Predicting the most appropriate London neighborhoods to open a Mediterranean fusion restaurant

Athanasios Makris

March 18, 2020

I. INTRODUCTION

This report describes a methodology on how to predict the most appropriate neighborhoods in London to open a new Mediterranean fusion restaurant. Mediterranean cuisine is getting a lot of traction as a diet which provides a plethora of health benefits, since it is based on large quantities of fresh fruits, vegetables, nuts, fish, and olive oil. Moreover, modern societies promote more and more a healthy and balanced way of life. This is especially profound in high population capitals of the western world in which the everyday routine is hectic and large segments of the population suffer from continuous stress and psychological issues. These people, especially the ones with higher income, continuously seek for new experiences that stimulate their senses in unique and unprecedented ways.

Thus, considering the above, if an entrepreneur would consider investing in the food business, a restaurant based on the Mediterranean cuisine would be an ideal candidate. Also, London is one of the most populated European capitals with a high percentage of people that are financially stable. Finally, modifying the restaurant cuisine with a fusion element, will further help the investment to succeed as it provides a niche variation in the taste and appearance of the menu.

However, what is the most appropriate neighborhood in London to open such a restaurant? To answer this question, we use the power of data and machine learning. In the following sections, we present the data we used and the methodology we applied in order to reach a conclusion. Finally, we discuss about our results and possible directions for future work.

II. DATA ACQUISITION AND PROCESSING

A. Data sources

In our analysis, we used multiple data sources in order to decide which are most suitable locations in London to open our restaurant. First, we extract the neighborhoods of London from Wikipedia [1]. Next, we use data from the YouGov's report [2] to exclude areas that this type or restaurant would not be an ideal option. Finally, we use Foursquare API [3] to analyze the existing venues in the most interesting neighborhoods and use these results to identify the most suitable neighborhoods for our investment.

B. Data processing

The first step in our analysis is to extract the neighborhoods of London from Wikipedia. In order to do this, we use the BeautifulSoup python library to scrape the webpage and load the data into a Pandas dataframe. An overview of the initial dataframe is presented in Figure 1.

	Location	London borough	Post town	Postcode district	Dial code	OS grid ref
0	Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2	020	TQ465785
1	Acton	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4	020	TQ205805
2	Addington	Croydon[8]	CROYDON	CR0	020	TQ375645
3	Addiscombe	Croydon[8]	CROYDON	CR0	020	TQ345665
4	Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728

Figure 1. Initial dataframe with London neighborhoods

Now, we have information about the neighborhoods of London in a structured format, but the data format can be improved in order to ease further processing. Specifically, we proceed to the following improvements:

- removal of unnecessary columns, such as *Dial code* or *OS grid ref*
- renaming of column names
- removal of reference numbers such as [7], [8], etc.
- replacement of 'and' with comma

An overview of the resulting dataframe is presented in Figure 2.

	Location	Borough	Post-town	Postcode
0	Abbey Wood	Bexley, Greenwich	LONDON	SE2
1	Acton	Ealing, Hammersmith, Fulham	LONDON	W3, W4
2	Addington	Croydon	CROYDON	CR0
3	Addiscombe	Croydon	CROYDON	CR0
4	Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14

Figure 2. London neighborhoods after initial processing

For the later steps in our analysis, we use the Foursquare API. Since its usage is not free, we need to limit the number of requests. Thus, we keep only the locations within the *Post-town* value equal to 'LONDON'. Then, we remove the Post-town column from the dataset as it is not necessary anymore. An overview of the final dataframe is presented in Figure 3.

	Location	Borough	Postcode
0	Abbey Wood	Bexley, Greenwich	SE2
1	Acton	Ealing, Hammersmith, Fulham	W3
1	Acton	Ealing, Hammersmith, Fulham	W4
6	Aldgate	City	EC3
7	Aldwych	Westminster	WC2

Figure 3. Final dataframe with London neighborhoods under consideration

III. METHODOLOGY

After the initial data acquisition and cleaning, we have the neighborhoods in London which are candidates for our investment. Next, we need to focus our analysis to only a subset of the neighborhoods. To achieve this, we use demographic data about the food habits of London habitants.

Next, we use the Foursquare API to extract the existing venues in the locations of interest. Foursquare API provides a lot of interesting data about a venue, however we focus only on the category of the venue and its pricing. Both of these attributes have direct relation to the type of the restaurant we want to open. Our restaurant's prices will be above the average since it will provide a unique fine dining experience. Also, we need to avoid neighborhoods in which there is a plethora of restaurants that have Mediterranean based cuisine.

