Investigation of Housing Price to District Amenities in Central London

Author: Yan Bin Pan

23 Aug 2020

Project Aims

- The project aims to investigate the correlation between housing price in each neighbourhood in London with local amenities density and type.

- The methodology developed would be able to identify overvalued and undervalued neighbourhoods in London based on amenties similary between different class of neighbourhoods

Data sources

- Location data of local amenities in London is obtained from the Foursquare API sandbox. This will be used to retrieve data regarding the most frequent and popular amenities by type for each neighbourhood and use this data to train the machine learning algorithm.
- Housing Data index will be scraped from the London Datastore and other sources affiliated regarding housing price data and district boundaries coordinates.
- Price data is web-scraped from Foxton public available data in London and used to assign market value price data to each of the neighbourhoods in London.
- London postcodes and neighbourhood data has been web-scraped from the wikipedia page, this will be used to identify and categorise each neighbourhood in London.

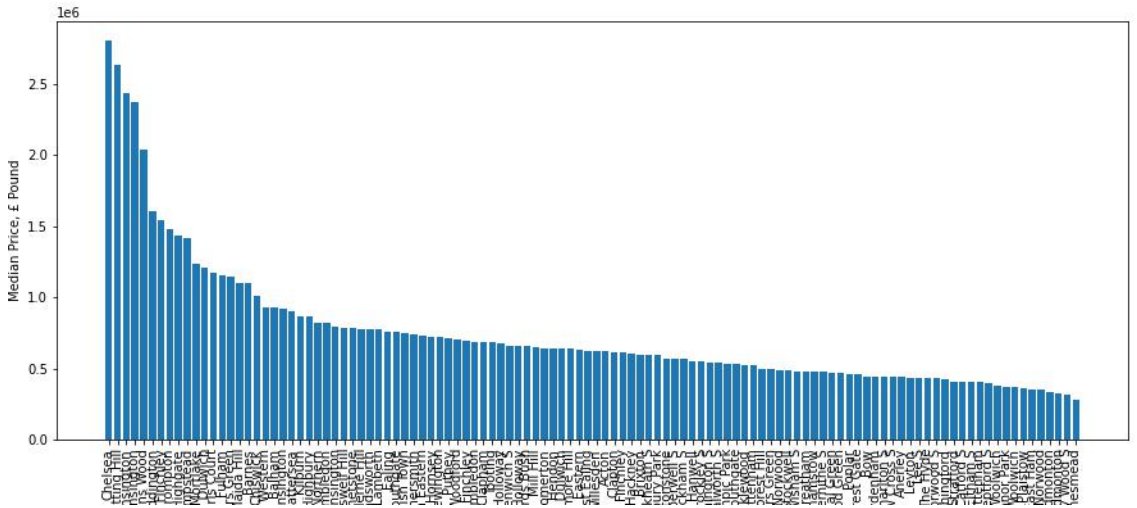
Data Engineering

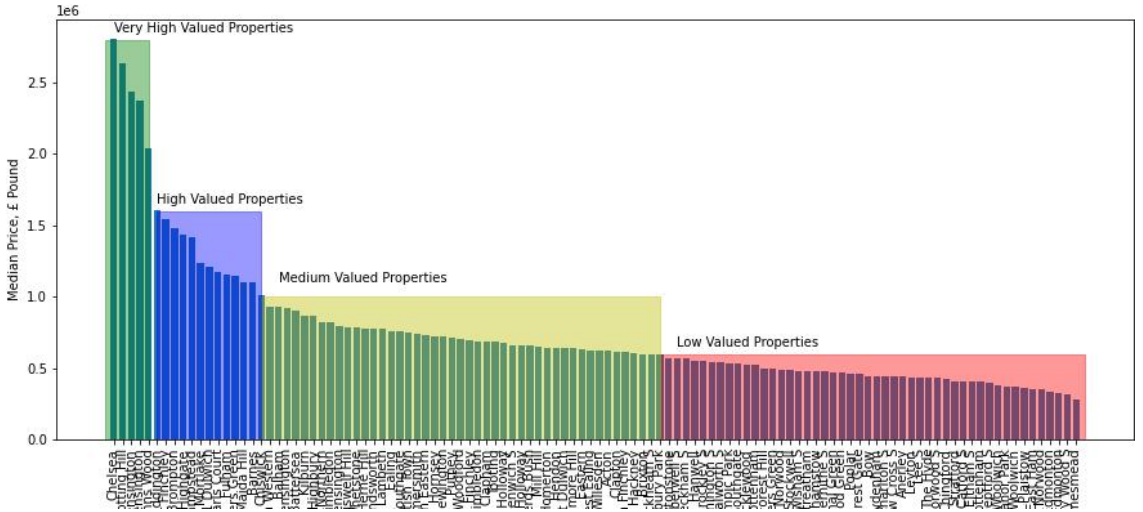
- The data extracted from the different data sources have been cleaned and formatted as to include the following information per neighbourhood
  - Coordinates
  - Postcodes
  - Market median price
  - Top 10 local amentities
  - Estimation of market based price category

