Week 5

IBM Data Science Professional Course Applied Capstone

Opening a New Halal Restaurant in London, England

By: Bilal Siddiqui

August 2019

Introduction

One of the most difficult things for certain people whether it is for religious reasons, specialty needs, or just because they want to eat healthy is locating a halal place to eat. They can always locate a vegetarian restaurant or go into a restaurant that does not provide halal options, but they do not get the same satisfaction as they could have had. Businesses are losing out on hundreds of potential customers because of a lack of halal options around the area. The city of London is one of the largest populated cities in the world. Opening a restaurant that caters to all dining needs including halal options will allow a business to increase revenue significantly. What else? Deciding where to open the restaurant is one of the most important decisions any business professional can make.

Business Problem

We have determined problem as the difficulty of finding halal options within an area to satisfy all customers needs. Our objective is to determine the best location in the city of London to open a new halal restaurant. Using the data science methodology steps we have learned throughout this course such as data mining, machine learning techniques, and more, we aim to help an individual answer this problem: Where is a good place to open a halal restaurant in the city of London?

Target Audience

We are looking to open a new halal restaurant in the city of London. This project would highly benefit business professionals specifically in the food business looking to take control of the halal consumer population and increase their revenue significantly. The population of Muslims who are the majority of halal consumers is continuously growing in London at about 12.5%

percent of England's total 21%. This analysis will help maximize the stakeholders benefits and deciding where to open the new restaurant.

Data

In order to open our new halal restaurant in London, England, we will need:

1) Gather the neighborhoods in London, the capital of England to define the scope of our problem.
2) Latitude and longitude coordinates of the neighborhoods we gathered for the city of London
3) Venue data on restaurants to perform machine learning algorithms on the data specifically clustering.

Sources

In order to get the neighborhoods for our analysis we will be using the following link: (https://www.airbnb.com/locations/london/neighborhoods) provided by Airbnb. Airbnb is an online hospitality service that members use to arrange lodging, temporary stay, our tourism experiences. The link provides 48 neighborhoods in London that we will use through web scraping from the 'BeautifulSoup' library to extract the data.

We will then using the Python geocode package to obtain the latitude and longitudes of the neighborhoods we have extracted.

Finally, we will use the Foursquare API to get the venue data on the London neighborhoods particularly restaurants.

Methodology

In the first step of our process of opening a new halal restaurant in London, England, we will need to get the list of neighborhoods provided to us in Airbnb page (https://www.airbnb.com/locations/london/neighborhoods). We will utilize web scraping using Python requests and the BeautifulSoup packages to extract the necessary information on neighborhood data. Following this, we will need the geographical coordinates which we will

obtain from the geocoder package that will allow us to convert the address into longitude and latitude coordinates which allows us to use the Foursquare API. Once we have gathered the data, we will utilize the pandas library to create a dataframe using the data and then visualize our findings using the Folium package.

Using the Foursquare API, we will get the top 100 venues that are within a 3000 meter radius. We will make API calls to Foursquare passing in the coordinates of the neighborhoods in a Python loop which will then return venue data in JSON format which we can then extract the venue name, category, and latitude and longitude. With the data we have received, we can determine how many venues were returned for each neighborhood and then analyze each by grouping the rows by neighborhood and taking the mean of the frequency of occurrence of each venue category. We will be gaining data specifically on “Indian Restaurants” because they hold the most similar cuisine that of most halal places.

Finally, we will perform clustering on the data through k-means clustering, an unsupervised machine learn algorithm to partition n number of observations and allocate each data point to the nearest cluster. We will cluster the neighborhoods into 3 clusters based on the frequency of “Indian Restaurant” because as mentioned before that is what we will be analyzing. The results will allow us to identify which neighborhoods has fewer number of restaurants and what neighborhoods has a larger number of restaurants which we can then determine to help us answer our question as to which neighborhoods are the best fit for opening a new halal restaurant.

Results

Based on the results we have gathered, we have come to three results based on our 3 clusters.

