IBM Applied Data Science Capstone - Final Project

By DANIELA SCOCCHIA

As a part of the final IBM Capstone Project, we have the opportunity to come up with an idea to leverage the Foursquare location data to explore or compare neighborhoods or cities of our choice or to come up with a problem that can use the Foursquare location data to solve.

In this report I will go through problem designing, data preparation and final analysis.

1. Discussion and Background of the Business Problem:

Problem Statement: Prospects of a Luxury Restaurant, in the wealthiest boroughs of London.

London is the capital and biggest city of England and the United Kingdom. It is one of the world's most important global cities and has been called the world's most powerful, most desirable, most influential, most visited, most expensive, sustainable, most investment-friendly, and most-popular-for-work city. London is definitely one of the best places to start up a new business. However, London is highly competitive and therefore, the need of a specific customer target for a restaurant it is really important.
For this project I have hypothesized a customer who intends to open a luxury restaurant in a wealthy borough of London. As a Data Scientist, I will firstly assess which are the 10 most populated borough by calculating the ratio between the estimated population and the borough's area. I will then assess the top 5 wealthiest and populated boroughs of London by looking at average properties value. Then, I will identify which are the ones have more restaurants and what are the top trending venues. This will indicate which borough is the most suitable to open a luxury restaurant where people tend to go out for eating and which ones are the ones with more competition.

Target Audience

1. Businesspeople who wants to invest or open a luxury restaurant in London.
2. Freelancer who loves to have their own restaurant as a side business. This analysis will give an idea of how beneficial it is to open a restaurant and what are the pros and cons of this business.
3. Budding Data Scientists, who want to implement some of the most used Exploratory Data Analysis techniques to obtain necessary data and analyze it.

2. Data description

Boroughs of London Data

Firstly, I need to look at what data are available online describing the boroughs of London. The Wikipedia page ("List of London boroughs", publicly available at https://en.wikipedia.org/wiki/List_of_London_boroughs) contains a table describing the 32 boroughs of London with their relative data about Local authority, Political control, Headquarters, Area (sq mt), Population (est 2013), geographical co-ordinates, and the number of the legend on the Wikipedia map.

I will use this table to determine the most populated boroughs by calculating the ratio between the Area (sq mt) and Population (est 2019).

Properties value data

In order to understand which is the average value of properties in the boroughs, I will use the table "London property value" publicly accessible from the government website (https://data.london.gov.uk/download/average-house-prices/b1b0079e-698c-4c0b-b8c7-aa6189590ca4/land-registry-house-prices-borough.csv).

The table contains information about area codes, name of the borough, the year, the measure (Mean, median, sales) and the Value. I will be focusing on mean Values from the year ending in December 2017.

Geographical and Venues Data

To get the geographical information of the borough I will use geolocator.geocode. I will then use foursquare to look at which boroughs have the highest number of restaurants and what are the top 5 trending venues. This will indicate the best borough where to open a luxury restaurant based on population, wealth, and trending venues.

3. Methodology

The aim of the project is to find an optimal location for opening a luxury restaurant in one of the wealthiest and most populated boroughs of London. Firstly, the data used for the analysis was collected from publicly available sources. The obtained information comprises of the names, geolocation, population, and average value for each of the borough of London. After borough data was scraped from Wikipedia, I put them into a data frame and then further transformed, so that unnecessary columns and rows are dropped, and column headers have a name that is easy to work with. Final data frame contained information about each of the borough, along with the respective population and area.

As the original table from Wikipedia contained the coordinates in the wrong they were converted by using geolocator.geocode.

The ratio between Population/Area was calculated to identify the highest densely populated boroughs. To find the housing price for each borough, official data from the government were used. The table provided data for each borough with the year and the value. However, different methods were used to calculate the value. I extracted the data concerning Mean values relative to the most recent year (year ending in 2017). The Geopy geocoders library was then used to gather the coordinates of each borough and this information was also put into a data frame.

