THE BATTLE OF NEIGHBORHOODS

Submitted By

Sharvari Upasani

November 24th, 2019

PREDICTING LOCATION TO OPEN RESTAURANT IN DUBAI

Table of Contents

1. Introduction

1.1 Background

Dubai is a city based in the United Arab Emirates. It is a cosmopolitan city with multiple cultures. It is a place where there are many expatriates. About 70% of Dubai's population consists of expatriates. Hence, there are many types of food/cuisines liked by people. Cuisines like Filipina, Italian, Indian, Pakistani etc., are some of the most common cuisines.

As, there are plenty of people and many different food types, the question arises, is there a need to open a restaurant in Dubai? If yes, where should it be located so that it shall be visited/utilized more and what different types of food should we sell at this restaurant?

There are plenty of people travelling by metro. Hence, the restaurant location can be decided considering the utility as well as convenience of people. If metro is Dubai's lifeline, let's consider a restaurant to be opened close to metro station. But before that, we need to analyze a few factors of doing so.

1.2 Problem

We do have data for metro stations & metro riders. We can also predict the restaurants that are located near metro stations using the Foursquare API. Our whole idea is to predict a place/cuisine that shall be feasible for utilization of our restaurant.

Hence, we need to modify/combine the data and our findings so that we can predict the actual frequency of places where there are eateries and their cuisines.

1.3 Primary Predictions & Why this Data.

Why do we use Metro stations data? There are many Metro users in Dubai. Also, as per Dubai Government's protocol - we cannot eat/drink at metro stations or in metro trains. Thus, it is a good idea to open a restaurant near a Metro Station, which will be easy for passengers to drop-in/take away their favorite food at their convenience.

2. Data

2.1 Dubai Metro Data

The data for Dubai metro stations is used from the xls file provided by https://bayanat.ae.

This data gives an idea of the Metro stations in Dubai. There are two types of lines – Red Line and Green Line metro stations. Dubai Metro Rail is distributed across these two lines. Hence, a list of all the metro stations with a description of Metro Line is shared in the data. Also, there are Latitude & Longitude Values of the Metro stations shared in the file.

Link: http://data.bayanat.ae/en_GB/dataset/metro-stations/resource/666412a8-3383-47df-b521-c70150698e6a

2.2 Dubai Metro Rider's Data

The data for Dubai metro riders is used from the xls file provided by https://bayanat.ae.

This data file consists of trips from a start station to end station on Red & Green Lines respectively. We need to calculate the number of trips from this data to check frequency of a station being used. We can calculate the trips from one station to another, hence to get a value of most used trip line and most visited station. Using the most used station, we will explore neighborhoods and try to build further datasets.

Link: http://data.bayanat.ae/en_GB/dataset/metro-ridership/resource/c9e51547-7f43-4475-9d35-17e0fc5e9eb2

2.3 Foursquare API

We'll collect different latitude longitudes of multiple metro stations. We'll try to compare their neighborhoods using Foursquare API. We'll try to get the data of various restaurants and their types. We'll also look at the distances at which the restaurants are located. Based on this data we'll explore the neighborhoods individually.

Using this data, we can get the place which has more number of restaurants. When it comes to restaurants, competition matters and also the cuisine. So based on those data sets we will try to fetch a viable cuisine or will try to find if Multi-cuisine is feasible or no.

Link: https://foursquare.com

2.4 Data Cleaning

Data received from the websites is in a rough format. That is, we don't need the whole data. We need to remove a few fields play with a few to make it compatible as per our dataset.

Here is the Metro stations Data. We will remove the fields that are not required and form the dataset newly.

Data frame showing Metro Line of Dubai

	station_name	line_name	station_location_longitiude	station_location_latitiude	loading_date	ingestion_id
0	World Trade Centre Metro Station	Red Metro Line	55.285061	25.224829	20171119	4431
1	Emirates Towers Metro Station	Red Metro Line	55.279821	25.217214	20171119	4431
2	Financial Centre Metro Station	Red Metro Line	55.275587	25.211030	20171119	4431
3	Burj Khalifa/ Dubai Mall Metro Station	Red Metro Line	55.269518	25.201401	20171119	4431
4	Business Bay Metro Station	Red Metro Line	55.260419	25.191275	20171119	4431

We'll apply Data Preprocessing, i.e., clean the data in a way that we can use it. We'll remove the loading date, ingestion id and update line name. We will also update column names so that it becomes more meaningful.

