Exploring London City Venues

Nicola Croon

28/02/2020

1. Introduction

1.1 Background

London City is just a small of part of London and in this area, there are many and many shops, such as coffee shops and restaurants. When I went to London one year ago, I noticed that in this area there were not so many pizza places, as a good Italian would desire. In this area there are thousands of people that eat every day because its centrality in the big town.

This fact gave me such idea. Could it be convenient to open a pizza place in London City?

1.2 Problem

With the help of the Foursquare API I would like to explore the venues of each ward in London City in order to understand if could be convenient or not to open a pizza place. My metric of decision will be the concentration of competition for such places, like pizza places itself and Italian restaurants.

2. Data Acquisition and Cleaning

2.1 Data sources

The data that I used for this project is available at the link:

https://www.doogal.co.uk/AdministrativeAreas.php?district=E09000001.

From this csv file I obtained all the Wards associated to latitude and longitude in London, as well as postal codes and some other data that, honestly, are not relevant for my analysis.

Thanks to this file I got the Wards and their geographical coordinates, but there are many different latitudes for each ward because of the different postal codes in each area. For simplicity I will use just one of them.

The dataset consists in 6799 observation, but the number is consistently reduced by selecting only the existing postal codes, that still in use.

In this notebook I selected only the Wards and their geographical coordinates, and I dropped all the other columns. Later on I will use the Foursquare API to explore each ward and look at the Venues in each Ward.

2.2 Data Cleaning

The data was really cleaned. I dropped all the columns that were not relevant in my analysis. As I said, I selected only the Wards and their coordinates. With the use of the Foursquare API I will get the venues of each one of them.

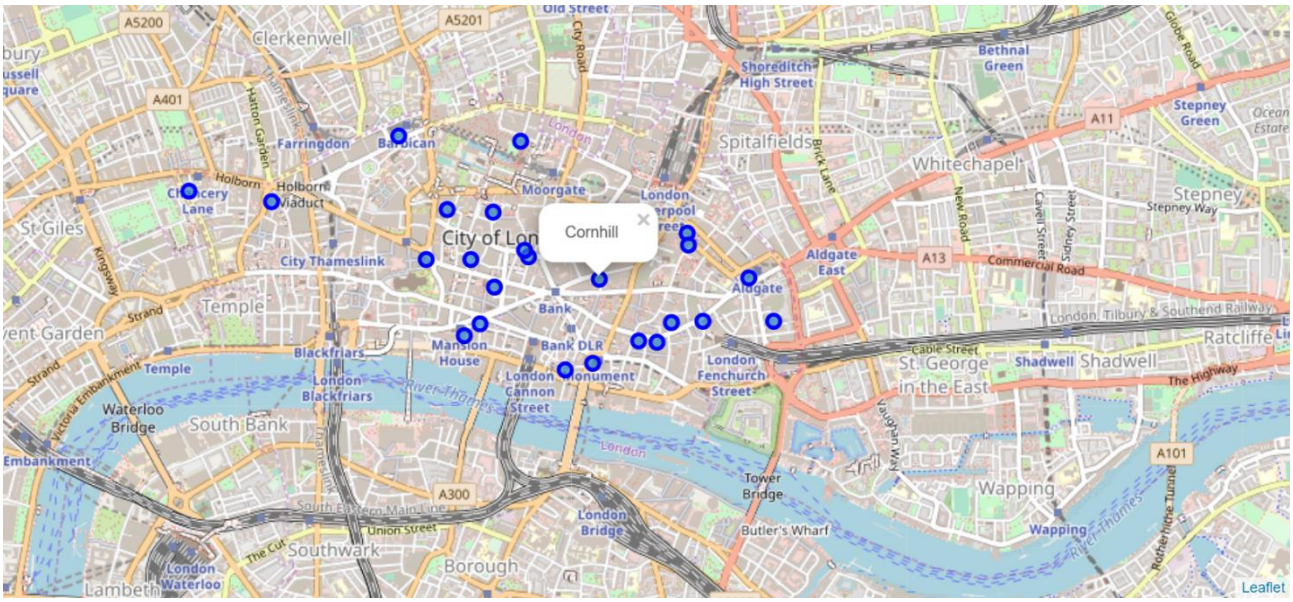
After data cleaning, the dimension of the database was substantially reduced to 25 rows and 3 columns, that were latitude, longitude and the Ward name.

This dataset will be used only for the purpose of discovering the venues in each ward and could or could not be considered as the principal dataset of the analysis.