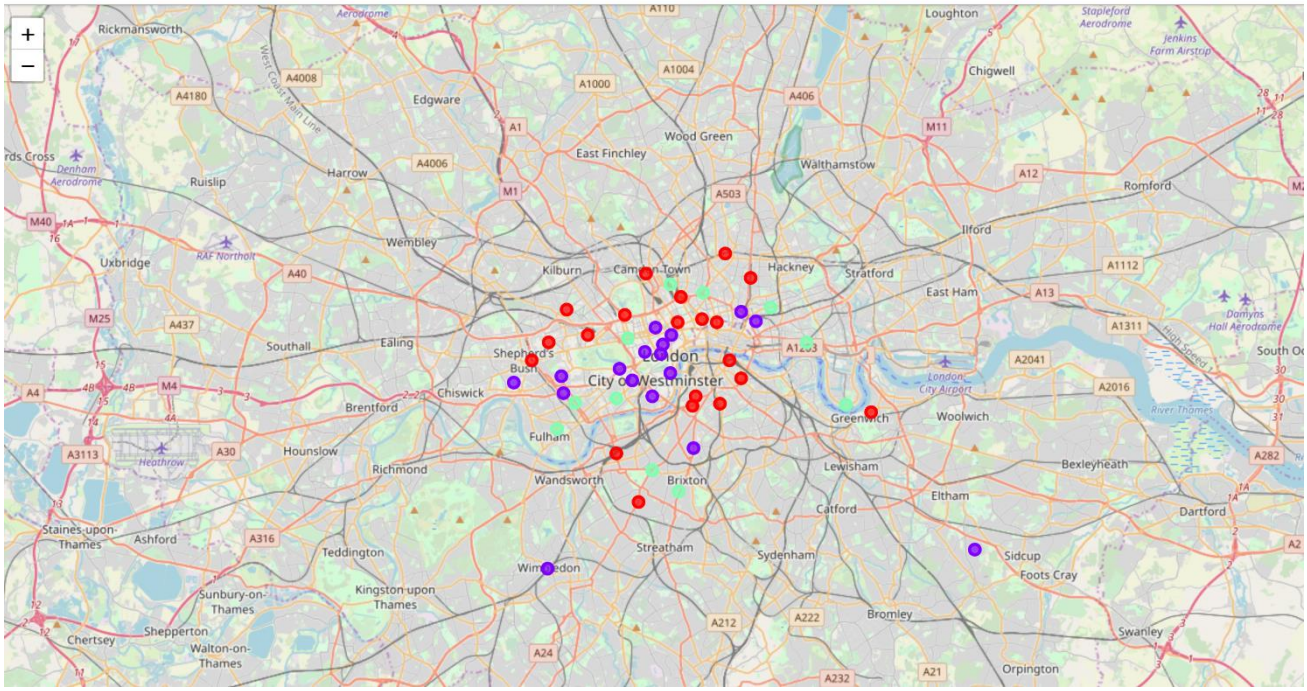
For the first cluster, categorized as ‘0’, we have very few Indian restaurants.

For the second cluster, categorized as ‘1’, we have the greatest number of Indian restaurants.

For the third cluster, categorized as ‘2’, we have a moderate number of Indian restaurants.

The clustering show on a map is below:

- Cluster 0 is shown by green dots
- Cluster 1 is shown by purple dots
- Cluster 2 is shown by red dots



As we can see, the greater number of Indian restaurants are focused within the inner city of London, England.

Discussion/Suggestions

One of the largest observations noticed is the significant amount of Indian restaurants in the inner city of London and the lack thereof in the outer part. Cluster 0 represents the outer for the most part and because this project was done without the additional research on factors such as population, it is recommended to open a new halal restaurant closer to the inner city,

but not around cluster 1 as there would be an intensive amount of competition due to surrounding restaurants.

In this project, we only considered one factor out of the many possible other ones such as rent cost, population, income, and more that could influence the opening of a new halal restaurant in London. In the future, it is advised to take a look at as many possible influences that might have an affect on opening a new restaurant. There is also a limited number of API calls that can be made with Foursquare free version, and so it is recommend to bypass this issue to gain the maximum amount of information possible. There is also the possibility of other restaurants such as Chinese, Thai, Pakistani, or others being halal or having halal options.

Conclusion

The main goal of this project was to answer the question: Where is a good place to open a halal restaurant in the city of London? Using Airbnb to gather the neighborhoods in London and using the Foursquare API to retrieve data about each of them, we have come to our conclusion. We have determined the different concentration of Indian restaurants for each cluster, and so what is the best place?

The final decision is up to the stakeholder who will be opening the restaurant. There are a number of different circumstances to consider. Maybe cluster 0 has a higher rent cost or cluster 2 has a larger population. These are just a couple. The information we have gathered here provides quality information to get started on where to open our new halal restaurant. Indian restaurants in cluster 1 are facing intensive competition due to the higher concentration, I would def recommend staying away from there especially as a new chain opening. Cluster 0 has very few Indian restaurants so that could mean businesses don't typically do well or rent cost is steep. This project recommends going with cluster 2, as there is moderate competition, with significant population to increase customers as well as the fact being a similar restaurant with very similar options, but halal gives them an edge. Nonetheless, stakeholders are not advised against going with cluster 0 as well.