BATTLE OF THE BRANDS: CAPSTONE FINAL PROJECT FIND OUT WHERE TO GO EAT IN GREAT CAIRO

By: Osama Mohamed Mostafa AbdElMageid Ghozlan

AGENDA

Introduction & Business Problem

Data: Requirements, Acquisition and Usage

Project Methodology

Results

INTRODUCTION

In Egypt, many outings revolve around having a delicious meal followed by spending some time in Cafe. With the large number of opening restaurants offering different cuisine, it becomes rather hard to determine which restaurants are worth going to and how far they are.

BUSINESS PROBLEM

This project aims to provide simple guide on where to eat based on Foursquare ratings, price, category and geographic location data for restaurants in Great Cairo. The restaurants would be grouped (clustered) based on their similarities so that it can be easily determined what type cuisine they are looking for and in its price range.

DATA REQUIRED

For this task, Foursquare API will be used to retrieve data for restaurants in Great Cairo:

- Venue Name
- Venue ID
- Venue Location
- Venue Category
- Rating
- Price
- Likes
- Phrases

DATA ACQUISITION

To acquire the needed data the following shall be done:

- Retrieve list of districts in Great Cairo
- Use geolocator to obtain coordinates (latitude and longitude) for fetched districts
- Retrieve a list of all venues in these districts
- Retrieve venue IDs for each venue
- Retrieve venue name, location, category, etc.. for each venue