Foursquare API was utilized to find restaurants in a radius of 1000m from the coordinates of each borough. Then I performed a clustering of the districts with an unsupervised learning K-means algorithm. Finally, based on the obtained clusters, outcomes were discussed and conclusions were drawn.

4. Results

Identification of the wealthiest boroughs of London

In order to identify the wealthiest borough of London I inspected the mean property values by analyzing the publicly accessible data (https://data.london.gov.uk/download/average-house-prices/b1b0079e-698c-4c0b-b8c7- aa6189590ca4/land-registry-house-prices-borough.csv).

Here, it resulted that the top 10 wealthiest boroughs are Kensington and Chelsea, Westminster, Camden, Hammersmith and Fulham, Wandsworth, Richmond Upon Thames, Islington, Southwark, Barnet, and Haringey.

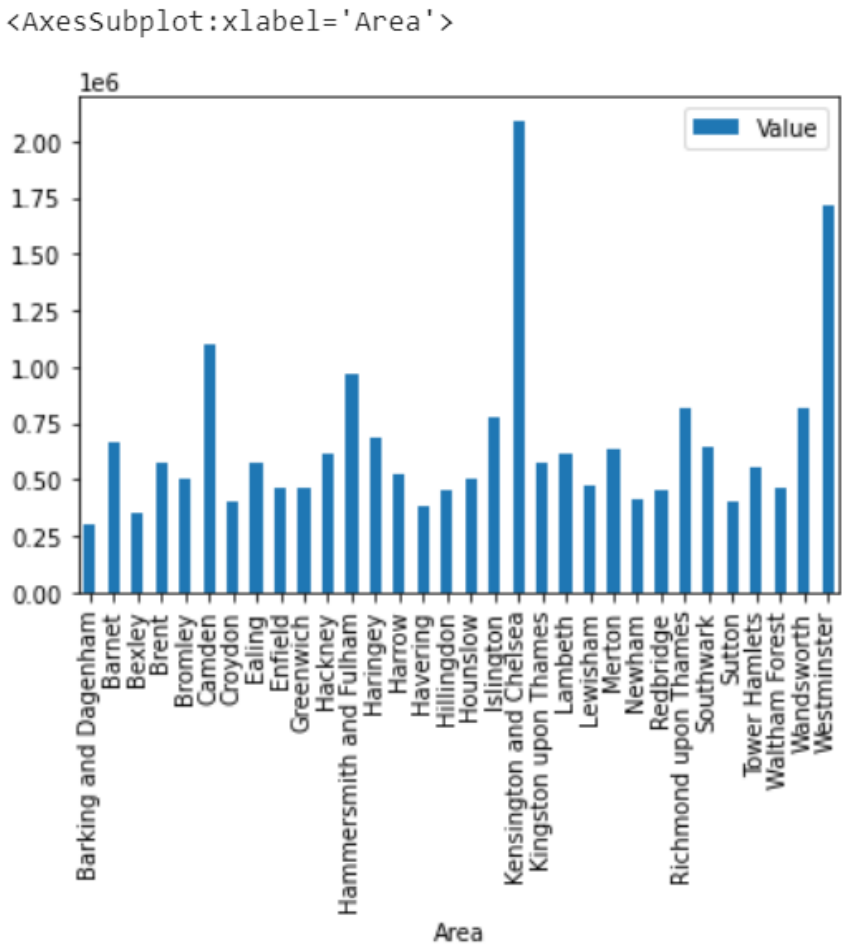

Figure 1: Mean property value in the London Boroughs.

Identification of the most densely populated and wealthiest boroughs of London

In order to identify the most densely populated and wealthiest borough of London I calculated the ratio population/area (mt sq) (Figure 2).

Looking at the figure below, it's easy to understand that the top 5 wealthiest boroughs resulted to be Kensington and Chelsea, Westminster, Camden, Hammersmith and Fulham, and Islington.