Using Foursquare data, we first explore a single location in order to evaluate if the data are useful for our analysis based on the assumptions we have made. The, we apply k-means clustering to all neighborhoods of interest in order to find common patterns in the venues located to each one. Through this method, we identify which is the final set of neighborhoods that we consider for our investment.

IV. ANALYSIS

As described in the previous section, the first step in our analysis is to use demographic data to find the most suitable locations that will be further explored. Through internet search, we found a report from YouGov [2] that maps the areas of London to most popular cuisines based on people's profiles.

In this report, we see that people in West London mostly prefer Mediterranean based cuisines (French, Italian). We also know that West London is one of the most expensive locations in London and the majority of its habitants have high income.

These factors are crucial to the success of our restaurant. Our potential customers should prefer the Mediterranean cuisine and should also be able to appreciate and afford the fine dining experience we will offer.

Thus, we keep only the neighborhoods of the West London in our dataframe. In practice, this means keeping the locations in which postal code starts with 'W'. An overview of the resulting dataframe is presented in Figure 4.

	Location	Borough	Postcode
0	Acton	Ealing, Hammersmith, Fulham	W3
1	Aldwych	Westminster	WC2
2	Bayswater	Westminster	W2
3	Bedford Park	Ealing	W4
4	Bloomsbury	Camden	WC1
5	Charing Cross	Westminster	WC2
6	Chinatown	Westminster	W1
7	Chiswick	Hounslow, Ealing, Hammersmith, Fulham	W4
8	Covent Garden	Westminster	WC2
9	Ealing	Ealing	W5
10	Fitzrovia	Camden	W1
11	Grove Park	Hounslow	W4

Figure 4. West London neighborhoods

More than one neighborhood have the same postal code. We process the dataframe in order to combine them to one row with the neighborhoods separated with a comma. The process is followed for boroughs. The result is presented in Figure 5.

	Postcode	Borough	Location
0	W1	Westminster, Camden	Mayfair, Chinatown, Soho, Marylebone (also St ...
1	W10	Kensington, Chelsea	North Kensington
2	W11	Kensington, Chelsea	Notting Hill
3	W12	Fulham, Hammersmith	White City, Shepherd's Bush, Wormwood Scrubs
4	W13	Ealing	West Ealing
5	W14	Fulham, Hammersmith	West Kensington
6	W2	Westminster	Bayswater, Paddington
7	W3	Fulham, Hammersmith, Ealing	Acton
8	W4	Fulham, Hammersmith, Ealing, Hounslow	Gunnersbury, Chiswick, Bedford Park, Grove Park
9	W5	Ealing	Ealing
10	W6	Fulham, Hammersmith	Hammersmith
11	W7	Ealing	Hanwell
12	W8	Kensington, Chelsea	Holland Park
13	W9	Westminster	Maida Vale, Little Venice
14	WC1	Islington, Camden	Holborn, Bloomsbury, St Pancras, King's Cross
15	WC2	Westminster, Camden	St Giles, Covent Garden, Aldwych, Charing Cross

Figure 5. Processed dataframe in which all neighborhoods and boroughs have been grouped by postal code

Next, we need to extract the coordinates for each location. In order to do this, we use the Geocoder python package. Initially, we used the OpenStreetMap provider. However, the results were not accurate and the vast majority of locations resulted with the same coordinates. We considered to use Google Maps API, but there is no free quota. Thus, we concluded to use the Bing Maps API. Its precision is good and it offers free quota. The resulting dataframe is presented in Figure 6 while a visual representation of the locations under consideration is depicted in Figure 7.

