Brexit: A Challenge or Opportunity for Catering Industry - Data Analysis of London Restaurant

Tsang Chiu Cheuk

March 01, 2020

Introduction

1.1 Background

Twenty-Twenty has a landmark ring to it. The Start of a new decade prompts bigger-than-usual thought about the future. Britain's reckoning with Brexit will leave a significant impact on the hospitality and food industry. However, Brexit does offer the potential for re-setting the direction of food business throughout the UK.

Throughout its history, London's dominance has often been portrayed as a 'problem' for its host nation. Apart from that, it is simultaneously the political, economic, business and cultural art centre. Therefore, London is very diverse. There are many different restaurants around London, including French, Italian, Asian, African, Middle Eastern and American ones.

1.2 Business Problem Description

As an investor agent, we are looking to take its opportunity to explore expansion after the Brexit. Lack of confidence, some restaurants are decided to shut down their business and moved back to their countries. Along with recent organic trend, many restaurants/cafés are available to rent in this moment. Traditionally, London is overwhelmingly seen as expensive and inaccessible. The preliminary target is to create a high-end / fine dining restaurants with organic mix and healthy. In order to survive in such competitive market, a strategic plan is essential. Wanna bet? Below factors will be explored in order to decide on the physical location in London:

1. London Population
2. London Demographics
3. The location competitors
4. Cuisine and Ingredient served by competitors
5. Segmentation of Market

When we consider all these factors, we can create a map and information chart where to establish the restaurant, ingredient used and each district is clustered according to the venue density.

1.3 Target Audience

Considering the diversity of London, there is a high multicultural sense. It ranges from Londoners, tourists and those who are passionate about organic food. To recommend the correct location, our company has appointed me to work on Data Science project for researching the best choice. This research would help anyone who wants to start their catering business on organic food in London after the Brexit.

1.4 Success Criteria

The success criteria of the project will be a good recommendation of borough/Neighborhood choice and nearest suppliers of ingredients for organic restaurant.

Data Description

2.1 Data Sources

To consider the problem, we can list the data as below:

One city will be analyzed in this project: London. And this project will base on the public data from Wikipedia and Foursquare.

In the Wikipedia, the London is treated as synonymous to the "Great London Area", meaning there are the areas within the London Area Postcode. In my project, we focus only in the neighborhoods within the London Post Code. Hence, the London area consists 32 Boroughs and the "City of London". The dataset exists for free on the web. Link to the dataset is https://en.wikipedia.org/wiki/List_of_areas_of_London

	Location	Borough	Post-town	Postcode	Dial-code	OSGridRef
0	Abbey Wood	Bexley, Greenwich	LONDON	SE2	020	TQ465785
1	Acton	Ealing, Hammersmith and Fulham	LONDON	W3, W4	020	TQ205805
2	Addington	Croydon	CROYDON	CR0	020	TQ375645
3	Addiscombe	Croydon	CROYDON	CR0	020	TQ345665
4	Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728

Diagram 1: Extract from Wiki page

Based on the wiki page, we have below **six** assumption:

Assumption 1: We will have multi-rows and assigned the same values from the other columns when the postcode is found more than one

	Location	Borough	Post-town	Dial-code	OSGridRef	Postcode
0	Abbey Wood	Bexley, Greenwich	LONDON	020	TQ465785	SE2
1	Acton	Ealing, Hammersmith and Fulham	LONDON	020	TQ205805	W3
1	Acton	Ealing, Hammersmith and Fulham	LONDON	020	TQ205805	W4
10	Angel	Islington	LONDON	020	TQ345665	EC1
10	Angel	Islington	LONDON	020	TQ345665	N1

Diagram 2: Assumption 1- Extract

Assumption 2: We will use the 'Location', 'Borough', 'Postcode', 'Post-town' fields and extracted into a new data frame

	Location	Borough	Postcode	Post-town
0	Abbey Wood	Bexley, Greenwich	SE2	LONDON
1	Acton	Ealing, Hammersmith and Fulham	W3	LONDON
2	Acton	Ealing, Hammersmith and Fulham	W4	LONDON
3	Angel	Islington	EC1	LONDON
4	Angel	Islington	N1	LONDON

Diagram 3: Assumption 2- Extract

Assumption 3: We assume that the restaurant open for the Boroughs with London Post-town. Therefore, all the non-post-town are dropped.