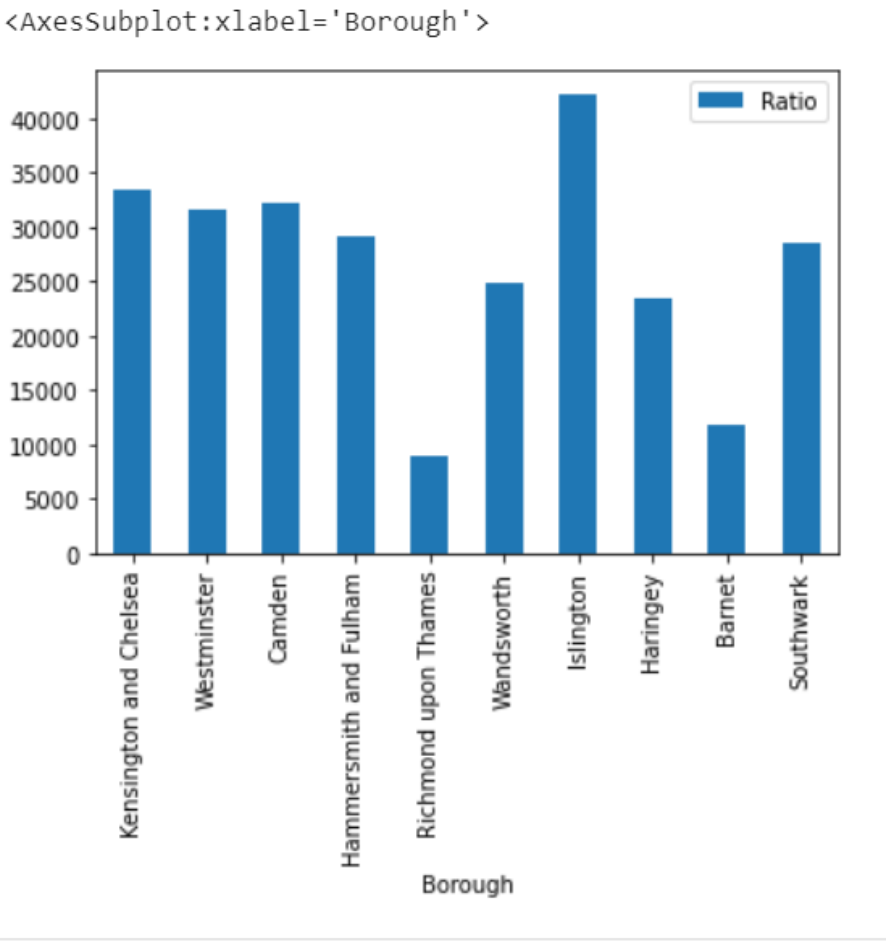

Figure 2: The most densely populated and wealthiest boroughs of London.

Identification of the borough with most restaurant

Inspection of the venues in each borough revealed that the borough with the highest number of restaurants is Camden with 9 restaurants, followed by Kensington & Chelsea with 8, Islington with 7, Hammersmith & Fulham, 4, and Westminster, with just 2 (Figures 3 and 4).