The real dataset is the one that we get after the implementation of the Foursquare API with all the venues divided by Wards.

3. Exploratory Data Analysis

There are not so many other exploratory data analysis techniques to implement on these types of data. What I did is to plot all the wards with the help of the folium library to give a geographical representation of where.



Ones I get all the venues and did some more EDA by looking at the number of pizza places that are present on the area. What I discovered is that the number of pizza places is not so high, instead the number of Italian restaurants is much higher.

My focus then will not be only on pizza places but also on Italian restaurants, that are considered as competitors.

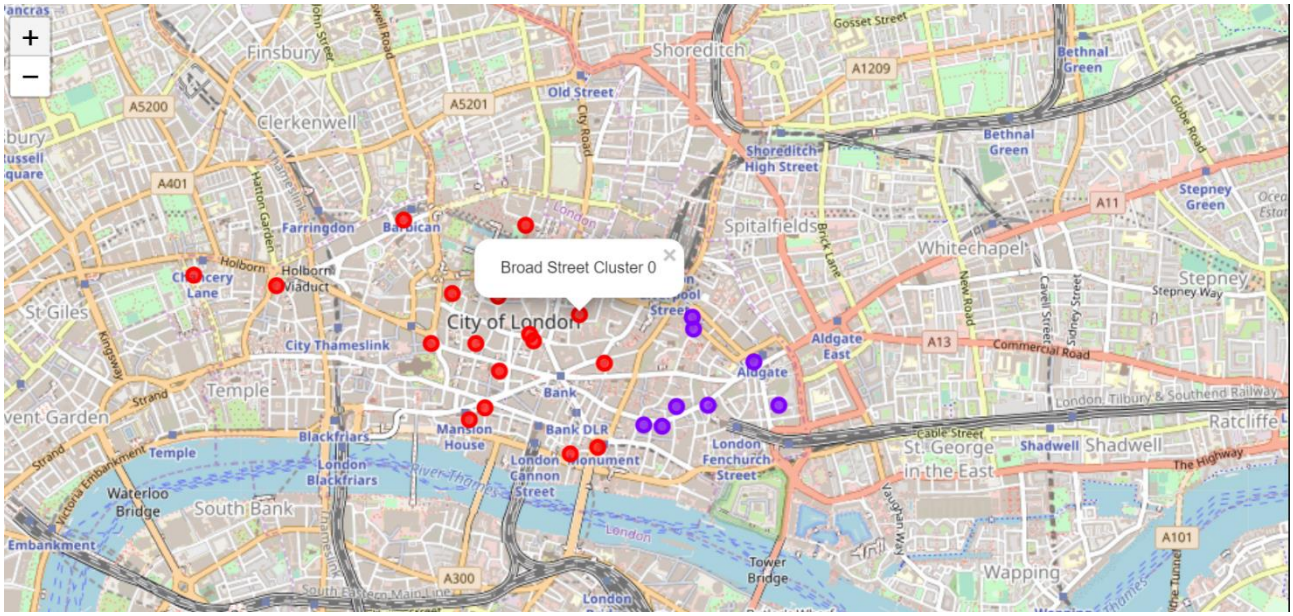
4. Modeling and Methodology

After the use of the API to get the venues of each ward I decided to put together all the results and cluster them in order to explore each cluster and understand which one is the most suitable for the opening of a new pizza place.

I used a K-means Clustering model to cluster the data. The initial number of clusters is 3 because of the dimension of the area, that is not so big.

5. Results

By exploring each cluster, I noticed that there is only one cluster in which the Italian restaurants occurs a lot.



The cluster that I considered eligible for my purposes of constructing a pizza place is the number zero, the one in red, because its high concentration of Italian restaurants in it. Although, not all the wards in it could be considered eligible. Only a few of them have a high number of Italian restaurants and pizza places inside.

After careful considerations I selected the suitable wards for my idea of opening a pizza places and we have:

- Vintry
- Bread Street
- Queenhithe
- Walbrook
- Cordwainer
- Coleman Street
- Aldersgate
- Bassishaw
- Cheap

6. Discussion

In these wards there is no direct competition from other pizza places because the most of them are chain stores, and this is not what means to me to open a pizza place. Although, there could be some competition by some other Italian restaurants but the product that will be offered is different, it could work!

7. Conclusions

There are further analysis that could be performed but, based on our initial business scope the analysis gave me good and meaningful results. In this analysis I considered only a small area of London and surely it could be done for the whole area, with a more complex analysis and with a higher computational cost.