Greater London Housing Prices

Nigel Brown

2020/10/13

Contents

1 Introduction

The goal of this project is to predict house prices in the Greater London Area based on the type of property and the area of London that the property resides in.

The dataset for the project is made from data sourced from HM Land Registry and FreeMapTools. This data has been pre-processed into a dataset titled lhd.rds. The data chosen is a 10 year period 2009-2019. The reason 2020 has been left out of the data is that is an incomplete year at the time of analysis. The data

contains 4 different types of property detached; flats/maisonettes; semi-detached and terraced. It doesn't contain the number of rooms or the total area of the property, therefore a 2 bedroom flat is categorized the same as a 4 bedroom flat and a prediction can only be made on the property type of a given district.

The price paid data set from the HM Land Registry is joined together with spatial data to add latitude and longitude for the properties based on their postcodes. This will allow for mapping of the data.

2 Data Wrangling

As a first step data, that is in logical form (true/false) is converted into integers 1 or 0. Next all columns containing character data are factorized.

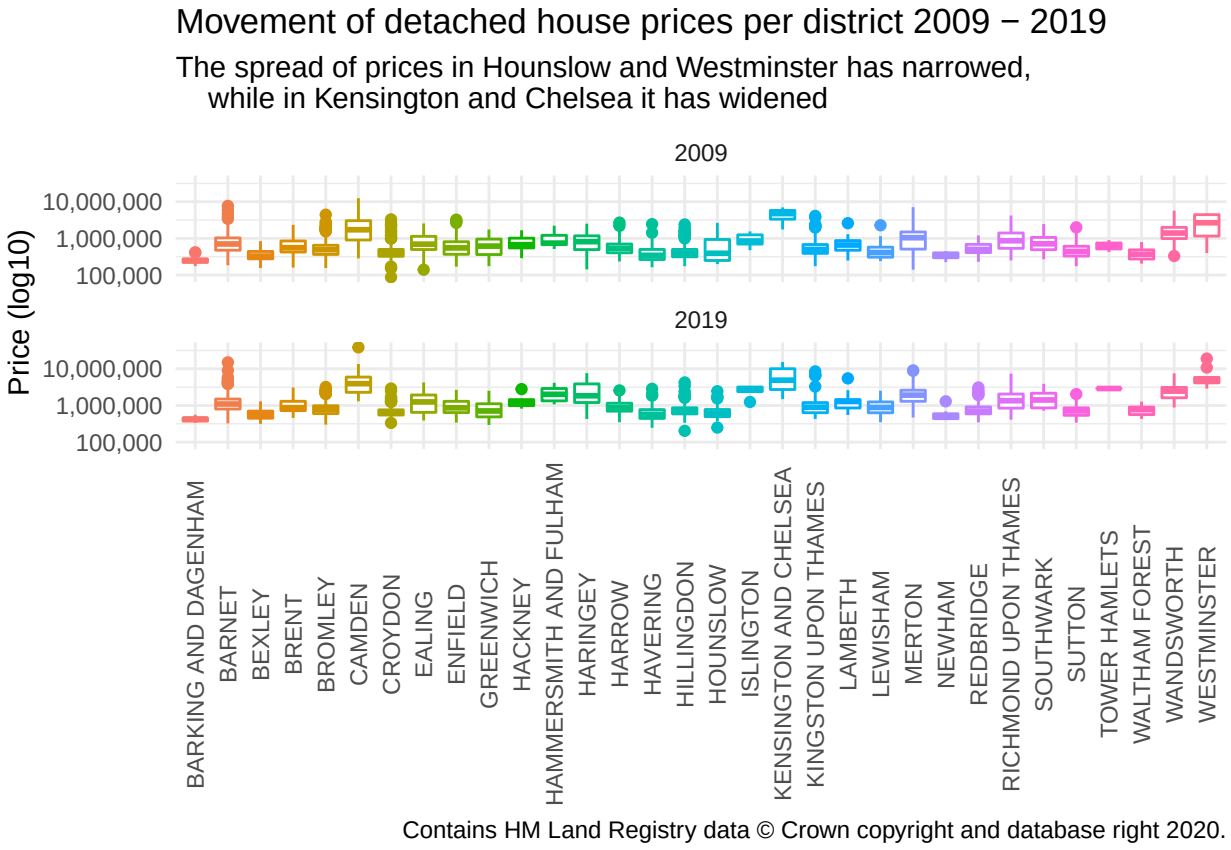
2.1 Features added to the data

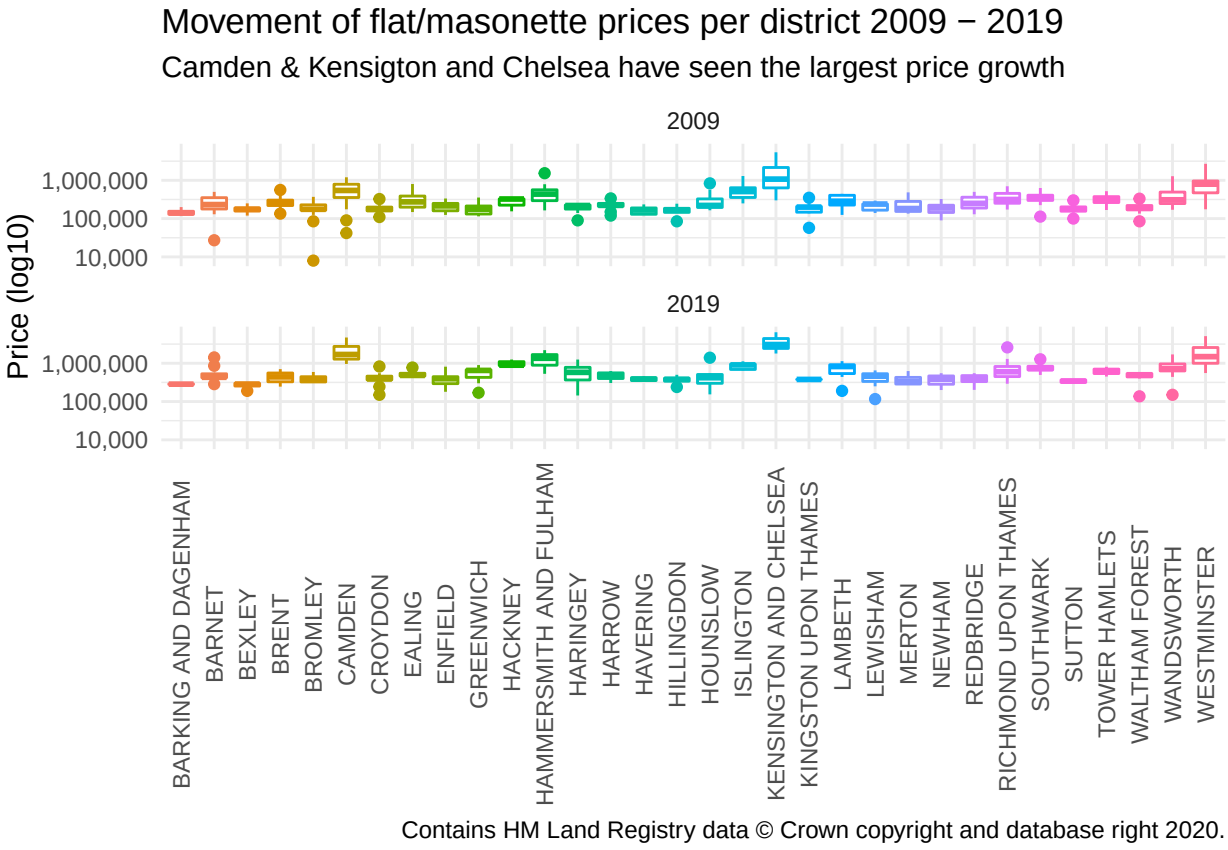
Next a new feature is created based on the outward code contained in a postcode i.e the sub-district and another feature containing the number of times a property was observed to be sold in the 10 year time period is added.