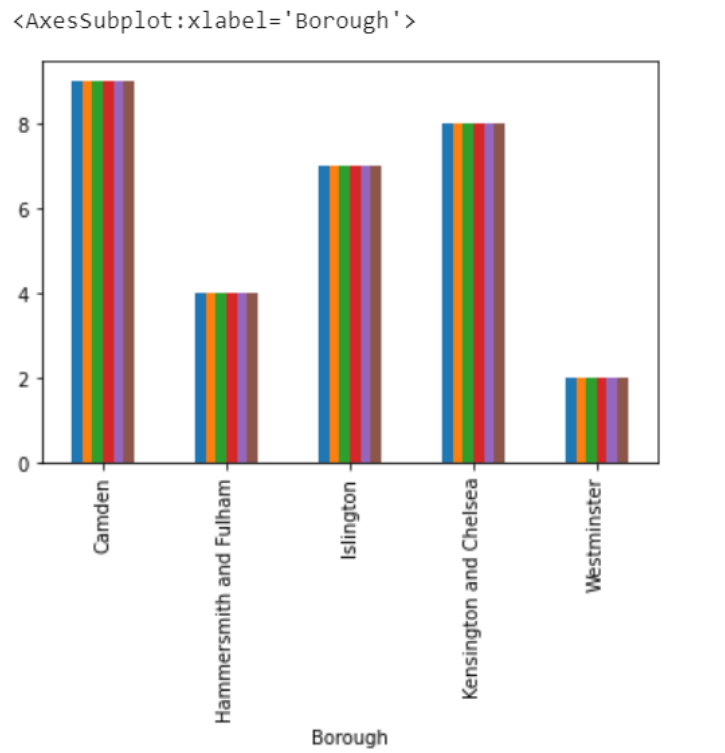

Figure 3: Number of restaurants in the wealthiest boroughs of London; box plot.

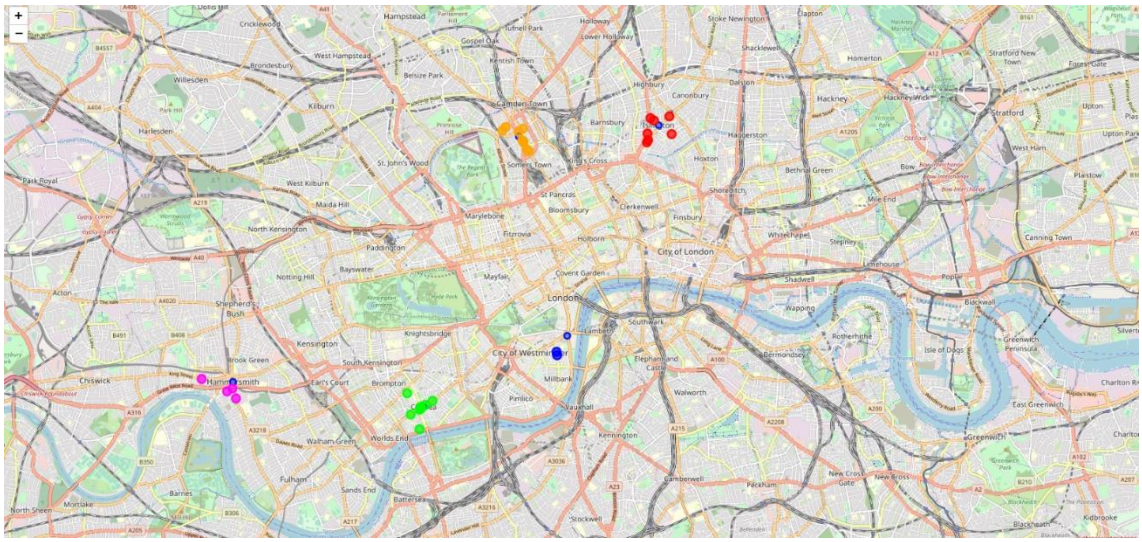

Figure 4: clustering of restaurants in the wealthiest boroughs of London; map view.

Clustering of borough according to their venues

Clustering of the boroughs showed that they are unique in terms of top 10 venues as every cluster resulted to correspond to a borough (Figure 5)