7
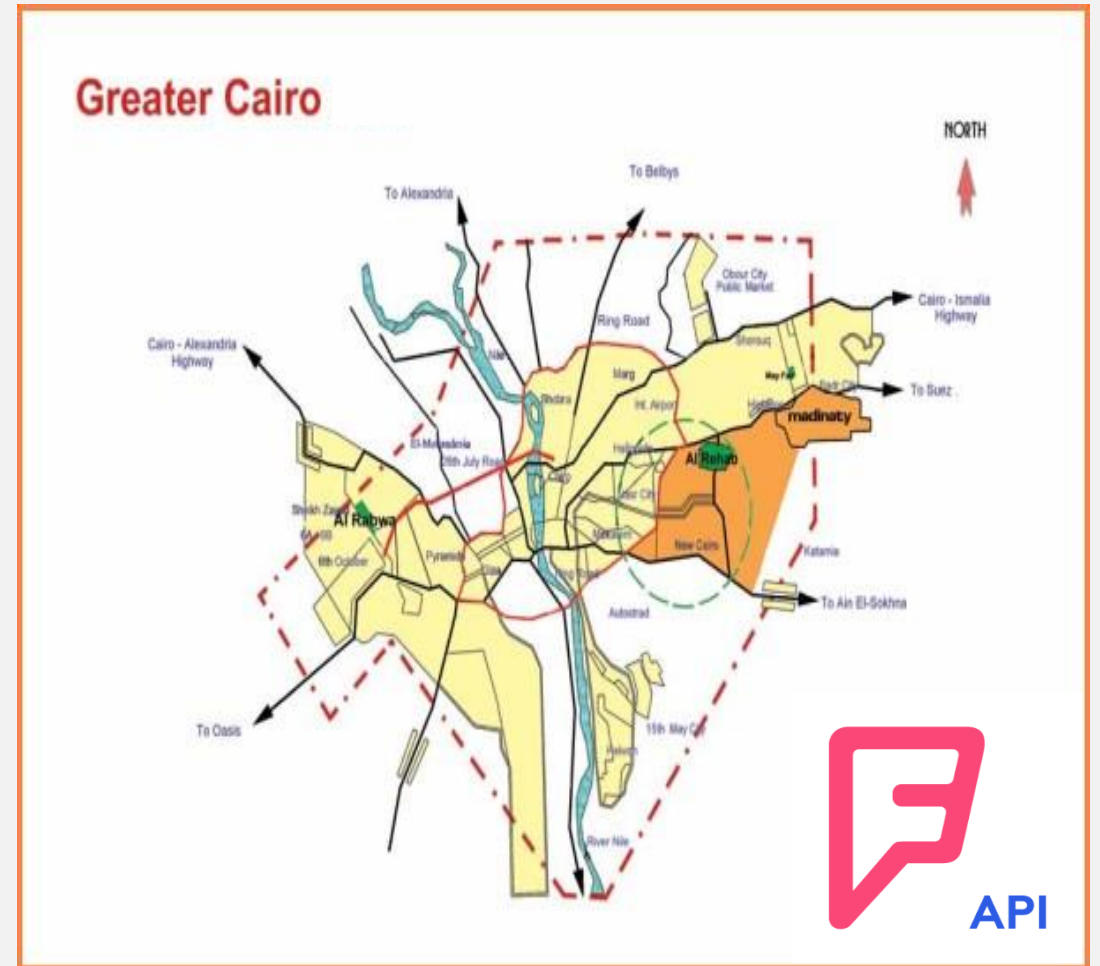
DATA USAGE

The aforementioned gathered data shall be fed into a clustering algorithim (k-means clustering algorithm being the one of choice) and using methods, the appropriate number of K shall be determined and consequently used allowing people looking to hang out whether it's in a cafe or restaurant to see which restaurants will quench their thirst and satisfy their crave
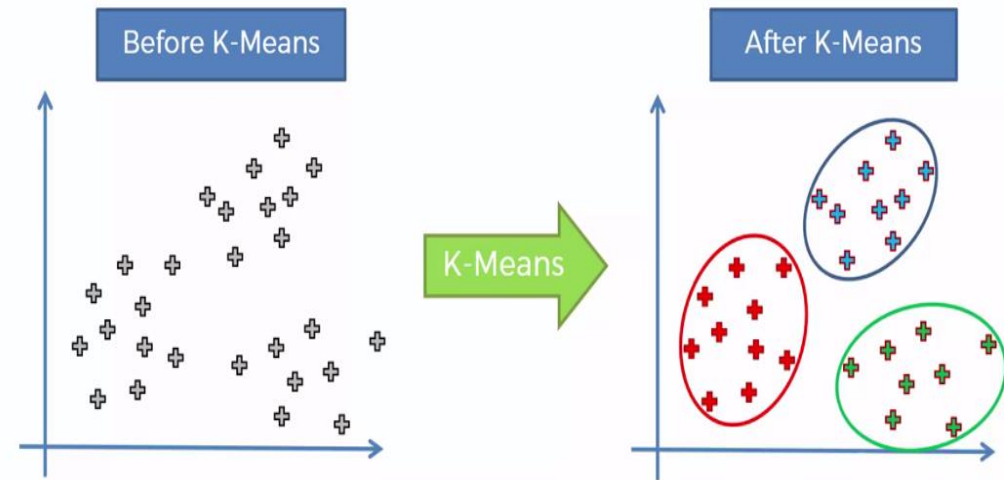
AGENDA

Introduction & Business Problem

Data: Requirements, Acquisition and Usage

Project Methodology

Results

METHODOLOGY

With the widespread of internet and its influence, people have become more aware and take more considerations on where they spend their money especially when they are seeking out quality. Restaurants and cafes are no exception but on the contrary many strive to improve their image throughout social media campaigns and so on in order to be more appealing (likeable) to people. These likes serve as an indicator of overall customer satisfaction, the more the likes/stars there are, the better the place is considered to be. Due to limitations of free API of Foursquare, data will be split into categorical bins based on the quality which allows clustering to be performed appropriately.

This will allow people to know what are types of food might be best to eat in Cairo if they are new to the area or just want to try out a different venue.

METHODOLOGY (CONT.) – OPTIMUM CLUSTER SIZE

The optimum number of clusters is always a challenge in clustering and thus, two methods were used to determine what that number should be. The first being Elbow Criterion Method and the second Silhouette Coefficient

METHODOLOGY (CONT.) – ELBOW CRITERION METHOD

- This method runs the clustering algorithm for a set number of values of K and for each value of K, the sum of squared errors (SSE) is calculated. The goal is to minimize SSE which tends to decrease towards 0 as the K value increases. If K is equal to the number of data points, then SSE will be zero as each point would be its own cluster and thus would have no error.

