Capstone Project - The Battle of Neighborhoods

Opening a new restaurant in London, UK

By,

Shriraam Muthuswamy

Jun-18-2020

1. Introduction

1.1 Background

During this pandemic time due to COVID-19, restaurants remains one of the most important essential services. Although, dine-in facility has been currently suspended, the takeout is still active. Despite this situation, one of my friends has decided to open a new restaurant as he believes that eventually business will pick up once this pandemic is over. He decides to open a restaurant based on some theme. Nowadays, restaurants with some themes attracts lot of people. It is one of the places where everyone would like to have some get together. He believes that his restaurant will be unique. Particularly, the location of the restaurant is one of the crucial decisions that will determine whether the restaurant will be a success or failure.

1.2 Business Problem

The main objective of this capstone project is to ***find the most suitable location for one of my friends to open a new themed restaurant in the London city***. By following data science methodology and machine learning algorithms such as Clustering, this project aims to recommend a suitable location to him to open up a themed restaurant in London based on K-Means Clustering.

1.3 Target Audience

One of my friends who wants to open a restaurant in London city

2. Data

2.1 Description of Data

To solve this problem, data source plays a huge role. After doing some research work, I found a data that I thought will be suitable to solve this problem. This data has been taken from the "https://data.world/makeovermonday/2018w51" which contains the information of the different borough's in London. This data contains all the boroughs from the London city from which the latitude and longitude data can be fetched using "Geocoder" package. After this, all the venues around this neighborhood can be fetched using "Foursquare API".

By doing the exploratory analysis on this data and transforming the data to the desired format in the preprocessing and data wrangling stages, this business problem can be solved by one of the machine learning algorithms. In our case, we are going to use "K-Means Clustering

3. Methodology

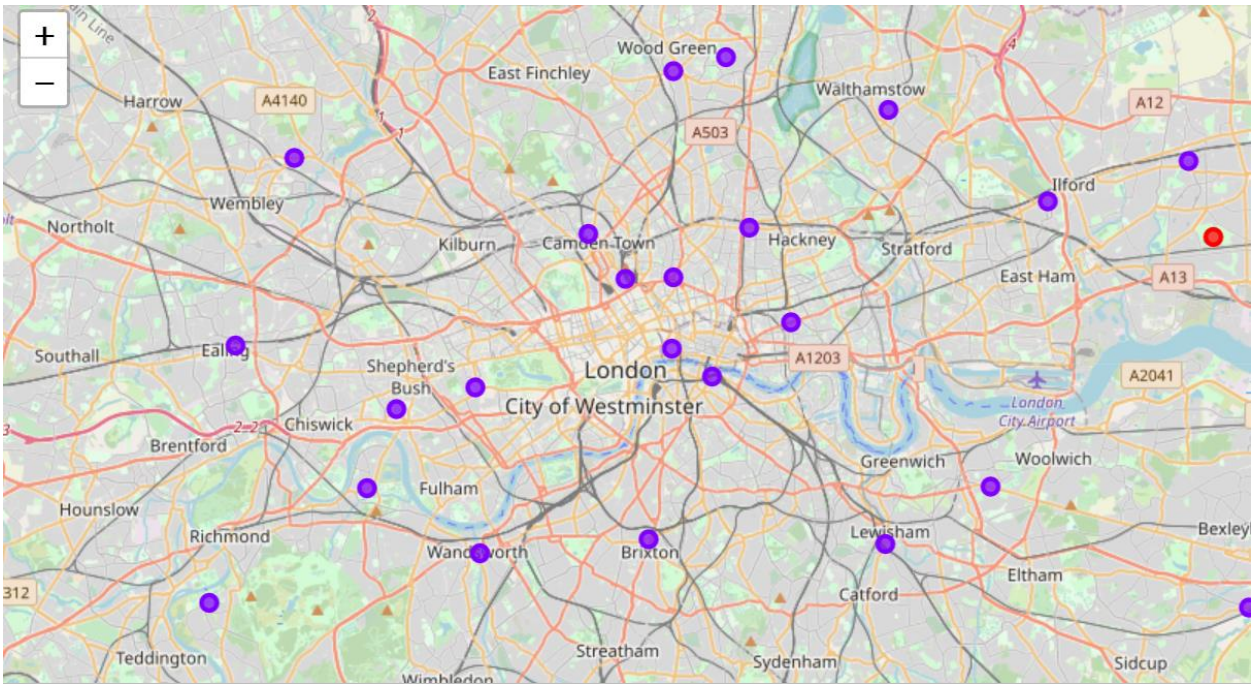
The first and most important step is to have the relevant data set to process. After looking for lot of data sources, the data set for London is found in the data world website (https://data.world/makeovermonday/2018w51). In this, we will have all the boroughs information in London. With this, we will use "Geocoder" package to get the latitude and longitude information and we will merge into the above dataset which we can use to process further. After gathering all this information, we will try to visualize the map of London city using "Folium" package.

Using "Foursquare API", we will pull list of venues in the neighborhood that are in the radius of 500m using the existing credentials of "Foursquare API" to retrieve the data. After this, we will group the venues by neighborhood before processing them to a technique called "one-hot encoding" where the variable is removed, and a new binary variable is added for each unique integer value. This can be done by One hot encoder module in the preprocessing package.

After this, we will try to get the top 5 venues for each neighborhood before processing to the modelling phase. In the modelling phase, we will build a model using the machine learning algorithm to recommend a location to build a new restaurant.

In this project, we have used a machine learning algorithm called “K-Means Clustering” which will form the clusters of similar type. After tuning to find the optimum “k” value, we will train our model and fit them into it. In our project, we have formed 2 clusters. After analyzing both the clusters, I will recommend my friend to open a new restaurant in the neighborhood of “Cluster 1” as there are only few restaurants in the place and opening up a new themed restaurant will have a very good business.

4. Results



The results clearly indicate that there are two clusters. Cluster 1 and Cluster 2.

Cluster 1 with few restaurants, indicated in **red color**

Cluster 2 with lot of restaurants, indicated in **blue color**

5. Recommendations

From the above results, my observation is that there are lot of venues in the neighborhood that are grouped in Cluster 2. So, I would recommend my friend to open a new themed restaurant in the neighborhood of Cluster 1, so that there will not be huge competition and his restaurant will be talk of the town in Cluster 1 if he is good enough in his dishes. He will have lot of opportunities to improve his business and he can taste the success soon. In contrast, if he's opening a restaurant in Cluster 2, since there are lot of venues, it will take a lot of time for him to establish his business.

6. Conclusion

In this project, we ***found the most suitable location for one of my friends to open a new themed restaurant in the London city*** by following data science methodology and machine learning algorithms such as Clustering.

7. References

- https://www.coursera.org/learn/applied-data-science-capstone/ungradedLti/f0QY7/segmenting-and-clustering-neighborhoods-in-new-york-city

- London City Dataset: https://data.world/makeovermonday/2018w51

- Latitudes and Longitudes : https://www.coursera.org/learn/applied-data-science-capstone/

- K-Means best fit : https://www.geeksforgeeks.org/elbow-method-for-optimal-value-of-k-in-kmeans/

- K-Means Clustering : https://www.coursera.org/learn/machine-learning-with-python/ungradedLti/bAtxN/lab-k-means