IBM Data Science Professional Certificate

Opening a Mexican restaurant in London, UK

Saulo Adonai Villaseñor Avalos

August, 2021

Introduction

London is one of the most multicultural cities in the world as well as one of the most touristic ones. The city has beautiful sights, a lot of history, beautiful cosmopolitan and classical areas with great architectural designs, plenty of museums, fantastic shopping malls with great brands and many other things. Thousands of people from every corner of the world come to visit the British capital every day and besides the mentioned above, people also want to enjoy a delicious meal which gives them the energy they need for their daily activities, or have a proper dinner after a busy and tiring day. Because of the multicultural background of London and the excitement of people to make the most out of their visit to London, there are a lot of places in the city for people to have a unique gastronomical experience. However, given the huge distance that separates London from the Latin-American countries, there are not many Mexican restaurants in the city, even though the Mexican cuisine is considered one of the best in the world. This give people a great opportunity to start a new and exciting business in the city.

Business problem

The Mexican cuisine is considered one of the best cuisines in the world and there is something in the Latin-American culture that is making people in Europe fond of it. This makes the opening of a Mexican restaurant an excellent option to invest in a business in London, the capital of England and the United Kingdom. However, it is essential to make an analysis of all the required data to find where is the best spot to open the restaurant since there are already some competition. Besides, the metropolitan area of London is composed of many neighbourhoods and people. So, how does this neighbourhoods relate between them? And what would be the right place to locate the next Mexican restaurant in London?

Target public for the project

This project is particularly useful to property developers and investors looking to open or invest in a new restaurant of Mexican food in the capital of the United Kingdom. According to the data obtained from the Foursquare API, there are currently 11 Mexican restaurants, so there is still plenty of opportunity for a business of this kind, but one should not give for granted the competition. Thus, the present work will help the investors to select the best place in the city for the opening of a Mexican restaurant.

Data

The following data will be used to solve this problem:

1. The list of neighbourhoods in London. This will define the scope of this project, which is confined to the city of London, capital city of England and the United Kingdom, in Europe. The information will be obtained via web scraping from the *List of areas of London* page of Wikipedia.
2. The latitude and longitude of the neighbourhoods. In obtaining the coordinates data of the neighbourhoods, the Geocoder package will be used with the arcgis_geocoder to obtain the latitude and longitude of the needed locations. This is required to plot the map and to get venues data.
3. Venue data related to each of the neighbourhood in London. This will be obtained from the Foursquare API by using the respective credentials and calls. This data will be used to perform the clustering of the neighbourhoods.

Sources of data and methods to extract the data

The *List of areas of London* Wikipedia page will be used to get the information of 531 neighbourhoods from Greater London, which is a metropolitan county in England that includes many cities around the boundaries of London and London itself at the centre. The information from the page is displayed in a table with information such as the name of the neighbourhoods, the London borough they belong to, their post town and their postal code, dial code and OS grid ref, which is a value that identifies the location any postal code in the UK. Web scraping will be used as a method to obtain the data from the Wikipedia page to our program, with the help of Python's BeautifulSoup and requests packages. Afterwards, the geographical coordinates of the neighbourhoods will be obtained by using the Geocoder package from python, which gives us the latitude and longitude of each neighbourhood.

After this, the Foursquare API will be used to get the venue data for those neighbourhoods. Foursquare has one of the largest databases of more than 105 million places and is used by over 125000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Mexican Restaurant category to help us to solve the presented business problem. This project will also use many data science techniques, such as web scraping,

working with APIs, doing data cleaning, data wrangling, applying K-means clustering (which a machine learning technique), and mapping venues by using Folium. In the next section, the methodology section will be presented. Here we will discuss the steps taken for this project, the data analysis made and the machine learning technique used to solve the problem.

