IBM DATASCIENCE CAPSTONE REPORT

ABSTRACT

FINDING SUITABLE APARTMENTS FOR TOURIST OR INDIVIDUALS SEEKING FOR HOUSES

Contents

Introduction

London is the most beautiful city in the world, where individuals, families and couples want to stay. This city in known for its beauty and attractions. However, individuals, couples and families find it hard to search for appropriate place or hotel to stay.

That may not be the only problem this groups are having. Also, they have to keep account of their pockets i.e. the cost of the house or hotel they want stay. Mind you they also have other offers they can consider except from staying in the city of London, which can suit their budget and the space they require for type of apartment.

Reasons that make many people love London includes, good services, such as transportation, community services and most importantly security, the freedom to walk free and safe. Good restaurant of that suits your need and adventurous parks.

Problem

Finding that specific house seems to be a lot of job for individuals and hence it takes time. This is where I come in handy as a data scientist with skills, I have been equipped with by coursera. I will help people who are struggling to find their desired house in way that, I will use techniques to find their desired house, which suits their need, i.e. budget, safety etc.

Interests

People like couples, tourists and families who are looking for new apartments to stay will be interested because this will help solve their problem and furthermore, it will save their time and finding the best apartments to stay. Business owners, apartment owners will also be interested because this will provide them the opportunity to gain more customers in their different businesses.

Data acquisition and cleaning

Data used

1. Foursquare Location Data

   I will be building an application that will interact with the foursquare platform. The data used will correspond to the dataset I obtained from *www.doogal.co.uk.* Which displays the latitude, longitude, postcode etc.

2. I will be using data from the London borough in south London

Source: www.doogal.co.uk

	Postcode	In Use?	Latitude	Longitude	Easting	Northing	Grid Ref	Ward	Parish	Introduced	Terminated
0	BR1 4BY	Yes	51.417289	-0.001741	539050	170591	TQ390705	Downham	Lewisham, unparished area	1980-01-01	NaN
1	BR1 4DN	Yes	51.418996	-0.002156	539016	170780	TQ390707	Downham	Lewisham, unparished area	1980-01-01	NaN
2	BR1 4EY	Yes	51.418477	0.005042	539518	170736	TQ395707	Downham	Lewisham, unparished area	1980-01-01	NaN

Above is an example of data table

Methodology

1. **Machine learning:**

   I will be using the famous K-means clustering to categorise postcodes into different types of clusters based on the venue categories. The data will be segmented into groups that are similar amongst them self, hence the difference will be the occurrence of venue. For the dataset, K-means

Clustering will append another column to the dataset which will depict the cluster number, similar sectors will be grouped together.

2. Data Wrangling

   The data I got contains a lot of data, so to make it short and precise I will use the following columns: postcode, latitude, longitude and ward, to obtain location, hence for those who don't know the postcode has two parts, i.e. District and sector e.g. (BRQ 4BY) → (District Sector). A new column which contain Sectors is created and added the new column. The data for Lee Green ward is filtered to minimize the calls to Foursquare API

3. Exploratory Analysis:

   This city has 18 unique wards; therefore, it is filtered on a single ward as a point of interest namely Lee Green. As explained, postcode is having components, district and sector.

   We display and plot all the sectors on the map and then 5 unique clusters based on the venue categories provide by the foursquare API data are created
   Lastly explore the top 5 venues in each sector.

   The aim of this project is to limit work for individual and find recommendations for the user. We use Foursquare to get many venues, their coordinates and category per section up to 500m radius and then we limit number of venues as 100 for each section

   As demonstrated in the Result and Discussion section, by clustering the area based on the venue categories and filtering the cluster to narrow down the search area can help a lot to simplify the house hunting process.
   Customers can then prefer to focus their house search by analysing various trade-offs. Those who are confused can efficiently compare various areas and choose the one which is most suitable for them.

   I have demonstrated everything in the code section, by clustering the area in responding to venue categories and filtering using Techniques like machine learning and others. Which saves people time from searching houses or

hotels. Customers can now efficiently compare various areas and choose the one which is most suitable for them.

Discussion

Cluster 0. Contains houses with proximity to Platforms, Train Stations, Café, Grocery Store, Gym/Fitness Center, etc.
Cluster 1. It contains houses with proximity to Pubs, Asian Restaurants, Supermarket, Indian Restaurant, etc
Cluster 2. The cluster hilights all houses in the proximity to Café, Pub, Grocery Store, Coffee Shop, Middle Eastern Restaurant, Fish & Chips Shop etc.
Cluster 3. It contains houses with proximity to Fast Food Restaurants, Veterinarian, Hotels, etc.
Cluster 4. It contains houses with proximity to Rental Car Location, Parks, Shopping Plaza, etc

Conclusion

I have demonstrated everything in the code section, by clustering the area in responding to venue categories and filtering using Techniques like machine learning and others. Which saves people time from searching houses or hotels. Customers can now efficiently compare various areas and choose the one which is most suitable for them.