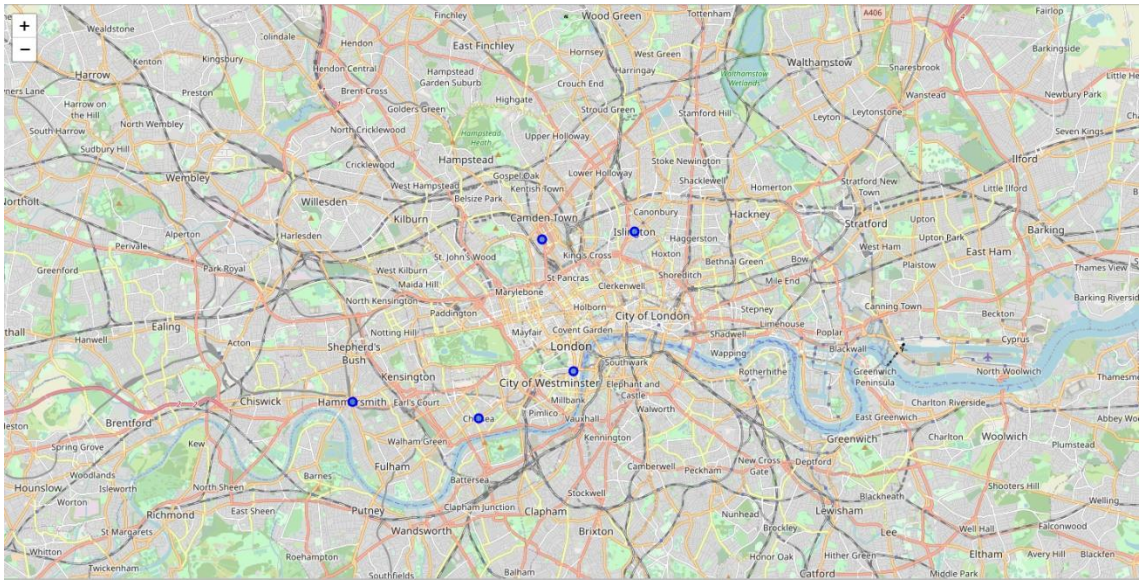

Figure 5: clustering of boroughs by top 10 venues; map view.

Identification of the trending venues in the Boroughs.

The inspection of most trending venues showed that in Kensington & Chelsea, food and hospitality venues are the top trending venues differently to Camden where also music venues are included (Figure 6).

	Borough	Value	Ratio	latitude	longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Islington	778290.0	42241.637631	51.538429	-0.099905	Pub	Bakery	Theater	Burger Joint	Yoga Studio
1	Kensington and Chelsea	2092485.0	33360.897436	51.487542	-0.168220	Bakery	Japanese Restaurant	French Restaurant	Ice Cream Shop	Gift Shop
2	Camden	1099876.0	32146.309524	51.536388	-0.140556	Coffee Shop	Tea Room	Music Venue	Yoga Studio	Falafel Restaurant
3	Westminster	1718124.0	31521.954162	51.500444	-0.126540	Hotel	Plaza	Garden	Historic Site	Coffee Shop
4	Hammersmith and Fulham	972231.0	29248.499210	51.492038	-0.223640	Pub	Coffee Shop	Hotel	Concert Hall	Plaza

Figure 6: Top 5 trending venues in the wealthiest boroughs of London.

5. Discussion

Property values analysis revealed the wealthiest boroughs where to open a luxury restaurant. By normalizing the data with the population density data (see Ratio values), it resulted that the areas of Islington, Kensington & Chelsea, Hammersmith & Fulham, Camden, and Westminster are the most densely populated and wealthiest of London.

Here, it resulted that in Camden there's highest number of restaurants. This could suggest that opening a luxury restaurant here could be a risk, as the client could incur in more competitiveness and less revenue.

Trending venues analyses showed, instead, that in Kensington & Chelsea area 4 out of the top 5 trending venues are all belonging to the category of food and hospitality. This indicates that, despite more competition, the hospitality sector in Kensington & Chelsea is flourishing. Wealth assessed on Mean property values also suggests high profile costumers. This suggests that this area could be a suitable spot for opening a luxury business.

Finally, we can say that also Westminster could be a very good candidate. As a matter of fact, it is the most central area of the 5, and with a high number of hotels (1st most common venue), suggesting a high number of visitors (residents, workers, tourists, etc..).

The lack of restaurants as top 10 venues in this area means that the customer could have a unique spot for their business to grow, with little competitivity.