- Given the information stated above, a small value of K that still has a low SSE should be chosen. The elbow represents where diminishing returns are gained by increasing the value of K.
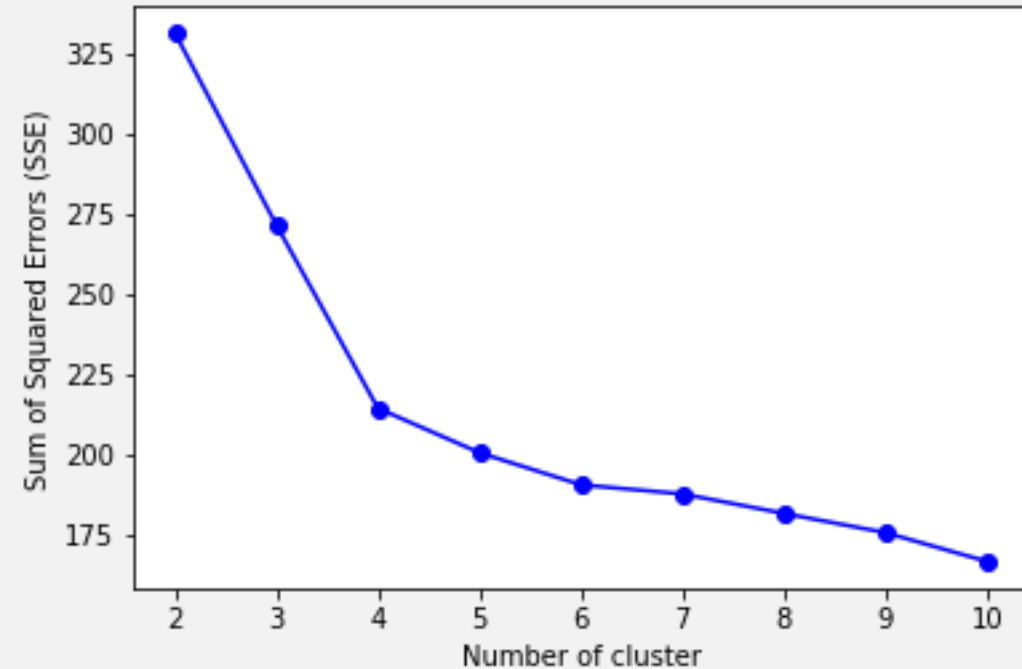
METHODOLOGY (CONT.) – SILHOUETTE COEFFICIENT

- The second method was the Silhouette Coefficient which also runs for a set value of K with the following equation denoting the coefficient:

$$s = \frac{b - a}{\max(a, b)}$$

Where:

- a denotes the mean distance between a sample and all other points in the same class.

- b denotes the mean distance between a sample and all other points in the next nearest cluster.

- The higher the Silhouette Coefficient is, the better the point matched to its own cluster and poorly matched to neighboring clusters.

Value of K	Silhouette Coefficient
2	0.1686757701533035
3	0.24787600957585568
4	0.326478132411156176
5	0.2928826800706246
6	0.2785346234758919
7	0.29641621861618744
8	0.2612483495589942
9	0.28945745396533984
10	0.31624532353747037

METHODOLOGY (CONT.)

Given the previous outputs, the gathered data was clustered into 4 (determined by using the Silhouette Coefficient) clusters using the k-means algorithm which facilitates find what cuisine is available and where in (Great) Cairo they can look to eat.

AGENDA

Introduction & Business Problem

Data: Requirements, Acquisition and Usage

Project Methodology

Results

RESULTS

Running the k-means clustering algorithm with K=4 and populating the unique venue categories in each cluster, it can be observed that Cafes seem to be the most abundant venue in all clusters as many Cafes now opt to additionally serve main dishes of different variety other than sandwiches, drinks, juices and beverages

