

Choose your own project submission

Nigel Brown

10/7/2020

Contents

1	Introduction	1
2	Data Wrangling	2
3	Exploritory Data Analysis	2
3.1	How does the avg price increase year on year ?	6
4	Methods/Analysis	6
5	Results	6
6	Conclusion	6
7	Disclaimers	6
8	Appendix	6

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 from 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 number of rooms or the area squared of the property, however some feature engineering is performed to try to improve results and enhance the data.

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.

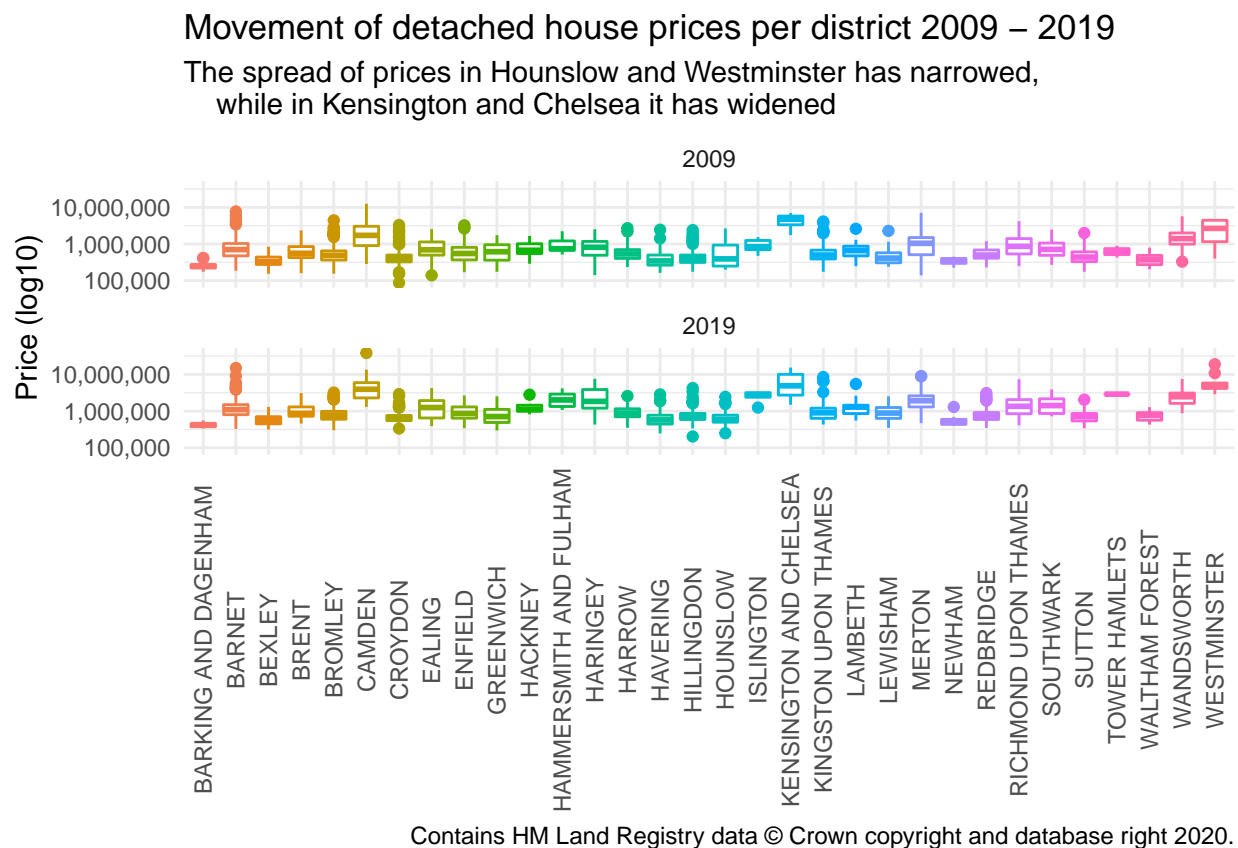
2 Data Wrangling

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. 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 Exploratory Data Analysis