Methodology

Firstly, we need to get the list of neighbourhoods in London. Fortunately, the list is available in the Wikipedia page *https://en.wikipedia.org/wiki/List_of_areas_of_London*. We will do web scraping using Python's requests and BeautifulSoup packages to extract the list of neighbourhoods data. However, it is important to note that this is just a list of names. We also need to get the geographical coordinates in the form of latitude and longitude in order to be able to use the Foursquare API. To do so, the Geocoder package will be used to convert addresses into geographical coordinates in the form of latitude and longitude. After gathering the data, the data will be populated into a pandas DataFrame and then we will visualise the neighbourhoods in a map using the Folium package. This will allow us to perform a sanity check to make sure that the data of the geographical coordinates returned by the Geocoder package is correctly plotted in the city of London.

Next, the Foursquare API will be used to get the top 100 venues that are within a radius of 500 meters. We need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare passing in the geographical coordinates of the neighbourhoods in a Python loop. Foursquare will return the venue data in JSON format and the venue name, venue category, venue latitude and venue longitude will be extracted. With the data, we will check how many venues were returned for each neighbourhood and examine how many unique categories can be curated from all the returned venues. Then, we will analyse each neighbourhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering. Since we are analysing the "Mexican Restaurant" data, we will filter by setting "Mexican Restaurant" as venue category for the neighbourhoods.

Lastly, we will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies K number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and most popular

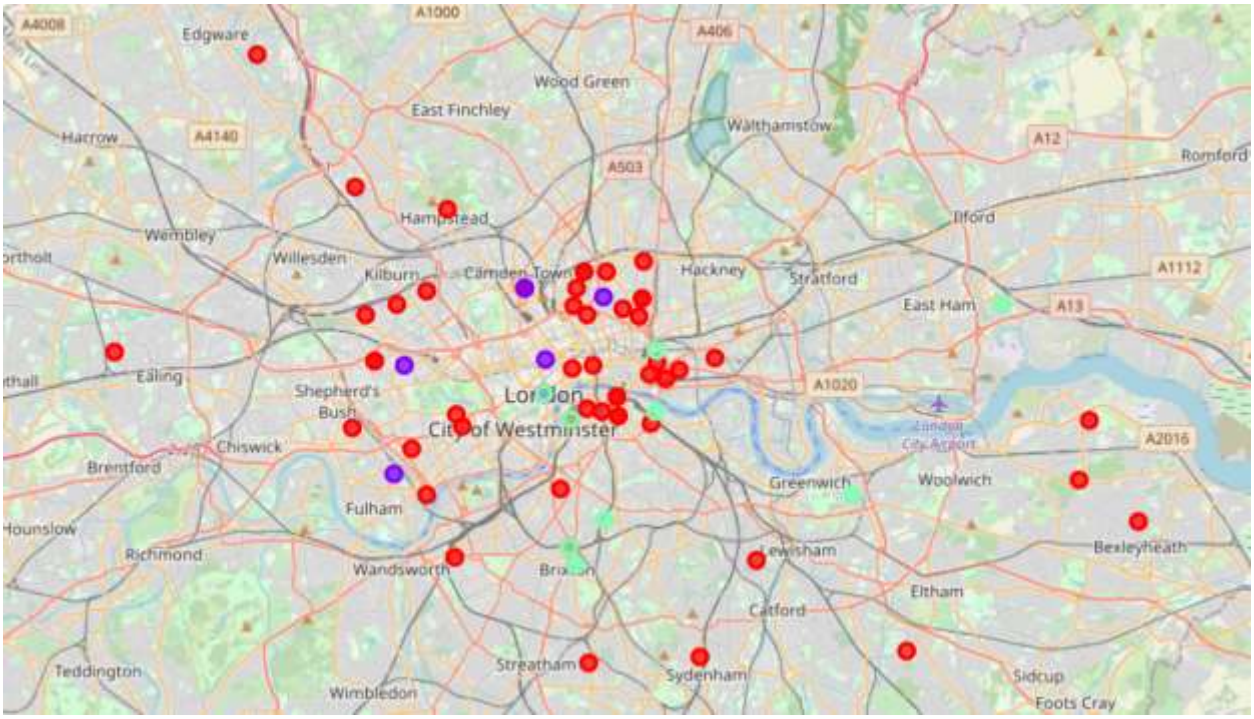
unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. The neighbourhoods will be clustered into 3 clusters based on their frequency of occurrence for “Mexican Restaurant”. These results will allow us to identify which neighbourhoods have higher concentration of Mexican restaurants and also which neighbourhoods have fewer number of restaurants. Based on the occurrence of Mexican restaurants in different neighbourhoods, it will help us to answer the question as to which neighbourhoods are most suitable to open the next Mexican restaurant in London.