Scrapped Data frame showing Metro Line of Dubai after Preprocessing

	Station	Line	Longitude	Latitude
0	World Trade Centre Metro Station	Red	55.285061	25.224829
1	Emirates Towers Metro Station	Red	55.279821	25.217214
2	Financial Centre Metro Station	Red	55.275587	25.211030
3	Burj Khalifa/ Dubai Mall Metro Station	Red	55.269518	25.201401
4	Business Bay Metro Station	Red	55.260419	25.191275

Here is our data for stations, line with latitude longitude values.

Now, we will try to analyze the data which gives us rider's info. Similarly, we'll try to scrap the data in a way that it becomes acceptable. For a particular date, we have the data of start location, end location, time of the metro rides as below.

Data frame showing Metro Riders List of Dubai

	txn_date	txn_time	start_location	end_location
0	2017-11-04	15:48:19	Burj Khalifa/ Dubai Mall Metro Station	Burj Khalifa/ Dubai Mall Metro Station
1	2017-11-04	13:55:22	Burj Khalifa/ Dubai Mall Metro Station	Burj Khalifa/ Dubai Mall Metro Station
2	2017-11-04	17:22:58	Burj Khalifa/ Dubai Mall Metro Station	Burj Khalifa/ Dubai Mall Metro Station
3	2017-11-04	17:50:51	Al Fahidi Metro Station	Al Fahidi Metro Station
4	2017-11-04	19:32:17	Burj Khalifa/ Dubai Mall Metro Station	Burj Khalifa/ Dubai Mall Metro Station

Here, we have grouped the data as per end station. We calculated the number of trips to the stations. This will give us an idea of which station is used more.

Scrapped Data frame showing Metro Trips after Preprocessing

	Date	Station	Trips
0	2017-11-04	Burj Khalifa/ Dubai Mall Metro Station	38862
1	2017-11-04	Burj Khalifa/ Dubai Mall Metro Station	38862
2	2017-11-04	Burj Khalifa/ Dubai Mall Metro Station	38862
3	2017-11-04	Al Fahidi Metro Station	43813
4	2017-11-04	Burj Khalifa/ Dubai Mall Metro Station	38862
5	2017-11-04	Burj Khalifa/ Dubai Mall Metro Station	38862
6	2017-11-04	Burj Khalifa/ Dubai Mall Metro Station	38862
7	2017-11-04	Burj Khalifa/ Dubai Mall Metro Station	38862
8	2017-11-04	Burj Khalifa/ Dubai Mall Metro Station	38862

We now need to get the latitude longitude values mapped with trips. Hence, we shall merge both the Datasets and form a new dataset.

Scrapped Data frame showing Metro Trips after Merging

	Station	Line	Longitude	Latitude	Date	Trips
0	World Trade Centre Metro Station	Red	55.285061	25.224829	2017-11-04	11982
1	Emirates Towers Metro Station	Red	55.279821	25.217214	2017-11-04	13963
2	Financial Centre Metro Station	Red	55.275587	25.211030	2017-11-04	14687
3	Burj Khalifa/ Dubai Mall Metro Station	Red	55.269518	25.201401	2017-11-04	38862
4	Business Bay Metro Station	Red	55.260419	25.191275	2017-11-04	27444
5	Noor Bank Metro Station	Red	55.228509	25.155727	2017-11-04	27655
6	FGB Metro Station	Red	55.207898	25.126721	2017-11-04	18579
7	Mall of the Emirates Metro Station	Red	55.200443	25.121231	2017-11-04	39320
8	Sharaf DG Metro Station	Red	55.190931	25.114809	2017-11-04	14505

3. Methodology

3.1 Querying the data

We have a dataset of Trips, Stations & Latitude, and Longitude of stations. From this data, we can get which stations on Red & Green lines have less number of trips. Those stations are used very less. Hence, if we want to setup a restaurant near metro stations, those stations can be ignored. However, let's explore the neighborhoods of those two metro stations so as to make sure, whether or not any restaurants are there. If incase there is a corporate hub near those stations, it is possible that they may have restaurants. We can explore the neighborhoods using Foursquare API to clear our presumptions.