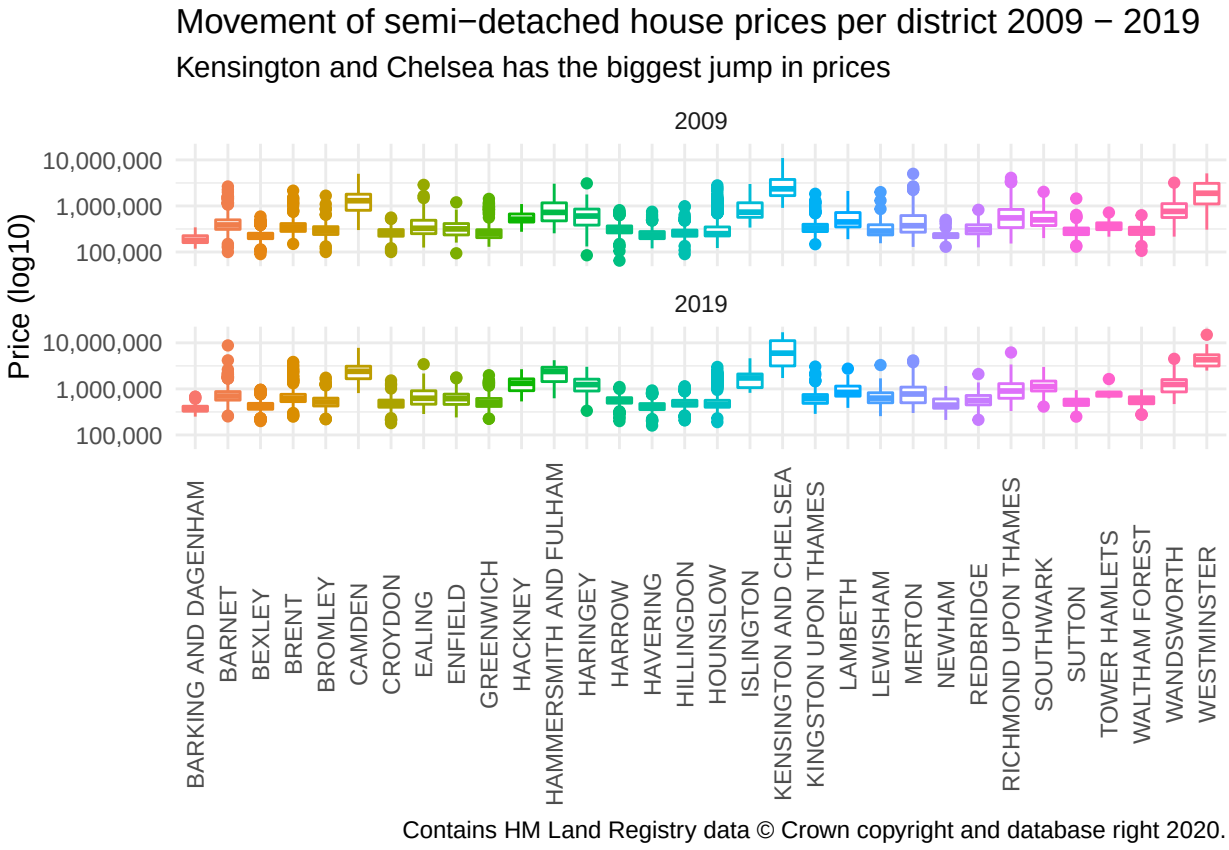
3 Exploritory Data Analysis

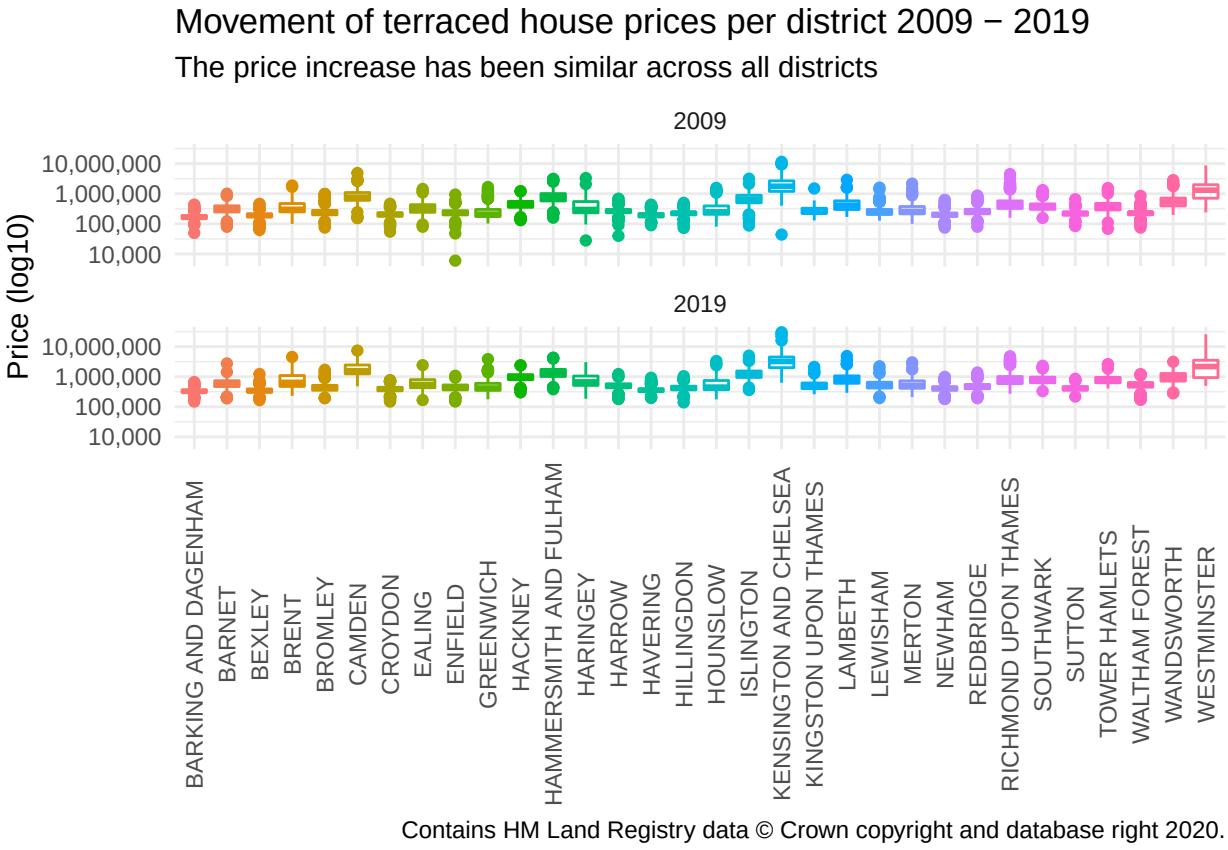
Now that the dataset is cleaned, an exploration of the data is performed. The data consists of 477382 rows of data, each row observes a single sale. The are 12 features in the dataset, with price being the outcome feature.