	Postcode	Borough	Location	Latitude	Longitude
0	W1	Westminster, Camden	Mayfair, Chinatown, Soho, Marylebone (also St ...	51.516479	-0.14816
1	W10	Kensington, Chelsea	North Kensington	51.523289	-0.21343
2	W11	Kensington, Chelsea	Notting Hill	51.512249	-0.20626
3	W12	Fulham, Hammersmith	White City, Shepherd's Bush, Wormwood Scrubs	51.506451	-0.23691
4	W13	Ealing	West Ealing	51.514530	-0.31951
5	W14	Fulham, Hammersmith	West Kensington	51.495739	-0.20980
6	W2	Westminster	Bayswater, Paddington	51.514938	-0.18049
7	W3	Fulham, Hammersmith, Ealing	Acton	51.513241	-0.26746
8	W4	Fulham, Hammersmith, Ealing, Hounslow	Gunnersbury, Chiswick, Bedford Park, Grove Park	51.489441	-0.26194
9	W5	Ealing	Ealing	51.514069	-0.29991
10	W6	Fulham, Hammersmith	Hammersmith	51.496151	-0.22942
11	W7	Ealing	Hanwell	51.508781	-0.33630
12	W8	Kensington, Chelsea	Holland Park	51.501610	-0.19175
13	W9	Westminster	Maida Vale, Little Venice	51.525871	-0.19526
14	WC1	Islington, Camden	Holborn, Bloomsbury, St Pancras, King's Cross	51.524502	-0.12273
15	WC2	Westminster, Camden	St Giles, Covent Garden, Aldwych, Charing Cross	51.516510	-0.11967

Figure 6. West London locations with coordinates

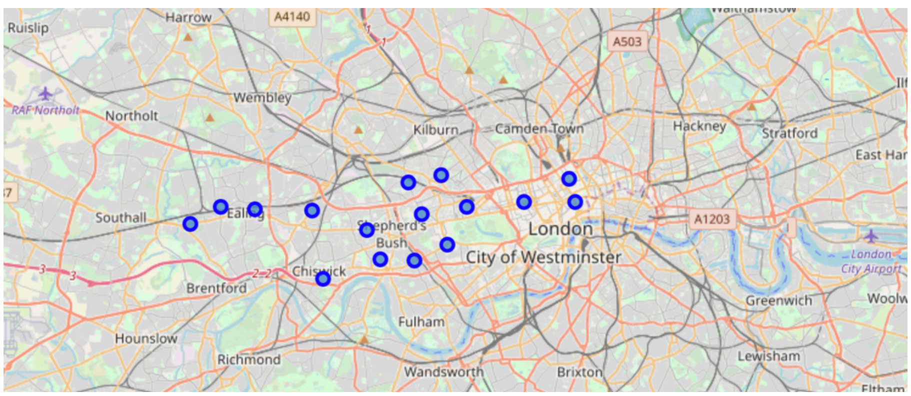

Figure 7. Visual representation of West London locations

In the next step in our analysis, we use the Foursquare API to extract the venues located in each of the locations under consideration. At this point we need to mention a very important parameter that affect the next parts of our analysis. Foursqaure API provides the option to filter venues by their pricing. It provides four pricing tiers 1-4, 1 being the least and 4 the most expensive. Based on the type of the restaurant we want to open and our pricing, we compete only with the most expensive restaurants that offer fine dining experience to their customer or similar. Thus, we filter the venues using pricing tiers 3 and 4.

In order to evaluate the validity of our approach, we first perform the analysis for only one postal code, which in our case is 'W1'. In Figure 8, an overview of venues for locations with postal code 'W1' is presented.

	name	categories	lat	lng
0	Le Relais de Venise L'Entrecôte	French Restaurant	51.518104	-0.151062
1	108 Brasserie	Lounge	51.517882	-0.150660
2	The Cavendish	Gastropub	51.519018	-0.149637
3	Roux at The Landau	French Restaurant	51.517932	-0.143651
4	Aubaine	French Restaurant	51.515092	-0.151904

Figure 8. Overview of venues in W1 postal code

We group the venues by category as this will help us check the competition and see what type of venues exist in the area. For locations with postal code 'W1', the venue categories are depicted in Figure 9.

From the result, we can see that in this location, the top venues are hotel bars and French restaurants. The number of hotel bars is very positive for us, as it's really common for someone to go out for dinner and then have a drink. Thus, this location would be of interest.

On the other hand, there are enough fine French restaurants with which we are direct competitors. So, we may need to hold back at the moment and explore all locations to see if there are any better ones.

	Count
Hotel Bar	5
French Restaurant	4
Lounge	3
Cocktail Bar	3
Modern European Restaurant	2

Figure 9. Grouping of venues in W1 postal code by category

Next, we extract the unique neighborhoods from the initial dataframe along with their respective coordinates. An overview of the resulting dataframe is presented in Figure 10. These data are used as input in our algorithm that extracts venues using the Foursquare API.