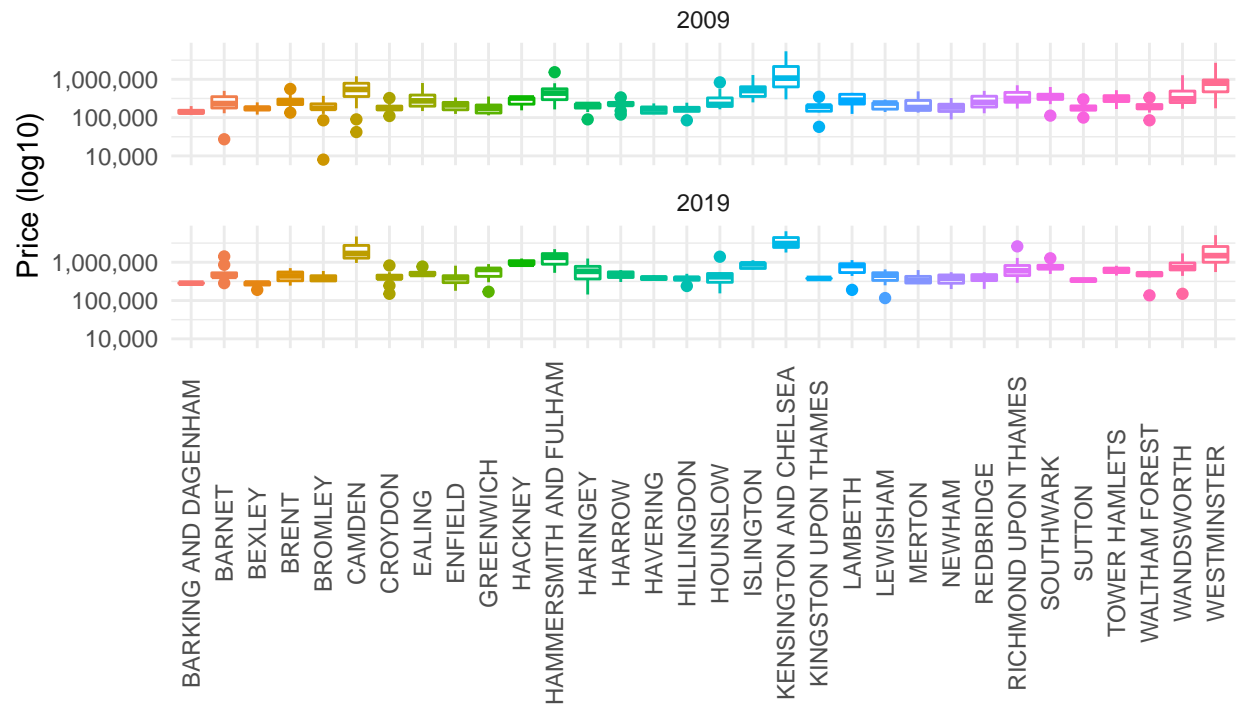
Now that the dataset is cleaned, an exploration of the data is performed. The data consists of 477382 rows of data, each row observes a single sale. The are 12 features in the dataset, with price being the outcome feature.