First the price movement for each property type is plotted





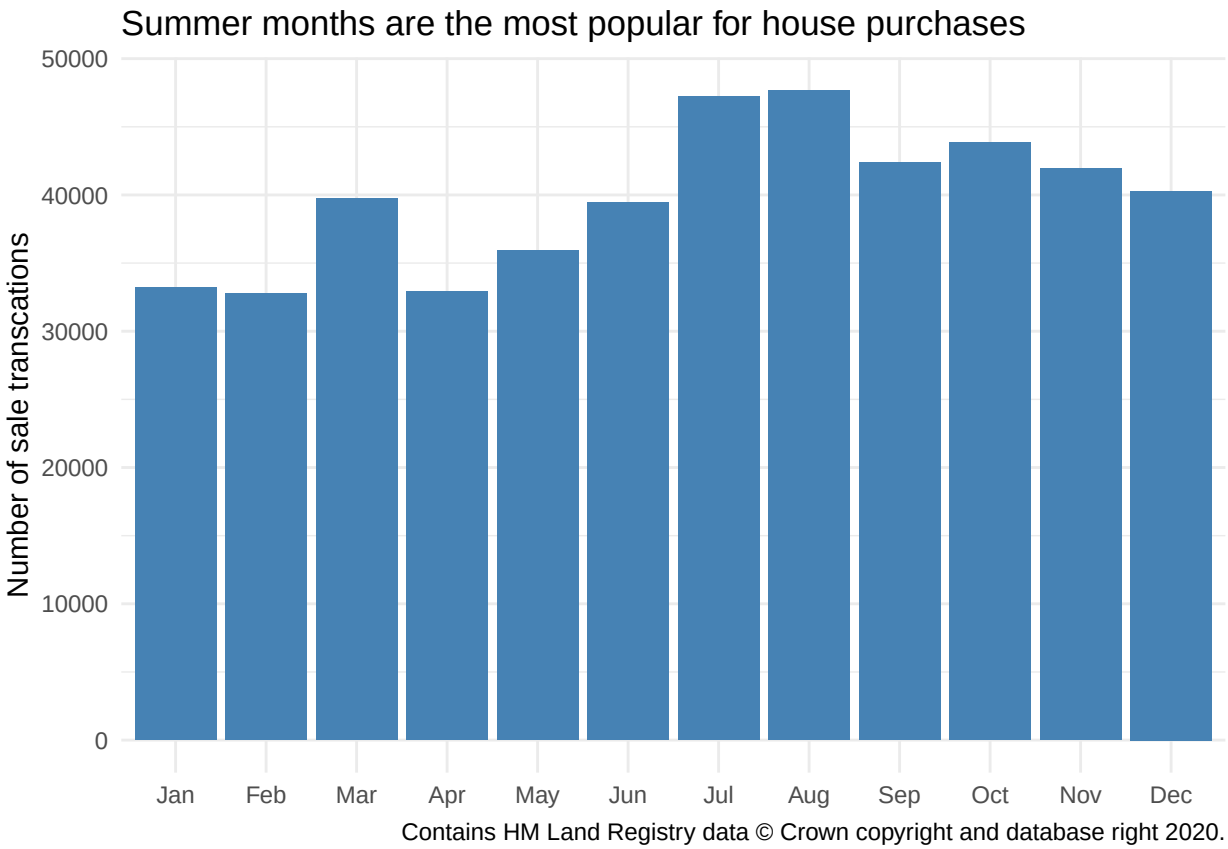




3.1 How does the avg price increase year on year ?

3.2 Is there a trend of which month most sales occur on ?

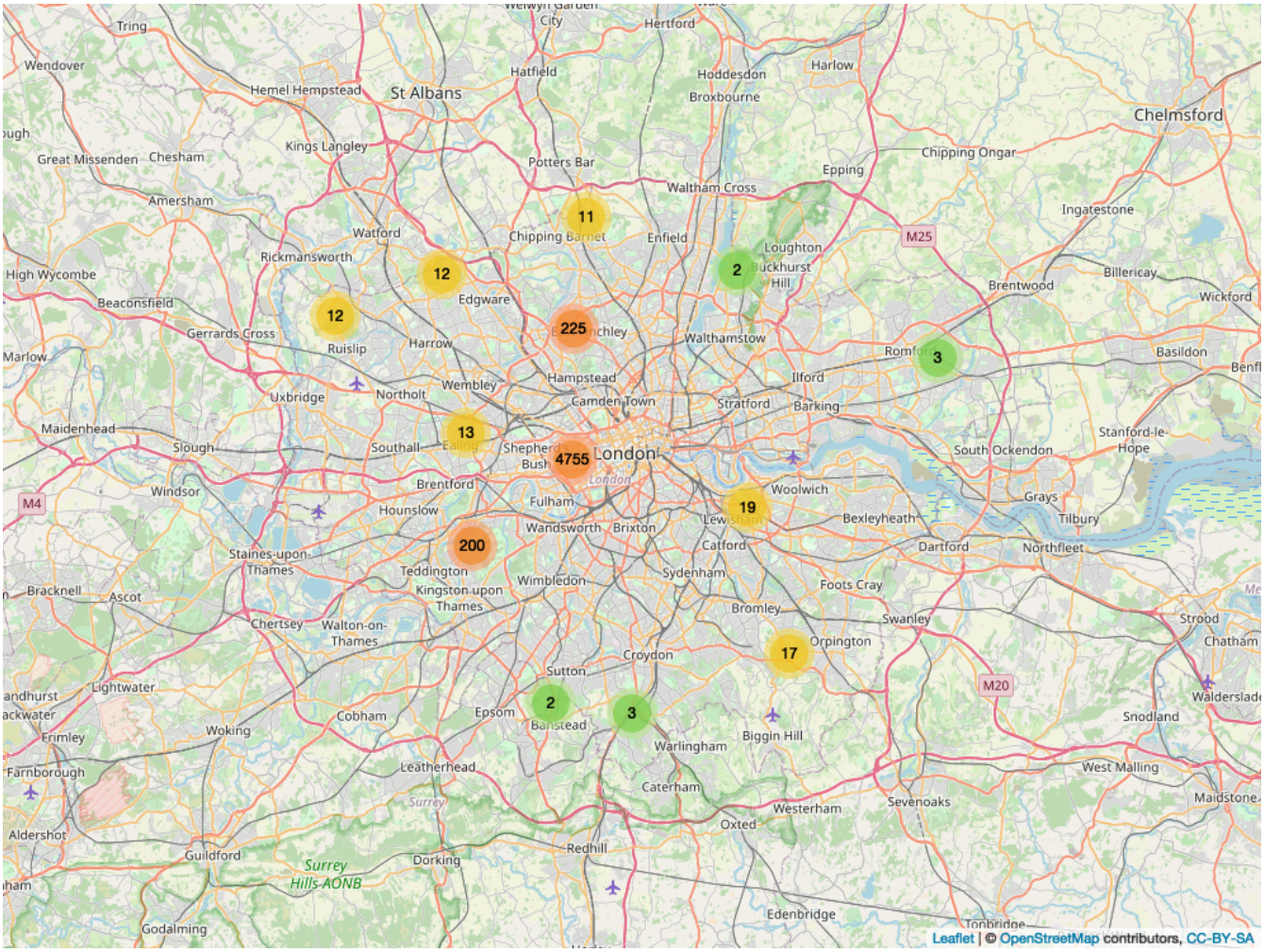
```
## `summarise()` ungrouping output (override with `.groups` argument)
```



3.3 Which are the 10 most expensive districts based on the number of properties sold over 3 million GBP?

district	properties_greater_than_3_million_GBP
KENSINGTON AND CHELSEA	1927
WESTMINSTER	958
CAMDEN	648
MERTON	287
RICHMOND UPON THAMES	268
WANDSWORTH	225
BARNET	217
HAMMERSMITH AND FULHAM	211
HARINGEY	102
ISLINGTON	96

3.4 How are properties sold for 3 million GBP or greater clustered ?



4 Which are the 10 least expensive districts based on the number of properties under 100000 GBP?

district	properties_less_than_100K_GBP
BEXLEY	31
WALTHAM FOREST	22
NEWHAM	21
ENFIELD	20
CROYDON	19
BROMLEY	17
HARROW	16
EALING	13
HARINGEY	12
BARNET	11

5 Methods/Analysis

6 Results

7 Conclusion

8 Disclaimers

9 Appendix