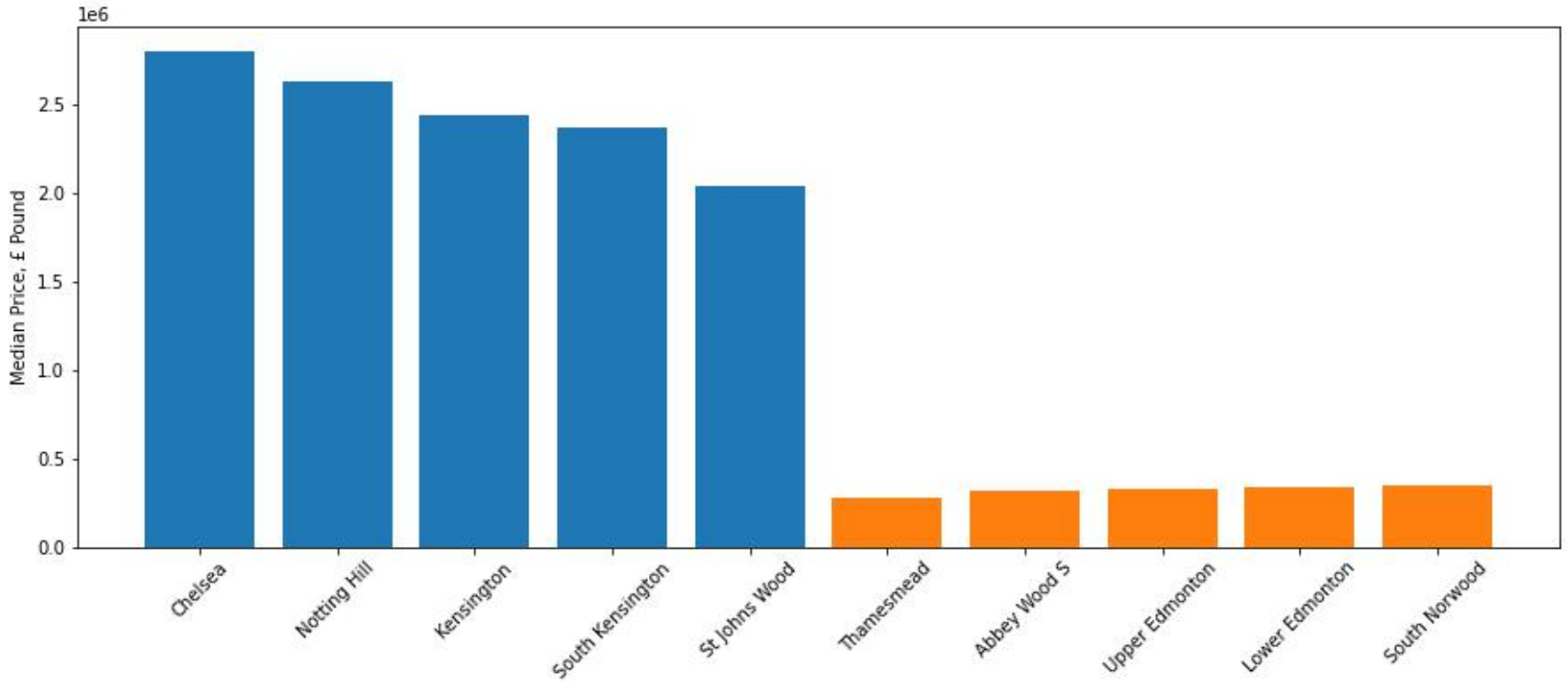
Methodology

- Divide our original dataframe now containing information of market price category and amenity based price category into 4 subsets defined by the market based price category.
- For each subset dataframe I would extract the 75% percentile amenity based cluster label for each neighbourhood.
- A comparison between the extracted labels and the market based price category would then be carried out.
- The code would then plot to identify any discrepancy between the predicted amenity based value and the current market based price value