	Location	Latitude	Longitude
0	Chinatown	51.516479	-0.14816
1	Marylebone (also St Marylebone)	51.516479	-0.14816
2	Fitzrovia	51.516479	-0.14816
3	Soho	51.516479	-0.14816
4	Mayfair	51.516479	-0.14816
5	North Kensington	51.523289	-0.21343

Figure 10. Overview of the unique locations and their respective coordinates

In our exploration, we search for the top 100 venues within 500m radius from the center of each location. The result is stored in a dataframe for further analysis. An overview of this dataframe is presented in Figure 11.

	Location	Location Latitude	Location Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Chinatown	51.516479	-0.14816	Le Relais de Venise L'Entrecôte	51.518104	-0.151062	French Restaurant
1	Chinatown	51.516479	-0.14816	108 Brasserie	51.517882	-0.150660	Lounge
2	Chinatown	51.516479	-0.14816	The Cavendish	51.519018	-0.149637	Gastropub
3	Chinatown	51.516479	-0.14816	Roux at The Landau	51.517932	-0.143651	French Restaurant
4	Chinatown	51.516479	-0.14816	Aubaine	51.515092	-0.151904	French Restaurant

Figure 11. Overview of the dataframe with the top 100 venues within 500m of each unique location

Now that we have extracted all venues, we perform some initial exploration to get a general idea about the venues, their distribution among the various areas and so on. Thus, we first investigate which are the most "populated" areas. An overview of the result is presented in Figure 12. Our analysis shows that the most populated areas (in terms of venues) are Chinatown, Fitzrovia, Marylebone, Mayfair, and Soho, each having 23 venues.

	Location Latitude	Location Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Location						
Aldwych	9	9	9	9	9	9
Bayswater	10	10	10	10	10	10
Bedford Park	1	1	1	1	1	1
Bloomsbury	4	4	4	4	4	4
Charing Cross	9	9	9	9	9	9

Figure 12. Overview of the dataframe with unique venues per location

Next, we check which are the most common venue categories. The result is depicted in Figure 13. We see that hotel bars is the most common venue followed by French restaurants and cocktail bars. Based on our assumptions and these results, our goal is to find the areas that maximize the number of drink related venues (bars, clubs, etc.) and have as few restaurants as possible.

	Count
Hotel Bar	40
French Restaurant	31
Cocktail Bar	29
Lounge	15
Gastropub	12
Seafood Restaurant	11
Modern European Restaurant	10
Steakhouse	10
Tapas Restaurant	9
Restaurant	9
Italian Restaurant	7
Nightclub	5
Cycle Studio	5
Turkish Restaurant	4
Szechuan Restaurant	4
Beer Bar	4
Brasserie	4
Dim Sum Restaurant	2
Social Club	1
Café	1
Grocery Store	1
Sushi Restaurant	1
Hotel	1
Chinese Restaurant	1
Asian Restaurant	1
Gourmet Shop	1
Movie Theater	1
Diner	1

Figure 13. Grouping per category of all venues returned by Foursquare API

The next step in our analysis is to attempt to cluster the locations based on venue categories. First, we perform one hot encoding. An overview of the result is presented in Figure 14.

	Location	Asian Restaurant	Beer Bar	Brasserie	Café	Chinese Restaurant	Cocktail Bar	Cycle Studio	Dim Sum Restaurant	Diner	...	Movie Theater	Nightclub	Restaurant	Seafood Restaurant	Social Club	Stea
0	Chinatown	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
1	Chinatown	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
2	Chinatown	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
3	Chinatown	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	
4	Chinatown	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	

Figure 14. One hot encoding based on the venue category

Then, we use result to group the locations based on the venue categories. The result is depicted in Figure 15.

	Location	Asian Restaurant	Beer Bar	Brasserie	Café	Chinese Restaurant	Cocktail Bar	Cycle Studio	Dim Sum Restaurant	Diner	...	Movie Theater	Nightclub	Restaurant	Se Resta
0	Aldwych	0.000000	0.111111	0.111111	0.000000	0.000000	0.111111	0.000000	0.0	0.000000	...	0.000000	0.000000	0.111111	0.1
1	Bayswater	0.000000	0.000000	0.000000	0.000000	0.000000	0.100000	0.000000	0.1	0.000000	...	0.000000	0.000000	0.200000	0.1
2	Bedford Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	...	0.000000	0.000000	0.000000	0.0
3	Bloomsbury	0.000000	0.000000	0.000000	0.000000	0.000000	0.250000	0.000000	0.0	0.000000	...	0.000000	0.000000	0.000000	0.0
4	Charing Cross	0.000000	0.111111	0.111111	0.000000	0.000000	0.111111	0.000000	0.0	0.000000	...	0.000000	0.000000	0.111111	0.1
5	Chinatown	0.000000	0.000000	0.000000	0.000000	0.000000	0.130435	0.043478	0.0	0.000000	...	0.000000	0.043478	0.000000	0.0