First the price movement for each property type is plotted



Movement of flat/masonette prices per district 2009 – 2019

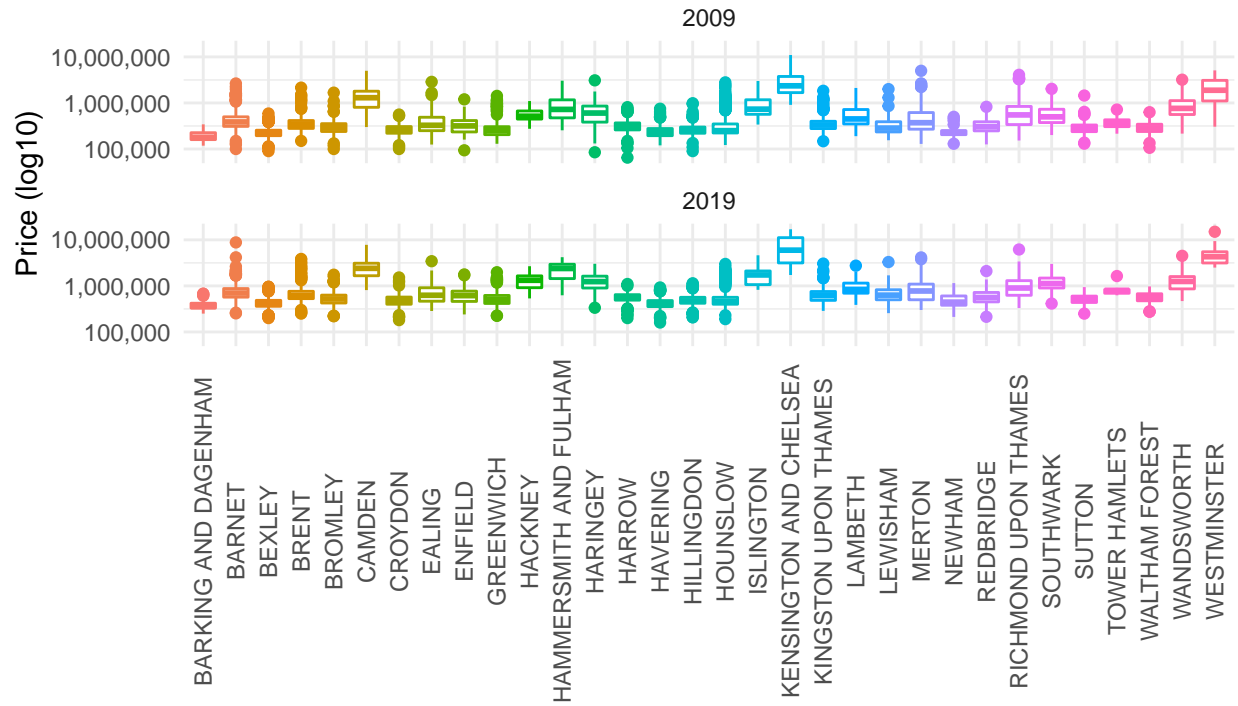
Camden & Kensington and Chelsea have seen the largest price growth



Contains HM Land Registry data © Crown copyright and database right 2020.

Movement of semi-detached house prices per district 2009 – 2019

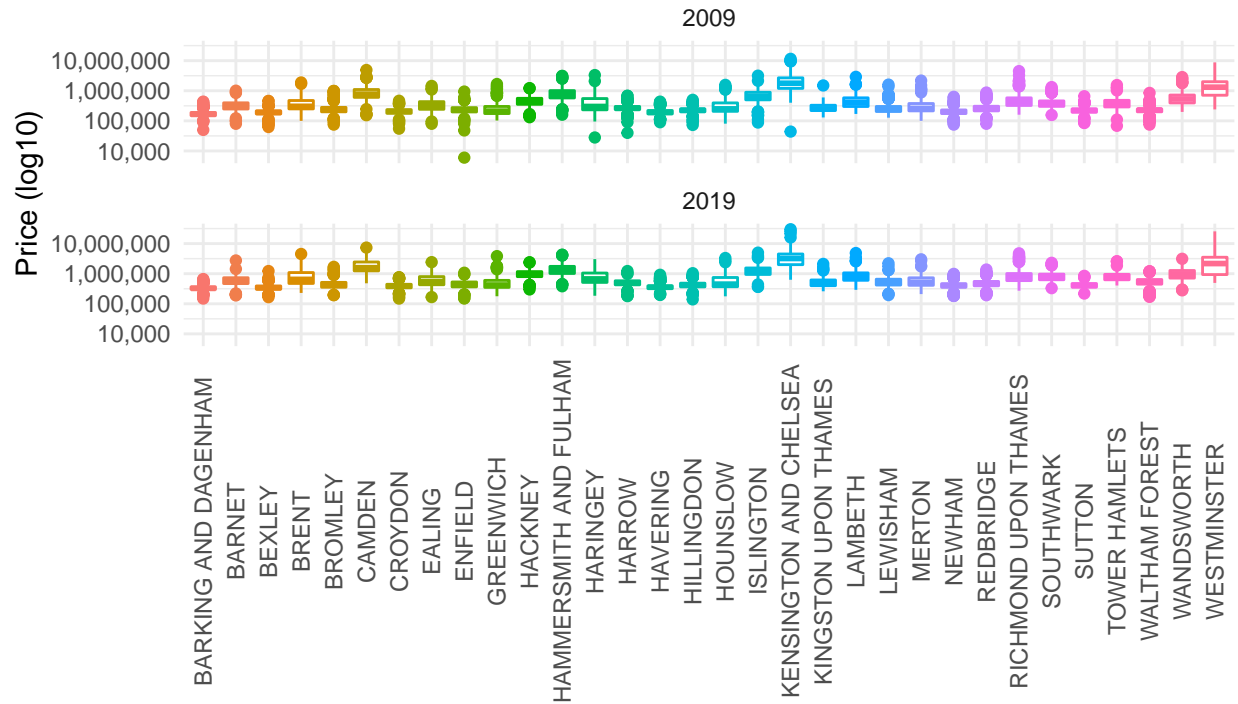
Kensington and Chelsea has the biggest jump in prices