	Venue Category	Count		Venue Category	Count		Venue Category	Count		Venue Category	Count
2	Café	14		Café	8		Café	12		Café	18
17	Middle Eastern Restaurant	4		Coffee Shop	8		Pizza Place	4		Restaurant	7
8	Falafel Restaurant	4		Lebanese Restaurant	5		Egyptian Restaurant	3		Coffee Shop	5
12	Ice Cream Shop	2		Restaurant	4		Dessert Shop	3		Egyptian Restaurant	4
11	Hookah Bar	2		Cupcake Shop	3		Fast Food Restaurant	3		Fast Food Restaurant	4
22	Syrian Restaurant	2		Middle Eastern Restaurant	3		Kebab Restaurant	3		Middle Eastern Restaurant	4
19	Restaurant	2		Fast Food Restaurant	3		Coffee Shop	2		Pizza Place	3
15	Kebab Restaurant	2		Burger Joint	2		Syrian Restaurant	2		Kebab Restaurant	3
14	Juice Bar	2		Syrian Restaurant	2		Food Truck	2		Asian Restaurant	2
24	Yemeni Restaurant	2		Sushi Restaurant	2		Thai Restaurant	2		Indian Restaurant	1
9	Fast Food Restaurant	2		Egyptian Restaurant	2		Pie Shop	2		Swiss Restaurant	1
5	Dessert Shop	2		Pub	1		Indian Restaurant	2		Steakhouse	1
10	Food Stand	1		Peruvian Restaurant	1		Middle Eastern Restaurant	2		Juice Bar	1
1	Burger Joint	1		American Restaurant	1		Asian Restaurant	2		Hot Dog Joint	1
13	Japanese Restaurant	1		Sandwich Place	1		Shawarma Place	1		Ice Cream Shop	1
7	Egyptian Restaurant	1		Juice Bar	1		Salad Place	1		Bakery	1
6	Dim Sum Restaurant	1		Seafood Restaurant	1		Korean Restaurant	1		Hookah Bar	1
16	Lebanese Restaurant	1		French Restaurant	1		Restaurant	1		Falafel Restaurant	1
18	Pizza Place	1		Indian Restaurant	1		Modern European Restaurant	1		Diner	1
4	Cupcake Shop	1		Fried Chicken Joint	1		Hookah Bar	1		Cupcake Shop	1
20	Sandwich Place	1		Bakery	1		Juice Bar	1		Cafeteria	1
21	Snack Place	1		Food Stand	1		Bakery	1		Yemeni Restaurant	1

1ST CLUSTER (CLUSTER 0)