Data Understanding

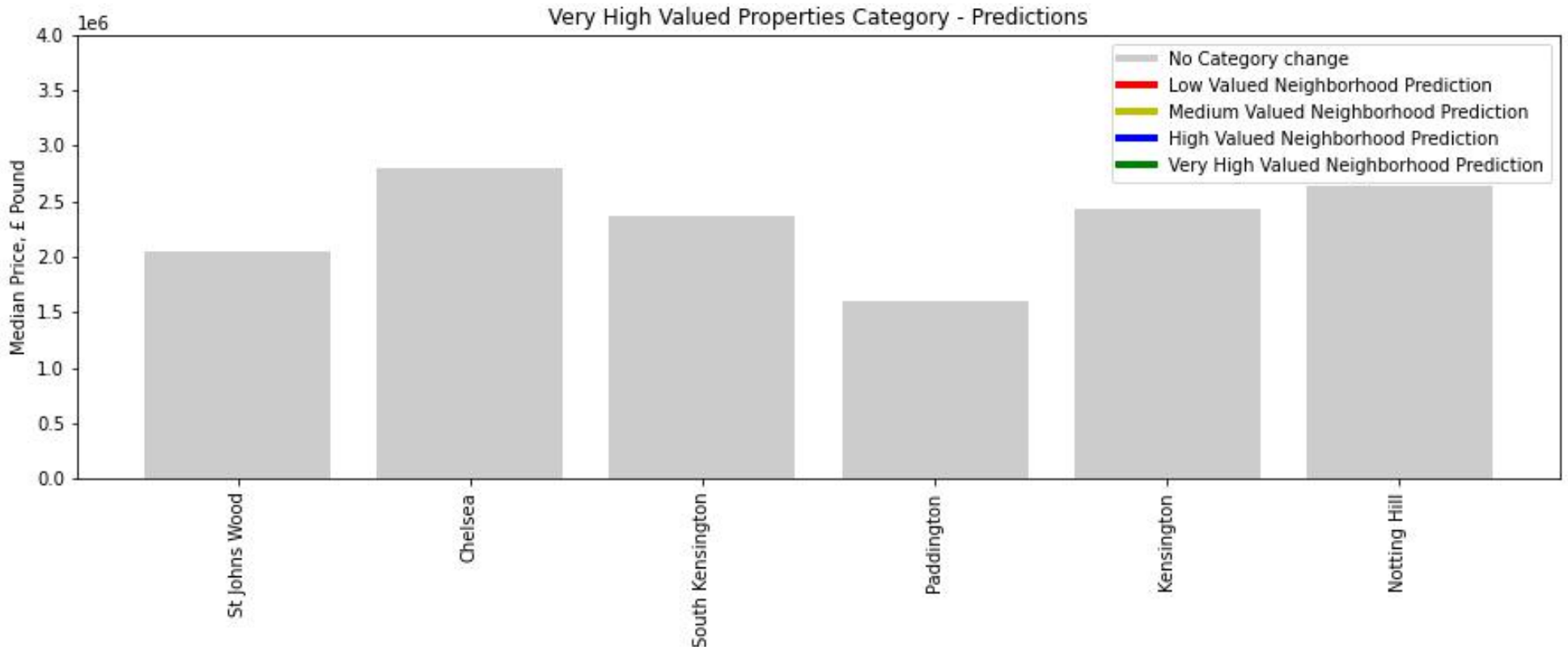





1e6
2.5
2.0
1.5
1.0
0.5
0.0
Median Price, £ Pound
Chelsea
Notting Hill
Kensington
South Kensington
