Data Science Specialization
IBM via Coursera

Feddah Alotaibi
fdalthaqeel@gmail.com

July 2019

1. Introduction:

Trying to find the best restaurant to have dinner at is difficult sometimes. Especially for people who are new to the city. The aim of this project is to leverage Foursquare location data in order to help tourists find a good place to eat. This report demonstrates the process of pulling data out of Foursquare and use it group restaurants in a Jeddah, one of the main cities in Saudi Arabia.

2. Business Problem:

Is it possible to group restaurants in Jeddah based on their similarities, can we tell which is best/average/worst? Can we help new visitors and people of Jeddah choose where to eat on a Friday night? This project aims to answer these questions by utilizing data extracted from Foursquare API. The main goal of this project is to explore restaurants and cafes in Jeddah city and cluster them based on similarities and define each of their characteristics.

3. Data Required:

In order to accomplish this project data has to be extracted from Foursquare API.
Data about Jeddah city include (venue name, venue ID, venue location, venue category and number of likes)

4. Methodology:

First of all, I used python jupyter notebook to extract data from foursquare.
The data required is the geolocator latitude and longitude for Jeddah City as well as venues and their information such as name, id, location, category, and. # of likes.
So, to get that, an application using https://foursquare.com/developers/ was created to access data.
K-means clustering algorithm used to group places (restaurants and cafes) to 4 and 5 clusters. To help people looking for a place to get food easily. What kind of restaurant and what did people say about it using number of likes. Since the number of likes says a lot about how good the place is.
At first this data was converted from json to dataframe and do some preparation like fixing columns names. The I created a remove list that includes anything that is not a restaurant or café so that it can be removed from the data. Such as: parks, hotels and bookstores. After that, each venue ID and number of likes were pulled. Then I created new categorical variables for the restaurants to better group them based on type of food.

4.1 Data preparation:

To prepare data for clustering, data from different lists was combined as well as creating new categorical data and encoding data. Descriptive statistics such as mean, max, min and median of total likes gave a clear picture of data. To understand the data more clearly, a histogram was plotted as well. Then each venue classified in one of these categories (poor, below avg, avg avg and great).

- Histogram

```
import matplotlib.pyplot as plt
jeddah_venues['total likes'].hist(bins=4)
plt.show()
```
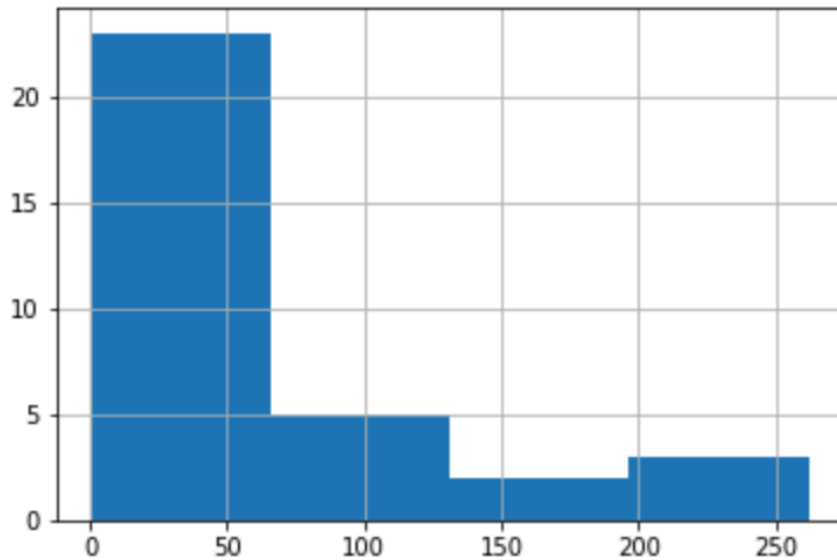

One hot encoding used so that each rating likes and restaurant category gets 0 or 1 as value.
Creating dummy variables for likes and categories.

	Name	Café	american food	euro asia indian food	italian food	mex southam food	middle eastern food	avg avg	below avg	great	poor
1	Hummus Refi (حمص ريفي)	0	0	0	0	0	1	0	0	1	0
2	McDonald's (ماكدونالدز)	0	1	0	0	0	0	0	0	1	0

4.2 Clustering:

At first, K-means clustering used to create 4 clusters and label each place a number from 0-3.