	District	District Latitude	District Longitude	Venue	Venue ID	Venue Category	Venue Latitude	Venue Longitude	Like Count	Rating	Cluster
23	Azbakeya	30.053889	31.245473	El Tabei El Domyati (التابعى الدمياطى)	4e856185cc2143d95495b059	Falafel Restaurant	30.056195	31.242615	9	Average	0
26	Azbakeya	30.053889	31.245473	La Poire Cafe	4fd9b0f2e4b06a851a06d49a	Dessert Shop	30.050434	31.242612	8	Average	0
39	Boulaq	30.063093	31.228926	Solana Restaurant	50580550e4b0b66b9511901f	Restaurant	30.065486	31.227720	8	Average	0
53	Daher	30.046661	31.323637	Cappuccino Cafe	4e39c21ad164102b0375fa23	Café	30.047037	31.324467	10	Average	0
56	Daher	30.046661	31.323637	Crystal Bakery (مخبز كريستال)	4f3e386ee4b0fbbbd74f33c1	Bakery	30.048606	31.327970	13	Average	0
59	Daher	30.046661	31.323637	Moka Coffee	55a857b8498ed9bf2ac04a74	Café	30.044007	31.325126	11	Average	0
61	Daher	30.046661	31.323637	El Sokarreya (السكرية)	4fcbacabe4b0c4b6a6cb90b0	Café	30.045573	31.324349	11	Average	0
68	El-Manial	30.027673	31.229000	Abo Hammam (ابو همام)	5190d61c498e61f72a5494f3	Juice Bar	30.024338	31.226007	8	Average	0
69	El-Manial	30.027673	31.229000	Koshary El Moemen (كشري المؤمن)	533bfac8498e176b286ea744	Egyptian Restaurant	30.025824	31.226286	8	Average	0
70	El-Manial	30.027673	31.229000	Saydala cafe (صيدلة كافيه)	52030780498effdba9b2e3c1	Café	30.027999	31.232707	7	Average	0
71	El-Manial	30.027673	31.229000	Hekaya	52ab4ca8498ee4de0c7e6d67	Coffee Shop	30.025359	31.226152	8	Average	0
82	Nasr City	30.052118	31.342205	Salah El Din Roastery (مظلة صلاح الدين)	4e7ceef729c2e8fcd9303c33	Snack Place	30.053479	31.338462	12	Average	0
86	Nasr City	30.052118	31.342205	Primo's Pizza (بريموس بيتزا)	5654b56d498efc7f706e085b	Pizza Place	30.054270	31.341996	10	Average	0
93	Nasr City	30.052118	31.342205	Storia	4e7fa0370aafbd4bf61fc80b	Café	30.055091	31.338568	11	Average	0
104	Nasr City	30.052118	31.342205	My Queen (ماي كوين)	4feb3f17e4b04246c6d12d7d	Fast Food Restaurant	30.051887	31.337060	12	Average	0
115	Rod El Farag	30.080570	31.245426	El Sahafa Cafe (قهوة الصحافة)	50a171bee4b0c6336f83699f	Café	30.080796	31.245187	9	Average	0
116	Rod El Farag	30.080570	31.245426	McDonald's	50788ab4e4b0b5e91f512674	Fast Food Restaurant	30.076986	31.245139	10	Average	0
141	Fustat	30.017961	31.237684	Kababgy Abo Ashraf (كبابجي ابو اشرف)	506f336de4b0f4cc1d4e8648	Kebab Restaurant	30.018590	31.234772	13	Average	0
142	Fustat	30.017961	31.237684	Soirée	4fae4891e4b01c0fe9840052	Restaurant	30.015229	31.241062	11	Average	0
161	Islamic Cairo	30.048747	31.263775	El Malky (المالكي)	4e3c61c545dd68e3272b00a0	Dessert Shop	30.047256	31.262646	11	Average	0
162	Islamic Cairo	30.048747	31.263775	Egyptian Pancakes	4bb20cbcf964a520a5b43ce3	Middle Eastern Restaurant	30.046655	31.261886	7	Average	0

2ND CLUSTER (CLUSTER 1)

