PROJECT REPORT (THE BATTLE OF NEIGHBORHOODS)

By Mohammad Asif

June 23, 2021

This report is about the final capstone project of my course “IBM Data Science Professional Certificate”. It includes 6 sections covering 3 pages. It is one of the 3 requirements for completing this course.

Note: For coding and depiction, please view the notebook on below link:

https://github.com/yunaf/Coursera_Capstone/blob/f4c0cec09598ca7dabea8bb5dc848d0e349a53ad/M%20ASIF%20Peer%20Graded%20Assignment--The%20Battle%20of%20Neighborhoods%20(Week%202).ipynb

1. Introduction to Business Problem (Who would be interested in this project)

As per Bloomberg News, the London Housing Market is in an inertia. It is now dealing with some of distinct headwinds, which include the possibility of increased taxes and an intimation from the Bank of England that U.K. domestic values may fall as plenty as 30 percent within the occasion of a disorderly separation from the European Union. More specifically, 4 disregarded cracks recommend that the London marketplace can be in worse form than many realize: hidden charge falls, record-low sales, homebuilder exodus, and tax hikes addressing abroad customers of houses in England and Wales.

Here, it is imperative to adopt machine learning tools in order to help customers in London to make wise and effective decisions. Consequently, the business problem we are working on is: how could we provide support to customers to purchase a suitable real estate in London in this ambiguous economic and financial scenario?

To solve this business problem, we are going to cluster London neighborhoods in order to recommend venues and the current average price of real estate where customers can afford a real estate investment. We will sort out profitable venues according to amenities and essential facilities covering such venues i.e. elementary schools, high schools, hospitals & grocery stores.

2. Data Description & Source.

The Data Source:

The data on London real estate were extracted from the HM Land Registry from the link (http://landregistry.data.gov.uk/). The following fields form the address data included with price paid data: Postcode; PAON (Primary Addressable Object Name). Normally the house number or name; SAON (Secondary Addressable Object Name). If there is a sub-building, for example, the building is divided into flats, there will be a SAON; Street; Locality; Town/City; District; County.

The Data Description:

This data will help us to play around with our coding because our all coding will depend upon it. To explore recommended locations across different venues according to the presence of amenities and essential facilities, we can access data through FourSquare API interface and arrange them as a dataframe for visualization. By combining data on London real estate and the relative price paid data from the HM Land Registry and data on amenities and essential facilities surrounding such properties from FourSquare API interface, we will be able to recommend profitable real estate investments.

3. Data Methodology/EDA/Inferential Statistical Testing

Here we need the concept of data methodology. The Methodology section will describe the main components of our analysis and predication system. The Methodology section comprises five stages which are:

1. Collect Inspection Data
2. Explore and Understand Data
3. Data preparation and preprocessing
4. Modeling
5. Data Visualization

The Foursquare API is used to access the venues in the neighborhoods. Since, it returns less venues in the neighborhoods, we would be analyzing areas for which countable number of venues are obtained. Then they are clustered based on their venues using Data Science Techniques. Here the k-means clustering algorithm is used to achieve the task. The optimal number of clusters can be obtained using silhouette score. Folium visualization library can be used to visualize the clusters.

4. Results Section:

Even though the London Housing Market is facing a temporary bad times, it is still an "ever-green" for business affairs. One can still invest in real estate market because from the data we can conclude that it is a temporary situation and overall the real estate market has a vast space to grow. The quality of London real estate market is that every person can invest according to his finical status.

5. Discussion/Observation/Recommendation:

We may elaborate our results under two main perspectives:

First Perspective:

One can examine them according to neighborhoods. It is interesting to note that, although West London (Notting Hill, Kensington, Chelsea, Marylebone) and North-West London (Hampsted) might be considered highly profitable venues to purchase a real estate according to amenities and essential facilities surrounding such venues i.e. elementary schools, high schools, hospitals & grocery stores, South-West London (Wandsworth, Balham) and North-West London (Isliington) are arising as next future elite venues with a wide range of amenities and facilities. In the same way, one might target under-priced real estates in these areas of London in order to make a business affair.

Second Perspective:

One can analyze our results according to the five clusters we have produced. Even though, all clusters could praise an optimal range of facilities and amenities. We found two main patterns, i.e. the first pattern is Clusters 0, 2 and 4, may target home buyers prone to live in 'green' areas with parks, waterfronts. Instead, the second pattern is Clusters 1 and 3, may target individuals who love pubs, theatres and soccer type activities.

6. Conclusion Section:

Summarizing the project we conclude that according to Bloomberg News, the London Housing Market is in a rut. It is now facing a number of different headwinds, including the prospect of higher taxes and a warning from the Bank of England that U.K. home values could fall as much as 30 percent in the event of a disorderly exit from the European Union. In this scenario, it is urgent to adopt machine learning

tools in order to assist home-buyers clientele in London to make wise and effective decisions.
The solution to this business problem was that we clustered London neighborhoods in order to recommend venues and the current average price of real estate where homebuyers can make a real estate investment. We recommended profitable venues according to amenities and essential facilities surrounding such venues i.e. elementary schools, high schools, hospitals & grocery stores.

End of Report-Thanks