Map showing Metro Line of Dubai

Map showing Red Metro Line of Dubai

Map showing Green Metro Line of Dubai

3.2 Cleansing the data

Out of the Latitude longitude points of Dubai Metro, we shall be using the ones which have lowest trips.

We find out the least used Metro Stations on both Green & Red lines and will explore the neighborhoods of both in the analyze section.

Data frame showing least metro trips Dubai

	Station	Line	Longitude	Latitude	Date	Trips
13	Nakheel Harbour & Tower Metro Station	Red	55.127173	25.057853	2017-11-04	1497
37	Creek Metro Station	Green	55.338953	25.218949	2017-11-04	2923

Furthermore, let's also explore the metro station having maximum trips. The intersection of both Green & Red Metro lines is Burjaman. This metro station has max number of trips.

Data frame showing most metro trips Dubai

	Station	Line	Longitude	Latitude	Date	Trips
32	BurJuman Metro Station	Green	55.304253	25.254856	2017-11-04	64269
33	BurJuman Metro Station	Red	55.304253	25.254856	2017-11-04	64269

We'll explore neighborhoods of all the three metro stations in next section.

3.3 Analyzing the data

Let's check the neighborhood of Red line metro station – Nakheel Harbour. This metro station has least number of trips.

Data frame showing Nakheel Harbour Metro station - Neighborhoods

	name	categories	lat	lng	distance
0	McDonald's	Fast Food Restaurant	25.057769	55.127391	23
1	Nakheel Harbour & Tower Metro Station (محطة نخ...	Metro Station	25.057951	55.127259	13
2	Canali Boutique	Men's Store	25.060925	55.128980	387
3	Arz's tent	Trail	25.061336	55.129579	457
4	Last Exit Dubai	Bistro	25.061697	55.129670	496

There are total places out of which hardly 2 are restaurants. Those two restaurants are also located at a greater distance. Even if we open a restaurant here, it would not be much in use. Yes, the level of competition is not more here. But, as we already know number of trips to this station is also low. Also there are no amusement parks, shopping centers etc. nearby so that it will be utilized more.

The next is Dubai Creek station, on Green line which has less number of trips.

Data frame showing Dubai Creek Metro station - Neighborhoods

	name	categories	lat	lng	distance
0	Creek Metro Station (محطة الخور)	Metro Station	25.218915	55.339029	8
1	Jadaf Pier	Harbor / Marina	25.219566	55.337019	206
2	Dubai Maritime city	Harbor / Marina	25.220886	55.337834	243
3	Al Jadaf Marine Transport Station	Boat or Ferry	25.216270	55.341010	363
4	Freez Ice Cream - Al Jaddaf	Ice Cream Shop	25.221217	55.335346	442
5	Premier Inn Al Jadaf	Hotel	25.218785	55.334404	458
6	Goldstate Hotel	Hotel	25.221015	55.334680	487

Again as assumed, there is less number of hotel – 3 out of 7. Also, there is no hotel near the metro station. One thing can be considered that though competition is less, we can have a plan to open a restaurant here as there are boat ferries and marina view. However, considering distance, latitude/longitude we can explore the cuisine and plan to get one here. But, there will be a lot of cultural background research needed so that one can open a proper cuisine restaurant here. The data doesn't show any such values, hence predicting on the basis of data shall be difficult.

However, let's keep the distance of hotels from metro station and a restaurant category at aside and explore the neighborhood with the maximum trips. It is expected to have many restaurants nearby and many different cuisines to give us a choice what and how we can open a restaurant.