	District	District Latitude	District Longitude	Venue	Venue ID	Venue Category	Venue Latitude	Venue Longitude	Like Count	Rating	Cluster
12	Azbakeya	30.053889	31.245473	Al Americaine (الأمريكين)	4eb2c4a7cc21adeaf89acb9a	Café	30.052561	31.244304	35	Excellent	1
42	Boulaq	30.063093	31.228926	Spectra	500857c6e4b0dbe2624e5406	Restaurant	30.065654	31.226778	156	Excellent	1
54	Daher	30.046661	31.323637	Nile Thailand Restaurant	4d23318e0901721e0726a9a5	Thai Restaurant	30.046354	31.327713	41	Excellent	1
79	Nasr City	30.052118	31.342205	Al-Awda (العودة)	4ee397d70e011d524554e8a4	Falafel Restaurant	30.053379	31.342359	58	Excellent	1
84	Nasr City	30.052118	31.342205	Awlad TaTa (اولاد طأطأ)	5075b693e4b0468f51822d71	Seafood Restaurant	30.054388	31.343465	39	Excellent	1
147	Islamic Cairo	30.048747	31.263775	Naguib Mahfouz Cafe (مقهى نجيب محفوظ)	4ee2696393ad88a6bed30288	Café	30.047963	31.261567	409	Excellent	1
148	Islamic Cairo	30.048747	31.263775	El Fishawy (قهوة الفيشاوى)	4b7bc840f964a520fe6d2fe3	Coffee Shop	30.047373	31.262556	421	Excellent	1
157	Islamic Cairo	30.048747	31.263775	Farahat (فرحات)	4ed7e33a0e0189600bb9f6a0	Egyptian Restaurant	30.046465	31.261270	48	Excellent	1
159	Islamic Cairo	30.048747	31.263775	Zainab Khatoon (زينب خاتون)	4e481a83483b03d090f653a8	Café	30.044588	31.263396	497	Excellent	1
183	Garden City	30.036417	31.231725	Zitouni	4cea21e20f196dcb8b1150ae	Lebanese Restaurant	30.036200	31.229422	35	Excellent	1
196	Garden City	30.036417	31.231725	Taboula (مطعم تبولة)	4c2f7d313896e21e6103e490	Lebanese Restaurant	30.039642	31.232854	145	Excellent	1
210	Heliopolis	30.100643	31.332511	IL Pennello Ceramic Cafe	4cf2a6f28333224be671118e	Café	30.100261	31.333240	44	Excellent	1
212	Heliopolis	30.100643	31.332511	Man'oucheh	4c1532acb7b9c9281a8eac37	Lebanese Restaurant	30.100304	31.334704	52	Excellent	1
226	Maadi	29.959840	31.258074	NOLA Cupcakes (نولا)	4f7b3f63e4b011878813148e	Cupcake Shop	29.958510	31.259526	124	Excellent	1
230	Maadi	29.959840	31.258074	Grapes Restaurant & Lounge	55f19d71498e570d3d052819	Restaurant	29.957948	31.259746	60	Excellent	1
231	Maadi	29.959840	31.258074	Maharani (ماهاراني)	5420740e498e871a2c13eecb	Indian Restaurant	29.958269	31.257651	36	Excellent	1
232	Maadi	29.959840	31.258074	Beano's Cafe (بينوس كافيه)	4e982052f79083717a963311	Café	29.958543	31.260475	185	Excellent	1
233	Maadi	29.959840	31.258074	Abou El Sid (ابو السيد)	4bf57ef794b2a5932bacabee	Middle Eastern Restaurant	29.957307	31.259536	56	Excellent	1
234	Maadi	29.959840	31.258074	Zööba (زووبا)	52e39e45498eff0338323322	Middle Eastern Restaurant	29.957894	31.259910	96	Excellent	1
236	Maadi	29.959840	31.258074	Cuba Cabana (كوبا كابانا)	4bba61f11261d13a3291ea98	Cuban Restaurant	29.956101	31.260071	96	Excellent	1

3RD CLUSTER (CLUSTER 2)

	District	District Latitude	District Longitude	Venue	Venue ID	Venue Category	Venue Latitude	Venue Longitude	Like Count	Rating	Cluster
6	Azbakeya	30.053889	31.245473	Latif Wassily (لطيف وسيلي)	4f7e9104e4b0e160b786d8dd	Bakery	30.054508	31.248270	28	Good	2
19	Azbakeya	30.053889	31.245473	Shalaby Hawawshi (حواوشي شلبي)	506486d8e4b0593606e72415	Egyptian Restaurant	30.053207	31.242935	24	Good	2
20	Azbakeya	30.053889	31.245473	McDonald's	4db0b701fa8ca4b3e9e83b90	Fast Food Restaurant	30.051956	31.241425	27	Good	2
25	Azbakeya	30.053889	31.245473	Akher Sa'a (أخر ساعة)	5047b0bfe4b0ea030c588261	Falafel Restaurant	30.053594	31.243780	16	Good	2
28	Azbakeya	30.053889	31.245473	Al Americaine (الأمريكين)	4db1e0725da32cf2df66c763	Café	30.052414	31.241670	22	Good	2
52	Daher	30.046661	31.323637	El Faraana Cafe (قهوة الفراعنة)	5159e314e4b086af7826a7d0	Café	30.047045	31.323439	15	Good	2
55	Daher	30.046661	31.323637	Hardee's	4cdeed4470bba1cd508063c4	Fast Food Restaurant	30.047803	31.328264	20	Good	2
58	Daher	30.046661	31.323637	Hawaii	4fa1a977e4b08da3595d0fc5	Hookah Bar	30.048023	31.322391	14	Good	2
60	Daher	30.046661	31.323637	Si Omar (سي عمر)	4e2f2d4ffa7614797de0c627	Café	30.044256	31.325532	34	Good	2
64	Daher	30.046661	31.323637	Shagarteen Cafe	4d2ba87c77a2a1cddff87eb7	Café	30.043662	31.326005	14	Good	2
80	Nasr City	30.052118	31.342205	Pizza Di Napoli	544534eb498e9b3efe220fb1	Pizza Place	30.054311	31.343071	16	Good	2
81	Nasr City	30.052118	31.342205	Peking	4da88ae0a86e68edb7dfc669	Asian Restaurant	30.055127	31.342944	19	Good	2
83	Nasr City	30.052118	31.342205	El Sultan (السلطان)	5276c07511d2b7efb2966b40	Pie Shop	30.050839	31.340002	27	Good	2
92	Nasr City	30.052118	31.342205	El Genany (الجناني)	51f8509d498e1d9eff5639b8	Syrian Restaurant	30.051426	31.347061	26	Good	2
97	Nasr City	30.052118	31.342205	Abo Mazen (أبو مازن)	4ebed46d8b81127c0abc1233	Shawarma Place	30.053991	31.340679	28	Good	2
154	Islamic Cairo	30.048747	31.263775	Al Hussein Cafes (مقاهي الحسين)	4f5120c8e4b00ad2ab1a4636	Café	30.047270	31.262593	26	Good	2
164	Islamic Cairo	30.048747	31.263775	El Hussein Pies (فطاطري الحسين)	5071953ee4b054850479ebd4	Pie Shop	30.046870	31.261858	29	Good	2
168	Islamic Cairo	30.048747	31.263775	Al Ahd Al Gaded (العهد الجديد)	4fa6d918e4b0a34ac8a07a4f	Egyptian Restaurant	30.047210	31.261588	15	Good	2
191	Garden City	30.036417	31.231725	Osmanly Restaurant	4c9cd87e9975a143fafb35cf	Turkish Restaurant	30.038657	31.230478	29	Good	2
200	Garden City	30.036417	31.231725	Nubian Village	50805c14e4b03db174e9a704	Middle Eastern Restaurant	30.035068	31.227350	16	Good	2
201	Garden City	30.036417	31.231725	Beymen Café	4cd5758994848cfaff63e6b1	Café	30.036058	31.229736	23	Good	2