Cluster 1:

Most of the places are Middle Eastern and American restaurants and their rate is average average.

```
jeddah_venues.loc[jeddah_venues['label']==0]
```

id	categories	lat	lng	total likes	total likes_cat	categories_new	label
1498e8000de659d45	Sushi Restaurant	21.581888	39.159764	35	avg avg	euro asia indian food	0
2e498ed111af319adc	Sandwich Place	21.585298	39.161312	20	avg avg	middle eastern food	0
9511d2f4f58c2e052d	Breakfast Spot	21.582303	39.159613	92	avg avg	middle eastern food	0
23e4b0ee26f2c2f339	BBQ Joint	21.582637	39.159683	18	avg avg	american food	0
7d498e7b1ebfae3260	Seafood Restaurant	21.584389	39.159995	28	avg avg	italian food	0
14be7bbcd4f4807a28	BBQ Joint	21.581687	39.159874	89	avg avg	american food	0
0b498e66916685e285	Middle Eastern Restaurant	21.585315	39.161336	19	avg avg	middle eastern food	0
56816af04d31cb3ec2	Fast Food Restaurant	21.580971	39.163867	21	avg avg	american food	0

Cluster 2:

Most of the places are Middle Eastern restaurants their rate is below average and poor.

```
jeddah_venues.loc[jeddah_venues['label']==1]
```

id	categories	lat	lng	total likes	total likes_cat	categories_new	label
a6ae5faf2869d4c7365	Middle Eastern Restaurant	21.581506	39.159932	14	below avg	middle eastern food	1
4a9a528bc5d8c632ce	Chinese Restaurant	21.580639	39.161305	9	below avg	euro asia indian food	1
9e4b0cbbce0cb2eed	Middle Eastern Restaurant	21.582178	39.166938	17	below avg	middle eastern food	1
82e4b0665ce9b44de2	Pizza Place	21.580888	39.160126	15	below avg	italian food	1
396498e6f2bedc79bff	Café	21.582139	39.167402	12	below avg	Café	1
35498e72388cd9e483	Middle Eastern Restaurant	21.581261	39.159999	4	poor	middle eastern food	1
c4498e172dc9469ce3	Egyptian Restaurant	21.582072	39.167643	4	poor	middle eastern food	1
2a38fa25270e18127c	Sandwich Place	21.582642	39.159527	4	poor	middle eastern food	1

Cluster 3:

Most of the places are café and Italian restaurants their rate is below average and poor

```
jeddah_venues.loc[jeddah_venues['label']==2]
```

id	categories	lat	lng	total likes	total likes_cat	categories_new	label
dedab46521c0a59c9a	Pizza Place	21.581727	39.159928	5	poor	italian food	2
:38920b713d97ea0dc	Asian Restaurant	21.580444	39.161462	6	poor	mex southam food	2
2de4b04b0c8d2046f9	Café	21.582180	39.167111	8	poor	Café	2
2a361aff5a344082e01	Pizza Place	21.581831	39.166933	4	poor	italian food	2
d25d891b0039de0c76	Café	21.582006	39.167202	6	poor	Café	2
e8e4b029502ba5107c	Fried Chicken Joint	21.580089	39.160578	1	poor	american food	2

Cluster 4:

Let's call this cluster the golden one, since it has great food and mostly Middle Eastern restaurants.

```
jeddah_venues.loc[jeddah_venues['label']==3]
```

| id	categories	lat	lng	total likes	total likes_cat	categories_new	label
f184cf820a13246edf4	Middle Eastern Restaurant	21.580490	39.162727	94	great	middle eastern food	3
7a18a8ce02fbbdb814	Fast Food Restaurant	21.580938	39.163797	262	great	american food	3
9b49e3562d7f6346fff8	Thai Restaurant	21.580301	39.161358	241	great	euro asia indian food	3
b3e4b048e6f6b578b7	Café	21.580188	39.161137	248	great	Café	3
8548b810332abf3ff98	Middle Eastern Restaurant	21.581661	39.167761	176	great	middle eastern food	3
954d7b0b1f7bc32179f	Middle Eastern Restaurant	21.582641	39.159502	110	great	middle eastern food	3
6b52d8468cfaa2f5ff6b	Fast Food Restaurant	21.580925	39.164004	100	great	american food	3
5752c5b5cb4f38f1f82	Middle Eastern Restaurant	21.581007	39.160128	194	great	middle eastern food	3

Cluster 5:

This cluster contains American restaurants that aren't good nor bad but below average

```
jeddah_venues.loc[jeddah_venues['label']==4]
```

id	categories	lat	lng	total likes	total likes_cat	categories_new	label
97fc73d40beda172e7	Burger Joint	21.581862	39.159764	10	below avg	american food	4
2498e01d9edb902d9	Burger Joint	21.585315	39.161432	10	below avg	american food	4
2e887719193960ef3a	Fried Chicken Joint	21.581639	39.159839	10	below avg	american food	4

The 5 clusters created using folium map are shown below:
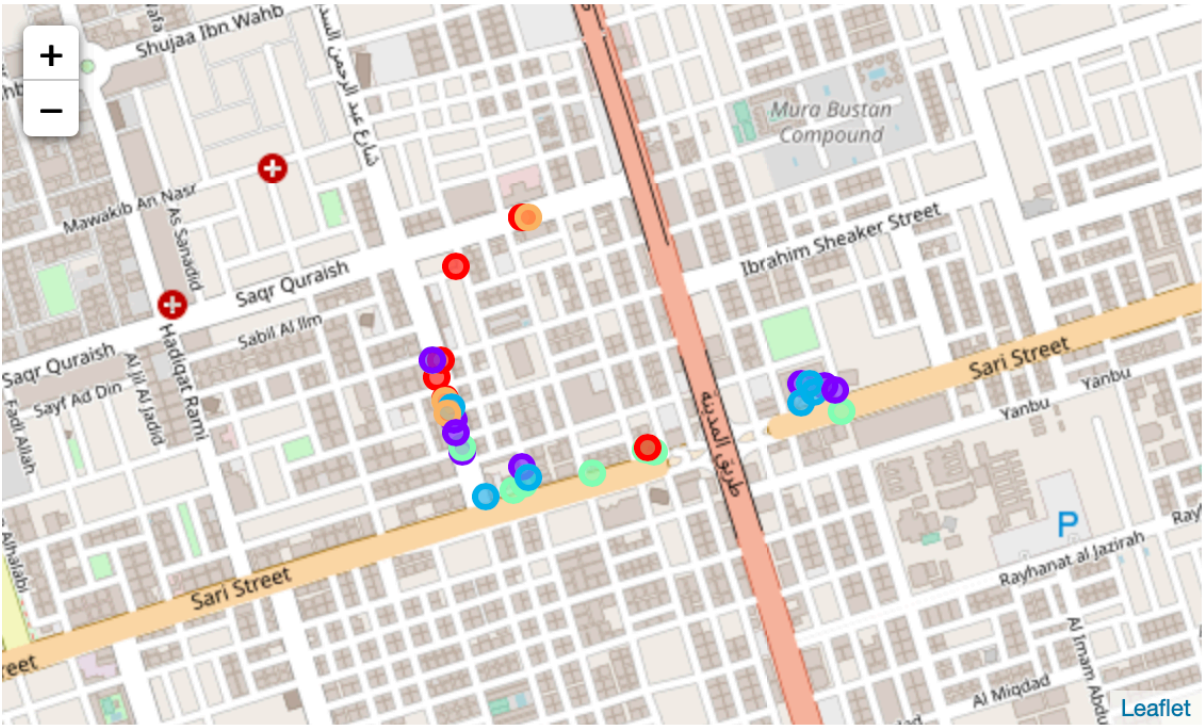
5. Result:

The result shows that the restaurants in the city of Jeddah were clustered into five groups each of which has its own characteristics.
In group one, it was clearly that most of the categories belongs to Middle Eastern and American food and people think that these are not the best ones but "Average average".
The story in group two is a little bit similar how ever there was no American food but most of them were Middle Eastern restaurants and based on people who experienced those restaurants. It is very clear that they are below average restaurants.
The worst restaurants in group three belongs to cafes and Italian foods.
Whereas, group four contains the best restaurants ever, mainly serve American and Middle Eastern food!
The last group have the restaurants that were classified as below average.

6. Discussion:

Based on the clusters resulted from this project. If a friend asked me for a place to eat dinner at, I would very much recommend AL-ZAWAQA or ElBiet ElHalbi if they in the mood for Middle Eastern food. However, if they're looking for a café, Barnies would be a great place. In addition, I will make sure that they never have Italian food here in Jeddah because it tastes so bad. For American food people of Jeddah agree that McDonalds is better than KFC.

7. Conclusion:

To summary, in this project data regarding places in Jeddah were extracted. A clustering algorithm was built to show five different groups that classify Jeddah's restaurants based on the number of likes and feedback from Foursquare.
Doing this project was a great experience to try and do my best. Putting out what we have learned throughout the course into a real-world problem was a great journey. Many thanks to Coursera and IBM for this course.