Results

The results from the k-means clustering show that we can categorize the neighbourhoods into 3 clusters based on the frequency of occurrence for “Mexican Restaurant”:

- Cluster 0: Neighbourhoods with no number of Mexican restaurants.
- Cluster 1: Neighbourhoods with low number to no existence of Mexican restaurants.
- Cluster 2: Neighbourhoods with high concentration of Mexican restaurants.

The results of the clustering are visualized in the map below with cluster 0 in red colour, cluster 1 in purple colour, and cluster 2 in mint green colour.



Discussion

Based on the observations from the map in the Results section, most of the Mexican restaurants are concentrated in the central area of London, corresponding to cluster 1 and cluster 2. This was already expected, since these areas correspond to the most touristic areas of London and a lot of people circulate around here. In cluster 0 there are no Mexican restaurants at all, so it would be a good idea to open a new one in a neighbourhood of this cluster. The most attractive spots for the restaurant are the areas which are closer to the centre of London and important zones like the City of London, which is the economical heart of London, or in the Southwark, another important area which is connected by bridges across the River Thames to the City of London and the London Borough of Tower Hamlets. So, as consultants, our recommendation for the best spot for our next Mexican restaurant would be Greenwich, which is in the heart of Southwark and is connected to several important highways; Blackfriars, which is in the City of London and is close to historical landmarks like Saint Paul's cathedral as well as some important highways; or Woodford, which is close to the London Bridge and next to a busy train station in London. These are our selected spots to open a Mexican restaurant in London, and they also are located to a fair distance from the other Mexican restaurants, making it even better since there will be no much competition and there are not many Mexican restaurants in London. There is no doubt that any of the selected spots will give an advantage to the owner of the restaurant over anyone who plans to open it in a different place.

Limitations and suggestions for future research

In this project we only considered one factor: the frequency of occurrence of Mexican restaurants. There are other factors such as population and income of residents that could influence the location decision of a new restaurant. However, to the best knowledge of this researcher such data was not available to the neighbourhood level required by this project. Future research could devise a methodology to estimate such data to be used in the clustering algorithm to determine the preferred locations to open a new restaurant. In addition, this project made use of the free Sandbox Tier Account of Foursquare API that came with limitations as to the number of API calls and results returned. Future research could make use of paid account to bypass these limitations and obtain more results. However, important assumptions were made when wrangling the data and the discussion takes into account the touristic activity of the neighbourhoods. This is explained and documented in the notebook code for the project.

Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 3 clusters based on their similarities and providing recommendations to the relevant stakeholders, i.e. property developers and investors regarding the best locations to open a new Mexican restaurant in London. To answer the business question that was raised in the introduction section, the answer proposed by this project is: The neighbourhoods in cluster 0 are the most preferred locations to open a new Mexican restaurant, and more specifically Greenwich, Blackfriars and Woodford due to their touristic location and closeness to the centre on London, besides the important roads and landmarks that are close to these neighbourhoods. The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding losses and making the most out of their investment when they take the decision of opening the next Mexican restaurant in London.

References

- En.wikipedia.org. (2021). *List of areas of London - Wikipedia*. [online] Available at: <https://en.wikipedia.org/wiki/List_of_areas_of_London>.
- Geographers' A-Z Map Company (2008), *London Postcode and Administrative Boundaries* (6 ed.), Geographers' A-Z Map Company, ISBN 978-1-84348-592-6
- Mills, A.D. (2001), *Dictionary of London Place Names*, Oxford, ISBN 0-19-280106-6-
- Royal Mail (2004), *Address Management Guide* (4 ed.), Royal Mail Group
- Foursquare. 2021. *Foursquare Developers Documentation*. [online] Available at: <https://developer.foursquare.com/docs/>.