	Location	Borough	Postcode	Post-town
0	Abbey Wood	Bexley, Greenwich	SE2	LONDON
1	Acton	Ealing, Hammersmith and Fulham	W3	LONDON
2	Acton	Ealing, Hammersmith and Fulham	W4	LONDON
3	Angel	Islington	EC1	LONDON
4	Angel	Islington	N1	LONDON

```
df21.shape
```

```
(762, 4)
```

Diagram 4: Assumption 3 - Extract

Assumption 4: Only the South East areas of London will be used for our analysis because of its more diverse outlook, proximity to afro-caribbean markets and accessible facilities. It is given that the South East areas has postcodes with SE.

	Location	Borough	Postcode
0	Abbey Wood	Bexley, Greenwich	SE2
1	Tulse Hill	Lambeth	SE24
2	Tulse Hill	Lambeth	SE27
3	Upper Norwood	Croydon	SE19
4	Walworth	Southwark	SE17
5	Well Hall	Greenwich	SE9
6	West Heath	Bexley	SE2
7	West Norwood	Lambeth	SE27
8	Westcombe Park	Greenwich	SE3
9	Woolwich	Greenwich	SE18

Diagram 5: Assumption 4 – Extract

Assumption 5: We will check the demography of London, mapping the proportion of races by London borough

	Local authority	White	Mixed	Asian	Black	Other
22	Lewisham	53.5	7.4	9.3	27.2	2.6
27	Southwark	54.3	6.2	9.4	26.9	3.3
21	Lambeth	57.1	7.6	6.9	25.9	2.4
11	Hackney	54.7	6.4	10.5	23.1	5.3
7	Croydon	55.1	6.6	16.4	20.2	1.8

Diagram 6: Assumption 5 - Extract

Assumption 6*: We got a significantly founding from above result, Lewisham, Southward, Lambeth, Hackney and Croydon have a relatively high results in Black. This may produce an opportunity for us to drill down on the catering and eating habit for the Black. Also, it produces a physical needs to set up Black restaurant*

	Location	Borough	Postcode
0	Abbey Wood	Bexley, Greenwich	SE2
1	Tulse Hill	Lambeth	SE24
2	Tulse Hill	Lambeth	SE27
3	Upper Norwood	Croydon	SE19
4	Walworth	Southwark	SE17
5	Well Hall	Greenwich	SE9

Diagram 7: Assumption 6 – Extract

There are now 92 instances in South East London area for Black analysis. Besides, the Geocoder package is used to obtain South East London data area such as the latitude and longitude for further analysis.

	Location	Borough	Postcode	Latitude	Longitude
0	Tulse Hill	Lambeth	SE24	51.45529	-0.09928
1	Tulse Hill	Lambeth	SE27	51.43407	-0.10375
2	Upper Norwood	Croydon	SE19	51.41990	-0.08808
3	Walworth	Southwark	SE17	51.48764	-0.09542
4	West Norwood	Lambeth	SE27	51.43407	-0.10375

Diagram 8: Add Latitude and Longitude

Besides, the Foursquare API will be used to get the geographical location for the London area, as well as leveraging to provision venues information for each neighborhood. Then, the venue will provide the categories needed for the analysis to determine the viability of the selected location for the restaurant. Link for the Foursquare API is https://developer.foursquare.com/


```
{'meta': {'code': 200, 'requestId': '5e5b6205542890001b0c25aa'},
 'response': {'suggestedFilters': {'header': 'Tap to show:',
   'filters': [{'name': 'Open now', 'key': 'openNow'}]},
  'headerLocation': 'Lewisham Central',
  'headerFullLocation': 'Lewisham Central, London',
  'headerLocationGranularity': 'neighborhood',
  'totalResults': 178,
  'suggestedBounds': {'ne': {'lat': 51.47996001800005,
    'lng': 0.021296961190459426},
   'sw': {'lat': 51.44395998200002, 'lng': -0.03637696119035749}},
  'groups': [{'type': 'Recommended Places',
    'name': 'recommended',
    'items': [{'reasons': {'count': 0,
       'items': [{'summary': 'This spot is popular',
         'type': 'general',
         'reasonName': 'globalInteractionReason'}]},
      'venue': {'id': '535823bc498ec8d8da9aad5f',
       'name': 'Street Feast Model Market',
       'location': {'address': '196 Lewisham High St',
        'crossStreet': 'entrance at Molesworth St',
```