St Johns Wood
Thamesmead
Abbey Wood S
Upper Edmonton
Lower Edmonton
South Norwood

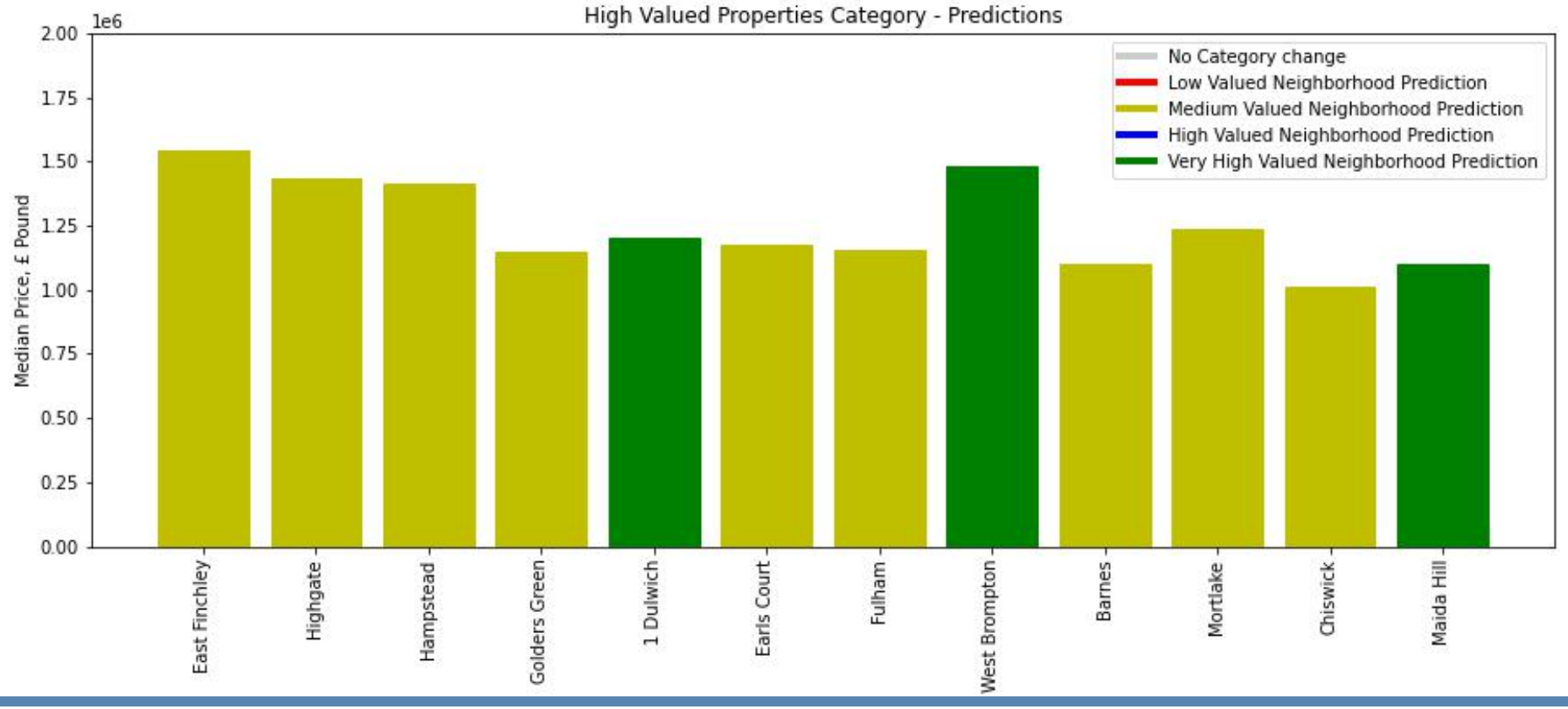
Very High Valued Neighbourhoods Predictions