Figure 15. Category grouping based on the ten most common venues per location

For each location, we keep only the top ten venue categories. An example is presented in Figure 16. We do this for all locations and we store the result to a new dataframe. This is depicted in Figure 17.

```
                venue  freq
0           Hotel Bar  0.22
1            Beer Bar  0.11
2    Tapas Restaurant  0.11
3  Seafood Restaurant  0.11
4          Restaurant  0.11
5  Turkish Restaurant  0.11
6           Brasserie  0.11
7        Cocktail Bar  0.11
8        Cycle Studio  0.00
9        Movie Theater 0.00
```

Figure 16. Example of category grouping for one location

	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Aldwych	Hotel Bar	Turkish Restaurant	Beer Bar	Brasserie	Seafood Restaurant	Restaurant	Cocktail Bar	Tapas Restaurant	Gourmet Shop	Café
1	Bayswater	Steakhouse	Restaurant	French Restaurant	Seafood Restaurant	Cocktail Bar	Dim Sum Restaurant	Hotel Bar	Turkish Restaurant	Gourmet Shop	Beer Bar
2	Bedford Park	French Restaurant	Turkish Restaurant	Tapas Restaurant	Beer Bar	Brasserie	Café	Chinese Restaurant	Cocktail Bar	Cycle Studio	Dim Sum Restaurant
3	Bloomsbury	Szechuan Restaurant	Cocktail Bar	Hotel Bar	Tapas Restaurant	Turkish Restaurant	Gourmet Shop	Beer Bar	Brasserie	Café	Chinese Restaurant
4	Charing Cross	Hotel Bar	Turkish Restaurant	Beer Bar	Brasserie	Seafood Restaurant	Restaurant	Cocktail Bar	Tapas Restaurant	Gourmet Shop	Café

Figure 17. Overview of the dataframe with the ten most common venues per location

At this point, we have all the necessary information to proceed to k-means cluster of the locations. However, first we need to decide what is the optimal number of clusters. There are various methods in the literature such as the Elbow [4] and the Silhouette [5] methods. In our analysis we use the Elbow method. We test from 1 to 10 clusters. As we can see in Figure 18, good values for k are 3 and 4. We select k equals to 4.

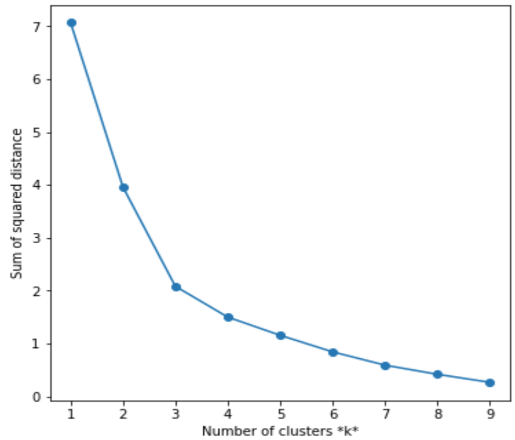

Figure 18. Elbow analysis for k-means algorithm

After applying the k-means algorithm, we enrich the dataframe with the most common venue categories per location with the cluster labels. An overview is presented in Figure 19.

	Cluster Labels	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	0	Aldwych	Hotel Bar	Turkish Restaurant	Beer Bar	Brasserie	Seafood Restaurant	Restaurant	Cocktail Bar	Tapas Restaurant	Gourmet Shop	Café
1	3	Bayswater	Steakhouse	Restaurant	French Restaurant	Seafood Restaurant	Cocktail Bar	Dim Sum Restaurant	Hotel Bar	Turkish Restaurant	Gourmet Shop	Beer Bar
2	1	Bedford Park	French Restaurant	Turkish Restaurant	Tapas Restaurant	Beer Bar	Brasserie	Café	Chinese Restaurant	Cocktail Bar	Cycle Studio	Dim Sum Restaurant
3	0	Bloomsbury	Szechuan Restaurant	Cocktail Bar	Hotel Bar	Tapas Restaurant	Turkish Restaurant	Gourmet Shop	Beer Bar	Brasserie	Café	Chinese Restaurant
4	0	Charing Cross	Hotel Bar	Turkish Restaurant	Beer Bar	Brasserie	Seafood Restaurant	Restaurant	Cocktail Bar	Tapas Restaurant	Gourmet Shop	Café
5	3	Chinatown	Hotel Bar	French Restaurant	Cocktail Bar	Lounge	Modern European Restaurant	Nightclub	Cycle Studio	Italian Restaurant	Gastropub	Seafood Restaurant

Figure 19. Overview of the dataframe with the ten most common venues per location enhanced with cluster labels

Finally, we enrich the dataframe with the coordinates of each location in order to be able to visualize the results on a map. An overview of the final dataframe is presented in Figure 20.