Diagram 9: Foursquare Extract

2.2 How the data used to solve the problem

Both datasets can consider the venue within the neighborhood of London Postcode area. Within a certain mile radius, the restaurant can be treated as the types of restaurant. Because of the Foursquare API constraint, the limit is 100 venues and the radius is 750 meter for each borough from their given latitude and longitude information. The accessibility and ease of supplies of ingredients are considered. Also, amenities and transportation connection are correlated.

Methodology

Our primary goal is to get the optimum location for new restaurant business in London area. The Foursquare workability can help to examine it. The Lewisham Borough postcode SE13 and Location can be studied.

3.1 Exploratory Data Analysis

First of all, the exploratory data analysis is executed. I load the data and explore the data from json file. Then, I use the pandas dataframe to transform the data into the nested dictionaries by python language in juypter notebook. More importantly, this dataframe has its geographical coordinates of London city

neighborhoods. The data of Foursquare can be used for analysis. Apart from that, London city map can be created with neighborhoods with geopy and folium libraries.

Then, the latitude and longitude values of London can be obtained. The South East London neighborhoods are zoomed to see the superimposed areas. Surprisingly, it is found that there is no African restaurant found in Lewisham. Based on the statistic, each neighborhood are regrouped and categorized.

Single Neighborhood

First of all, we will set up a single Neighborhood within the London area was done to examine the Foursquare workability. The Lewisham Borough postcode SE13 and Location - Lewisham is our case.

	name	categories	lat	lng
0	Street Feast Model Market	Street Food Gathering	51.460209	-0.012199
1	Maggie's Kitchen	Café	51.465380	-0.011213
2	Levante restaurant	Restaurant	51.462072	-0.009491
3	Corte	Coffee Shop	51.459776	-0.011554
4	Manor House Gardens	Park	51.456686	0.004684
5	Dirty South	Pub	51.458846	-0.002666
6	Levante Pide Restaurant	Turkish Restaurant	51.459848	-0.011476

Diagram 10: Single Neighborhood - Extract result

	Count
Pub	14
Café	9
Gastropub	7
Park	5
Garden	4

Diagram 11: Group the categories in Single Neigbourhoods

Multiple Neighbourhoods

	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighbourhood						
Bankside	200	200	200	200	200	200
Bellingham	134	134	134	134	134	134
Bermondsey	200	200	200	200	200	200
Blackheath	178	178	178	178	178	178
Brixton	200	200	200	200	200	200
Brockley	200	200	200	200	200	200
Camberwell	200	200	200	200	200	200
Catford	134	134	134	134	134	134

Diagram 12: Multiple Neighorhood Result – Extract

There are 197 unique categories.

	Count
Pub	904
Coffee Shop	604
Café	546
Park	412
Grocery Store	292

Diagram 13 Group the categories in Multiple Neigbourhoods

	Count
count	197.000000
mean	42.923858
std	97.401647
min	2.000000
25%	8.000000
50%	14.000000
75%	38.000000
max	904.000000

Diagram 14: The statistical summary of Multiple Neighourhoods

	Neighbourhood	African Restaurant	American Restaurant	Antique Shop	Aquarium	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Ath & S
2387	Lewisham	0	0	0	0	0	0	0	0	0	0
2388	Lewisham	0	0	0	0	0	0	0	0	0	0
2389	Lewisham	0	0	0	0	0	0	0	0	0	0
2390	Lewisham	0	0	0	0	0	0	0	0	0	0
2391	Lewisham	0	0	0	0	0	0	0	0	0	0
2392	Lewisham	0	0	0	0	0	0	0	0	0	0

Diagram 15: No African restaurant is found

3.2 Predictive Modeling

Finally, I create clusters of the neighborhood using the k-means to cluster the neighborhood into 5 clusters. The **top 10** venues can be labelled. To select the optimal k, the elbow method is used, together with the Sum of the Squared Error (SSE). In this method, the cluster is selected with the minimum SSE. After trying 500 iteration, optimal k is 3. In order to prevent the statistical bias method, Silhouette coefficient is tried. The rule is that the highest Sihouette Coefficent gives the best cluster. The cluster value is suggested to five.