- Perfect agreement between market and model predictions



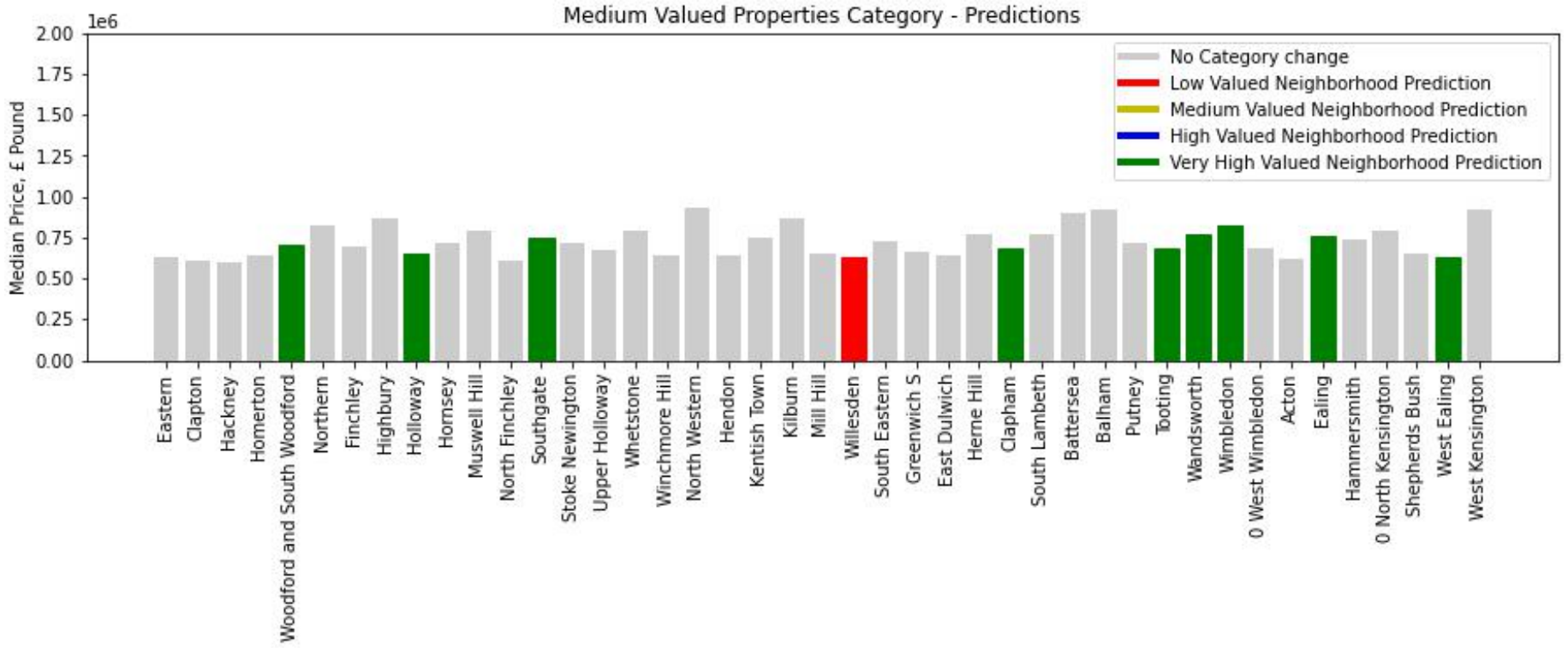
High Valued Neighbourhoods Predictions

- Most high valued postcodes are overpriced according to the model.