	Cluster Labels	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Latitude	Longi
0	0	Aldwych	Hotel Bar	Turkish Restaurant	Beer Bar	Brasserie	Seafood Restaurant	Restaurant	Cocktail Bar	Tapas Restaurant	Gourmet Shop	Café	51.516510	-0.1
1	3	Bayswater	Steakhouse	Restaurant	French Restaurant	Seafood Restaurant	Cocktail Bar	Dim Sum Restaurant	Hotel Bar	Turkish Restaurant	Gourmet Shop	Beer Bar	51.514938	-0.1
2	1	Bedford Park	French Restaurant	Turkish Restaurant	Tapas Restaurant	Beer Bar	Brasserie	Café	Chinese Restaurant	Cocktail Bar	Cycle Studio	Dim Sum Restaurant	51.489441	-0.2
3	0	Bloomsbury	Szechuan Restaurant	Cocktail Bar	Hotel Bar	Tapas Restaurant	Turkish Restaurant	Gourmet Shop	Beer Bar	Brasserie	Café	Chinese Restaurant	51.524502	-0.1
4	0	Charing Cross	Hotel Bar	Turkish Restaurant	Beer Bar	Brasserie	Seafood Restaurant	Restaurant	Cocktail Bar	Tapas Restaurant	Gourmet Shop	Café	51.516510	-0.1
5	3	Chinatown	Hotel Bar	French Restaurant	Cocktail Bar	Lounge	Modern European Restaurant	Nightclub	Cycle Studio	Italian Restaurant	Gastropub	Seafood Restaurant	51.516479	-0.1

Figure 20. Overview of the dataframe with the ten most common venues per location enhanced with cluster labels and coordinates

Figure 21 presents the clusters of the various locations based on the venue category types. In order to identify what differentiates each cluster, we will extract a separate dataframe for each of them. The resulting dataframes are presented in Figure 22 (cluster 0), Figure 23 (cluster 1), Figure 24 (cluster 2), and Figure 25 (cluster 3).