Diagram 16 Add marker to neighborhood

The top venue are labelled as below:

```
----Bankside----
                venue  freq
0         Coffee Shop  0.10
1                 Pub  0.07
2               Hotel  0.07
3  Italian Restaurant  0.05
4             Theater  0.05
5      Scenic Lookout  0.03
6        Cocktail Bar  0.03
7          Art Museum  0.03
8  Seafood Restaurant  0.03
9        Burger Joint  0.02


----Bellingham----
                 venue  freq
0         Grocery Store  0.12
1                  Park  0.09
2           Supermarket  0.07
3                 Café  0.06
4  Fast Food Restaurant  0.04
5    Italian Restaurant  0.04
6                   Pub  0.04
7           Coffee Shop  0.04
8         Train Station  0.03
9           Gas Station  0.03


----Bermondsey----
                 venue  freq
0          Coffee Shop  0.10
1                  Pub  0.07
2                Hotel  0.07
3  Italian Restaurant  0.05
4              Theater  0.05
5      Scenic Lookout  0.03
6        Cocktail Bar  0.03
7           Art Museum  0.03
8  Seafood Restaurant  0.03
9        Burger Joint  0.02


----Blackheath----
                 venue  freq
0                  Pub  0.15
1          Coffee Shop  0.06
2                 Park  0.06
3        Grocery Store  0.06
4                 Café  0.04
5       Clothing Store  0.04
6          Supermarket  0.03
7               Bakery  0.03
8  Italian Restaurant  0.03
9               Garden  0.03
```

	Location	Borough	Postcode	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th I Com Venu
0	Tulse Hill	Lambeth	SE24	51.45529	-0.09928	1	Coffee Shop	Pub	Café	Grocery Store	Bakery	Park
1	Tulse Hill	Lambeth	SE27	51.43407	-0.10375	1	Coffee Shop	Pub	Café	Grocery Store	Bakery	Park
2	Upper Norwood	Croydon	SE19	51.41990	-0.08808	4	Pub	Café	Park	Coffee Shop	Italian Restaurant	Groc Store
3	Walworth	Southwark	SE17	51.48764	-0.09542	3	Pub	Café	Coffee Shop	Park	Pizza Place	Italia Rest