4TH CLUSTER (CLUSTER 3)

	District	District Latitude	District Longitude	Venue	Venue ID	Venue Category	Venue Latitude	Venue Longitude	Like Count	Rating	Cluster
0	Ain Shams	30.130476	31.316749	El Wally Cafe	50d766bae4b023bf8cc4191c	Café	30.130768	31.317696	2	Poor	3
3	Ain Shams	30.130476	31.316749	Mo'men (مؤمن)	4f98614fe4b0fc88e6fea24d	Fast Food Restaurant	30.127758	31.320648	1	Poor	3
22	Azbakeya	30.053889	31.245473	Koshary Sayed Hanafy (كشرى سيد حنفى)	53400cf9498e9792e60acab6	Egyptian Restaurant	30.053527	31.242905	5	Poor	3
29	Azbakeya	30.053889	31.245473	La Chesa	4bec4468976ac928e6c2600b	Swiss Restaurant	30.051197	31.242703	4	Poor	3
30	Azbakeya	30.053889	31.245473	Alfi Bek (الفى بيك)	4e7cc021dab4ed68da56ab60	Middle Eastern Restaurant	30.053536	31.243486	4	Poor	3
31	Azbakeya	30.053889	31.245473	Om Kalthom Cafe	56b2830a498ee7829498f442	Café	30.055456	31.242710	4	Poor	3
32	Azbakeya	30.053889	31.245473	Layalina (ليالينا)	4f7e4bf9e4b00dcb6975b023	Coffee Shop	30.053238	31.242396	5	Poor	3
41	Boulaq	30.063093	31.228926	Kamala Asian Bar & Dining	4ec54b8c0aaf9233edc5f9f2	Asian Restaurant	30.065569	31.227818	5	Poor	3
44	Boulaq	30.063093	31.228926	Pinci Coffe	505cc8e1e4b0707f48b41cb7	Café	30.065985	31.226835	5	Poor	3
46	Boulaq	30.063093	31.228926	The Lounge Restaurant	50ca02c1245f2d4aa8c2b6ad	Restaurant	30.062000	31.227800	0	Poor	3

MAP OF CLUSTERS FOR USERS