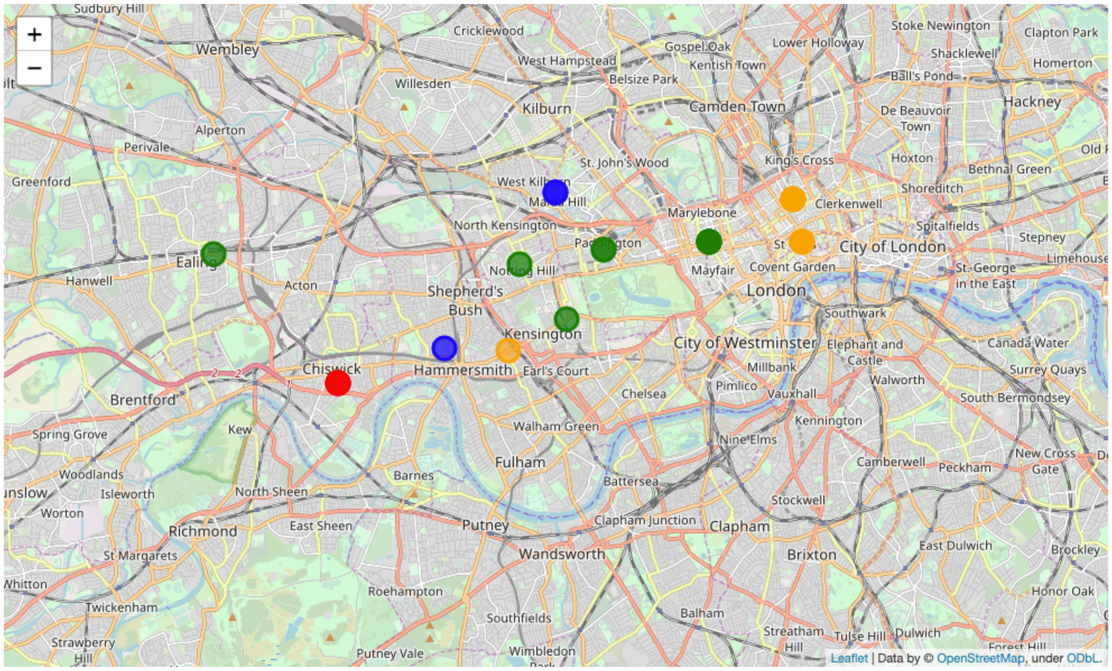

Figure 21. Visual representation of the resulting clusters

	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Aldwych	Hotel Bar	Turkish Restaurant	Beer Bar	Brasserie	Seafood Restaurant	Restaurant	Cocktail Bar	Tapas Restaurant	Gourmet Shop	Café
3	Bloomsbury	Szechuan Restaurant	Cocktail Bar	Hotel Bar	Tapas Restaurant	Turkish Restaurant	Gourmet Shop	Beer Bar	Brasserie	Café	Chinese Restaurant
4	Charing Cross	Hotel Bar	Turkish Restaurant	Beer Bar	Brasserie	Seafood Restaurant	Restaurant	Cocktail Bar	Tapas Restaurant	Gourmet Shop	Café
7	Covent Garden	Hotel Bar	Turkish Restaurant	Beer Bar	Brasserie	Seafood Restaurant	Restaurant	Cocktail Bar	Tapas Restaurant	Gourmet Shop	Café
13	Holborn	Szechuan Restaurant	Cocktail Bar	Hotel Bar	Tapas Restaurant	Turkish Restaurant	Gourmet Shop	Beer Bar	Brasserie	Café	Chinese Restaurant
15	King's Cross	Szechuan Restaurant	Cocktail Bar	Hotel Bar	Tapas Restaurant	Turkish Restaurant	Gourmet Shop	Beer Bar	Brasserie	Café	Chinese Restaurant
23	St Giles	Hotel Bar	Turkish Restaurant	Beer Bar	Brasserie	Seafood Restaurant	Restaurant	Cocktail Bar	Tapas Restaurant	Gourmet Shop	Café
24	St Pancras	Szechuan Restaurant	Cocktail Bar	Hotel Bar	Tapas Restaurant	Turkish Restaurant	Gourmet Shop	Beer Bar	Brasserie	Café	Chinese Restaurant
25	West Kensington	Hotel	Cocktail Bar	Restaurant	Hotel Bar	Grocery Store	Beer Bar	Brasserie	Café	Chinese Restaurant	Cycle Studio

Figure 22. Cluster 0 most common venue categories

	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	Bedford Park	French Restaurant	Turkish Restaurant	Tapas Restaurant	Beer Bar	Brasserie	Café	Chinese Restaurant	Cocktail Bar	Cycle Studio	Dim Sum Restaurant
6	Chiswick	French Restaurant	Turkish Restaurant	Tapas Restaurant	Beer Bar	Brasserie	Café	Chinese Restaurant	Cocktail Bar	Cycle Studio	Dim Sum Restaurant
10	Grove Park	French Restaurant	Turkish Restaurant	Tapas Restaurant	Beer Bar	Brasserie	Café	Chinese Restaurant	Cocktail Bar	Cycle Studio	Dim Sum Restaurant
11	Gunnersbury	French Restaurant	Turkish Restaurant	Tapas Restaurant	Beer Bar	Brasserie	Café	Chinese Restaurant	Cocktail Bar	Cycle Studio	Dim Sum Restaurant

Figure 23. Cluster 1 most common venue categories

	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
12	Hammersmith	Gastropub	Tapas Restaurant	Cocktail Bar	Turkish Restaurant	Beer Bar	Brasserie	Café	Chinese Restaurant	Cycle Studio	Dim Sum Restaurant
16	Little Venice	Gastropub	Turkish Restaurant	Tapas Restaurant	Beer Bar	Brasserie	Café	Chinese Restaurant	Cocktail Bar	Cycle Studio	Dim Sum Restaurant
17	Maida Vale	Gastropub	Turkish Restaurant	Tapas Restaurant	Beer Bar	Brasserie	Café	Chinese Restaurant	Cocktail Bar	Cycle Studio	Dim Sum Restaurant