Diagram 17: Data frame for analysis

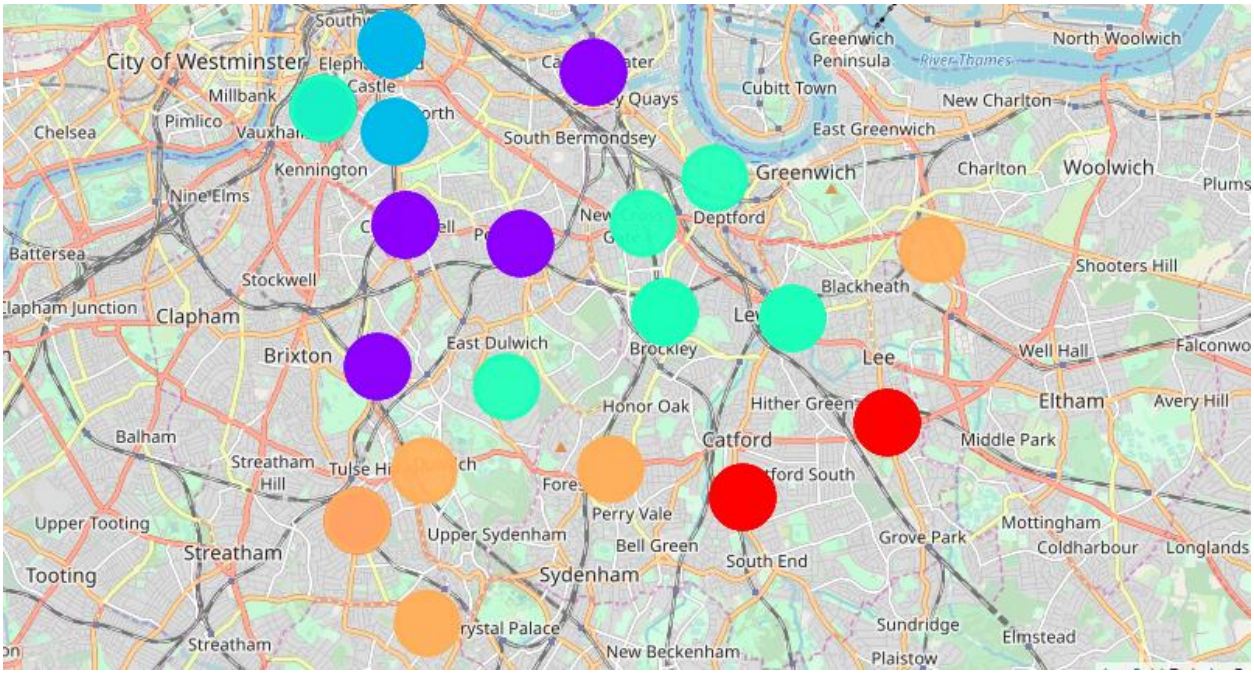

Diagram 18 Visualizing clusters in map

To select the optimal k, the **elbow method** is used, together with the Sum of the Squared Error (SSE). In this method, the cluster is selected with the minimum SSE. After trying 500 iteration, optimal k is 3

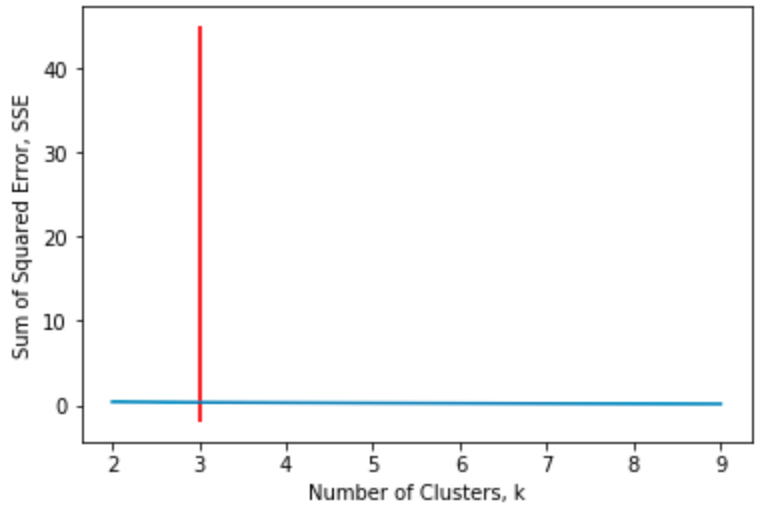

Diagram 19: Elbow method result

In order to prevent the statistical bias method, **Silhouette coefficient** is tried. The rule is that the highest Sihouette Coefficent gives the best cluster.

Cluster 1:

	Borough	Cluster Labels	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
17	Lewisham	0	Grocery Store	Pub	Park	Café	Supermarket	Gym / Fitness Center	Italian Restaurant	Indian Restaurant	Train Station	Coffee Shop
24	Lewisham	0	Grocery Store	Pub	Park	Café	Supermarket	Gym / Fitness Center	Italian Restaurant	Indian Restaurant	Train Station	Coffee Shop
31	Lewisham	0	Grocery Store	Park	Supermarket	Café	Coffee Shop	Pub	Fast Food Restaurant	Italian Restaurant	Train Station	Gas Station
37	Lewisham	0	Grocery Store	Park	Supermarket	Café	Coffee Shop	Pub	Fast Food Restaurant	Italian Restaurant	Train Station	Gas Station

Diagram 20: Cluster 1 Result – Extract

Cluster 2:

	Borough	Cluster Labels	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	Lambeth	1	Coffee Shop	Pub	Café	Grocery Store	Bakery	Park	Pizza Place	Brewery	Gym / Fitness Center	Tapas Restaurant
1	Lambeth	1	Coffee Shop	Pub	Café	Grocery Store	Bakery	Park	Pizza Place	Brewery	Gym / Fitness Center	Tapas Restaurant
6	Southwark	1	Café	Coffee Shop	Pub	Park	Cocktail Bar	Italian Restaurant	Middle Eastern Restaurant	Brewery	Grocery Store	Pizza Place
18	Lambeth	1	Coffee Shop	Café	Pub	Pizza Place	Brewery	Market	Bakery	Burger Joint	Restaurant	Cocktail Bar
29	Southwark	1	Pub	Café	Pizza Place	Coffee Shop	Park	Bar	Indie Movie Theater	Cocktail Bar	Art Gallery	Italian Restaurant

Diagram 21 Cluster 2 Result – Extract

Cluster 3:

	Borough	Cluster Labels	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
10	Southwark	2	Pub	Café	Coffee Shop	Hotel	Theater	Park	Italian Restaurant	Street Food Gathering	Pizza Place	Art Gallery
11	Southwark	2	Pub	Café	Coffee Shop	Hotel	Theater	Park	Italian Restaurant	Street Food Gathering	Pizza Place	Art Gallery
12	Southwark	2	Pub	Café	Coffee Shop	Hotel	Theater	Park	Italian Restaurant	Street Food Gathering	Pizza Place	Art Gallery
13	Southwark	2	Coffee Shop	Pub	Hotel	Italian Restaurant	Theater	Scenic Lookout	Cocktail Bar	Seafood Restaurant	Art Museum	Pizza Place

Diagram 22 Cluster 3 Result – Extract

Cluster 4:

	Borough	Cluster Labels	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
3	Southwark	3	Pub	Café	Coffee Shop	Park	Pizza Place	Italian Restaurant	Art Gallery	Theater	Cricket Ground	Brewery
5	Lewisham	3	Pub	Coffee Shop	Café	Park	Bar	Gastropub	Indian Restaurant	Cocktail Bar	Italian Restaurant	Hotel
7	Lewisham	3	Pub	Coffee Shop	Café	Bar	Park	Gastropub	Pizza Place	Cocktail Bar	Italian Restaurant	Indie Movie Theater
9	Southwark	3	Pub	Café	Pizza Place	Italian Restaurant	Coffee Shop	Park	Gastropub	Garden Center	Cocktail Bar	Mediterranean Restaurant

Diagram 23 Cluster 4 Result – Extract

Cluster 5:

	Borough	Cluster Labels	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	Croydon	4	Pub	Café	Park	Coffee Shop	Italian Restaurant	Grocery Store	Train Station	Gastropub	Bakery	Breakfast Spot
4	Lambeth	4	Pub	Coffee Shop	Grocery Store	Café	Park	Bakery	Pizza Place	Gym / Fitness Center	Pharmacy	Indian Restaurant
8	Southwark	4	Pub	Café	Grocery Store	Coffee Shop	Bakery	Park	Brewery	Pizza Place	Farmers Market	Italian Restaurant
14	Lewisham	4	Pub	Coffee Shop	Grocery Store	Café	Park	Supermarket	Gym / Fitness Center	Pizza Place	Japanese Restaurant	Italian Restaurant

Diagram 24 Cluster 5 Result – Extract

Result

Based on the above analysis,

1) The favorite restaurant in South East London is pubs, café and coffee shop.
2) There is a “blue market” for Africans in Lewisham area, South East Area.
3) Apart from that, Italian restaurant are famous in South East London Area such as Lambeth area and Southward

Discussion

It is important to note that the Clusters have variation, the most dominance around London is pubs. In Cluster 2 and 3, it discovered that there is a ‘golden nutrient’ to create African restaurant. It is attributed to the proximity to other amenities and accessibility to stations. As there is such a complexity, very different approaches can be tried in clustering and classification studies. For more detailed and accurate guidance, the data set can be expanded and the details of the neighborhood or street can be extended. I ended the study by visualizing the clustering information in South East London.

Conclusion

In conclusion, this project is performed within limited data. This may be right or may be wrong. It would credit a better results if there were more data within the area, traffic access and allowance of more venues exploration with the Foursquare. Also, the ratings and feedbacks of the current restaurants within the clusters would have helped in providing more insights into the best location after Brexit.