Contains HM Land Registry data © Crown copyright and database right 2020.

Movement of terraced house prices per district 2009 – 2019

The price increase has been similar across all districts

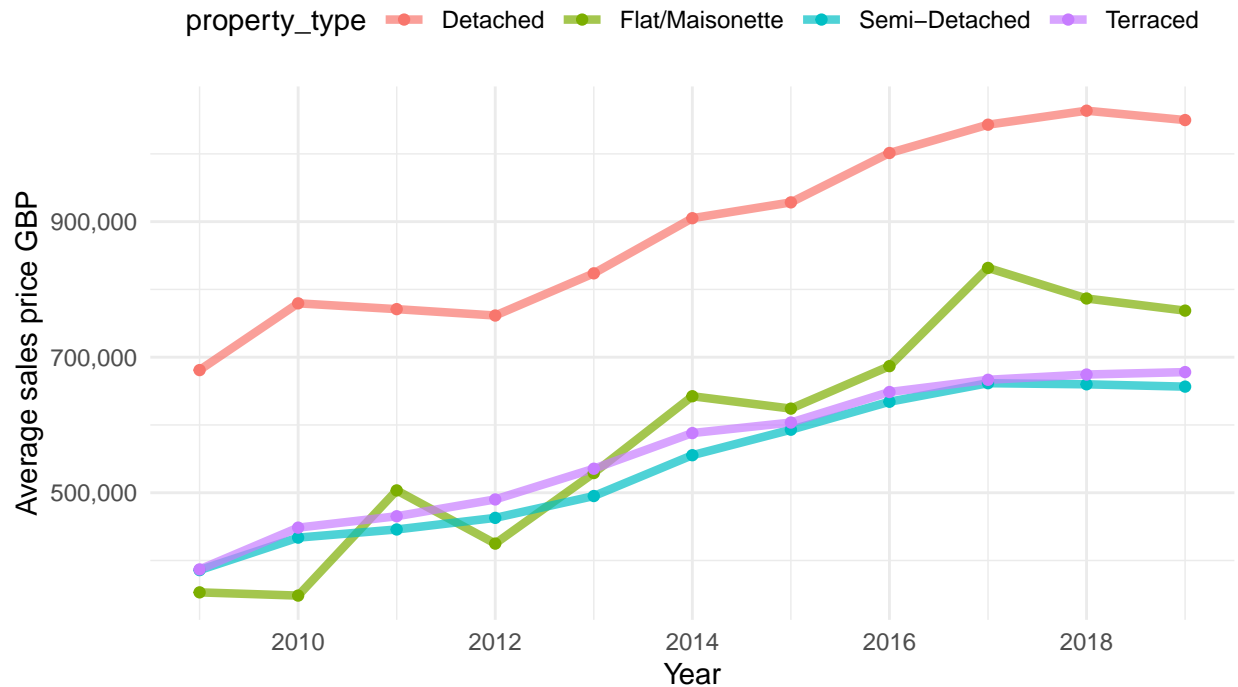


Contains HM Land Registry data © Crown copyright and database right 2020.

3.1 How does the avg price increase year on year ?

The increase average housing cost by type in Greater London

A distinct slowdown in the increase since 2017



Contains HM Land Registry data © Crown copyright and database right 2020.

4 Methods/Analysis

5 Results

6 Conclusion

7 Disclaimers

Contains HM Land Registry data © Crown copyright and database right 2020. This data is licensed under the Open Government Licence v3.0

FreeMapTools

Contains Ordnance Survey data © Crown copyright and database right 2020.

Contains Royal Mail data © Royal Mail copyright and database right 2020.

Source: Office for National Statistics licensed under the Open Government Licence v3.0

8 Appendix