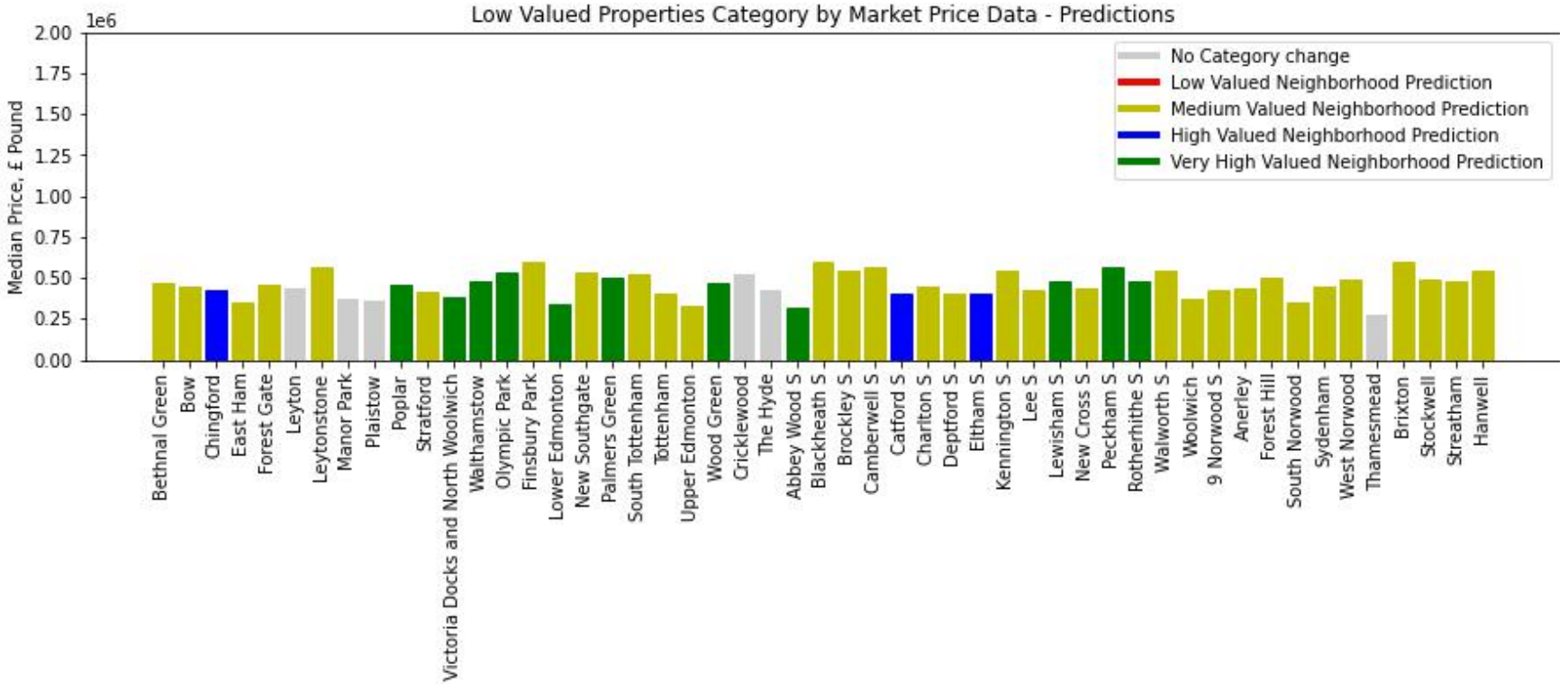
Medium Valued Neighbourhoods Predictions

- Good agreement is found between the model and the market price data. Few exceptions point at higher value.



Low Valued Neighbourhoods Predictions

- Model shows that most neighbourhoods in this category are undervalued and should actually have higher value.



Conclusion and Future Work

- The model seems to be able to accurately predict the value of very high valued properties.

- Most Low valued neighbourhoods present characteristics that are very similar to medium valued properties and some go all the way to very high valued neighbourhoods.
- Most High valued properties are overvalued compared to the amenitites they offer

- Future work will consider including metrics such as TFL and public transport access.
- Distance to high street venues.