Figure 24. Cluster 2 most common venue categories

	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Bayswater	Steakhouse	Restaurant	French Restaurant	Seafood Restaurant	Cocktail Bar	Dim Sum Restaurant	Hotel Bar	Turkish Restaurant	Gourmet Shop	Beer Bar
5	Chinatown	Hotel Bar	French Restaurant	Cocktail Bar	Lounge	Modern European Restaurant	Nightclub	Cycle Studio	Italian Restaurant	Gastropub	Seafood Restaurant
8	Ealing	Steakhouse	French Restaurant	Turkish Restaurant	Grocery Store	Beer Bar	Brasserie	Café	Chinese Restaurant	Cocktail Bar	Cycle Studio
9	Fitzrovia	Hotel Bar	French Restaurant	Cocktail Bar	Lounge	Modern European Restaurant	Nightclub	Cycle Studio	Italian Restaurant	Gastropub	Seafood Restaurant
14	Holland Park	French Restaurant	Grocery Store	Sushi Restaurant	Chinese Restaurant	Cocktail Bar	Italian Restaurant	Turkish Restaurant	Beer Bar	Brasserie	Café
18	Marylebone (also St Marylebone)	Hotel Bar	French Restaurant	Cocktail Bar	Lounge	Modern European Restaurant	Nightclub	Cycle Studio	Italian Restaurant	Gastropub	Seafood Restaurant
19	Mayfair	Hotel Bar	French Restaurant	Cocktail Bar	Lounge	Modern European Restaurant	Nightclub	Cycle Studio	Italian Restaurant	Gastropub	Seafood Restaurant
20	Notting Hill	Asian Restaurant	Gourmet Shop	Café	Social Club	Cocktail Bar	Movie Theater	Diner	Italian Restaurant	Gastropub	Beer Bar
21	Paddington	Steakhouse	Restaurant	French Restaurant	Seafood Restaurant	Cocktail Bar	Dim Sum Restaurant	Hotel Bar	Turkish Restaurant	Gourmet Shop	Beer Bar
22	Soho	Hotel Bar	French Restaurant	Cocktail Bar	Lounge	Modern European Restaurant	Nightclub	Cycle Studio	Italian Restaurant	Gastropub	Seafood Restaurant

Figure 25. Cluster 3 most common venue categories

This part concludes our analysis. In the next section, we present the results of our exploration and conclude to the locations that are the most promising (based on the data we have available) to consider for our investment.

V. RESULTS

From the individual dataframes per cluster, we can identify what are the common attributes of the venues in the cluster. Thus, we have the following:

In cluster 1, we see that the most common venue is Gastrobub followed by Turkish restaurant as the second most common venue. Locations in cluster 1 are not very interesting for us as the most common venues serve food.

In cluster 2, the three most common venues are French, Turkish, and Tapas restaurants. Again, these are not the locations that we would consider.

In cluster 3, the most common venue is Lounge. However, the second most common is Turkish restaurant. So, we may consider locations in this cluster, but further analysis is required to closely evaluate the competition and decide to proceed or not.

In cluster 4, we see that the most common venue is either a hotel bar or Asian restaurant. There are also various types of venues as second most common ones. This makes this cluster the most interesting of all, since there is space for the new restaurant we want to open. The locations in this cluster are the first we would further examine in order to reach a final decision.

VI. CONCLUSION

Using date from a variety of sources, we evaluated the neighborhoods of West London in order to find the most appropriate ones in order to open a Mediterranean fusion restaurant.

In our analysis, we concluded to the subset of areas that seems the most promising. However, these data are not enough for a final decision and work more as a guide. Further analysis, both quantitative and qualitative, is required in order to limit the number of

potentials areas of interest.
Such analysis would include computations regarding distances among competing restaurants, distances from means of public transport, distances from suppliers, etc. Also, the cost of lease and the availability of spaces to open a restaurant are impotent factors.
Finally, qualitative metrics such as customer satisfaction from existing restaurants, customer perception of each neighborhood, etc. would also affect the final decision.

REFERENCES

[1] Wikipedia, List of areas of London, https://en.wikipedia.org/wiki/List_of_areas_of_London
[2] YouGov, London's food tastes – mapped, https://yougov.co.uk/topics/politics/articles-reports/2015/10/13/mapped-londons-food-tastes
[3] Foursquare API, https://developer.foursquare.com/docs
[4] Elbow method, https://en.wikipedia.org/wiki/Elbow_method_(clustering)
[5] Silhouette method, https://en.wikipedia.org/wiki/Silhouette_(clustering)