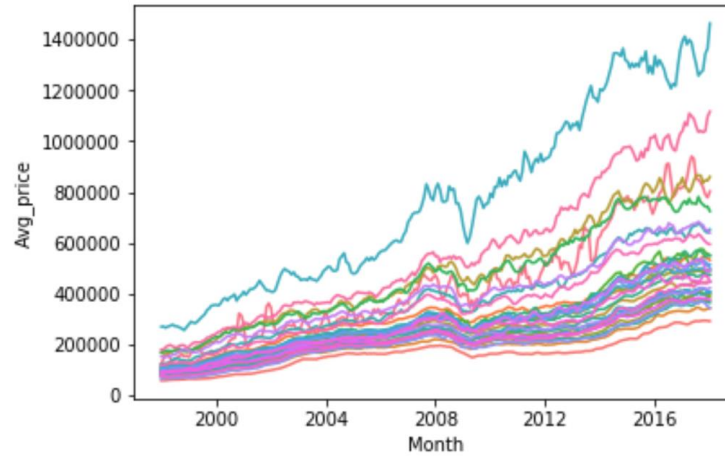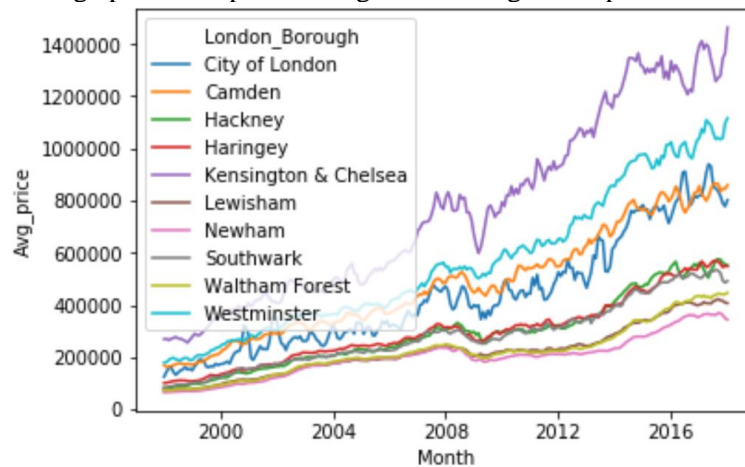**Fig.2:** General look at Properties average price of London's 33 boroughs over 1998-2018

Fig.3: Properties average price of top 10 boroughs with the greatest price increase over 1998-2018

Results

Insight 1

The borough that had the greatest increase in housing prices, on average, between 1998 and 2008 is **Hackney** borough.

Insight 2

The top 10 boroughs that have seen on average the greatest increase in housing prices between1998 and 2008 are:

	London_Borough	Price_Ratio
0	Hackney	6.278842
1	Waltham Forest	5.913104
2	Westminster	5.857866
3	Kensington & Chelsea	5.453241
4	City of London	5.418622
5	Lewisham	5.391205
6	Southwark	5.383056
7	Haringey	5.106149
8	Newham	5.080441
9	Camden	5.036017

Insight 3

The boroughs that have seen on average the greatest increase in housing prices are not necessarily the most expensive boroughs, as we can see by 'figure 3' and 'Insight 2'.

Discussion

Directions in which we can continue to investigate data further:

- We can check whether there is any correlation between the increase in property price and the initial borough's properties average price. For example, whether the boroughs that showed the greatest price increase were the ones that were also in demand in 1998 or the opposite.

CAREER TRACK UNIT 4 CHALLENGE

- We can load the additional data that is in our Excel file source and use the data about Sales volume by borough and date. We could check whether there was any influence of the price increase on the property's sales volume and if there was, how did it influence it.

Additional Challenges and How they've been overcome

- Compare multiple trends through different boroughs on the same line plot -
  Mentor introduced me to a Seaborn library.

- Having a deeper understanding of how to approach the following question: What will be the best way to visualize current data.

- Knowing which questions should I ask myself before choosing a certain plot (kind of plot)

- Working with Git and GitHub and performing commits - Mentor will have a separate meeting on this topic, at our next meeting.