Data frame showing Burjuman Metro Station - Neighborhoods

	name	categories	lat	lng	distance
71	UAE Exchange	Currency Exchange	25.254847	55.304011	24
13	Olive Gourmet	Middle Eastern Restaurant	25.254631	55.303768	54
16	Share Tea since 1992	Tea Room	25.254716	55.303699	57
3	Dôme Café	Café	25.254770	55.303663	60
0	Tim Hortons	Coffee Shop	25.254623	55.303647	66
24	Mr Chaat	Indian Restaurant	25.254322	55.304618	69
45	Max	Clothing Store	25.254809	55.304952	70
11	Aryaas Veg Restaurant	Vegetarian / Vegan Restaurant	25.255373	55.303743	77
63	Ravi Darbar	Asian Restaurant	25.254391	55.304898	83
46	Pita Restaurant	Asian Restaurant	25.254432	55.305111	98
10	Carrefour	Supermarket	25.254032	55.303600	112
39	GRK Fresh Greek Burjuman	Greek Restaurant	25.254017	55.303518	119
41	Al Ustad Special Kabab	Persian Restaurant	25.255458	55.303137	130
32	Costa Coffee	Coffee Shop	25.254008	55.303347	131
44	YO! Sushi	Sushi Restaurant	25.253792	55.303534	138
2	California Pizza Kitchen	Pizza Place	25.253810	55.303486	139

Here, we see a lot of restaurants. They are actually with different cuisines, different tastes and located at nearby distances. There shall be a tough competition of hotels to sustain here, but also it will be used more. We might need other data to explore the prices to open hotel.

Hence, as per the data, it is clear that based on given values this place can be considered feasible to open a restaurant.

However, let's calculate the accuracy of the data & model we have been using. As we have Dubai Metro Stations as a primary place to open a restaurant nearby let's consider that are the distance, latitude, longitude and our neighborhood data proper to gain insights and predict a cuisine of restaurant?

4. Results

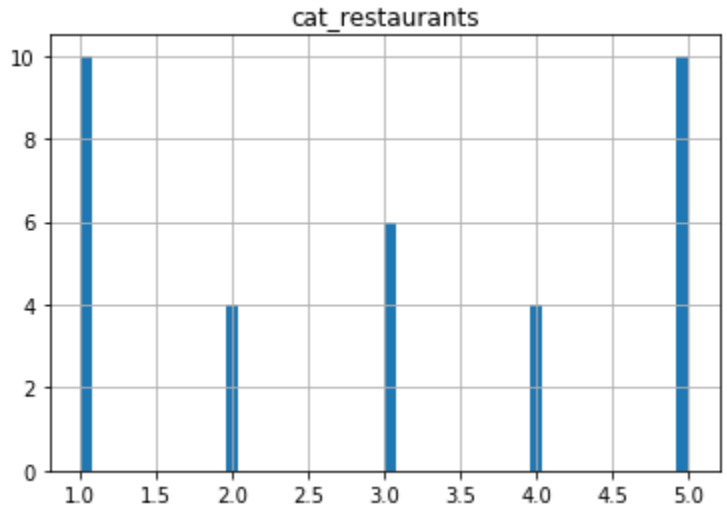
4.1 K-Nearest Neighbor Classification Model

We have a list of Restaurants with the latitude, longitude, distance. We'll give a numeric value to categories.

Data frame showing Burjuman Metro Station – Restaurants

	name	categories	lat	lng	distance	cat_restaurants
13	Olive Gourmet	Middle Eastern Restaurant	25.254631	55.303768	54	3
24	Mr Chaat	Indian Restaurant	25.254322	55.304618	69	5
11	Aryaas Veg Restaurant	Vegetarian / Vegan Restaurant	25.255373	55.303743	77	1
63	Ravi Darbar	Asian Restaurant	25.254391	55.304898	83	5
46	Pita Restaurant	Asian Restaurant	25.254432	55.305111	98	5
39	GRK Fresh Greek Burjuman	Greek Restaurant	25.254017	55.303518	119	1
41	Al Ustad Special Kabab	Persian Restaurant	25.255458	55.303137	130	1
44	YO! Sushi	Sushi Restaurant	25.253792	55.303534	138	1
56	KFC	Fast Food Restaurant	25.253985	55.303098	151	3
53	Nando's	African Restaurant	25.253682	55.303324	160	1
28	McDonald's	Fast Food Restaurant	25.253920	55.302908	170	3
40	Teriyaki Boy	Japanese Restaurant	25.253289	55.303350	196	1
27	Rocky's Cafe	Asian Restaurant	25.252996	55.302884	199	5

Graph showing Burjuman Metro Station Restaurant Categories

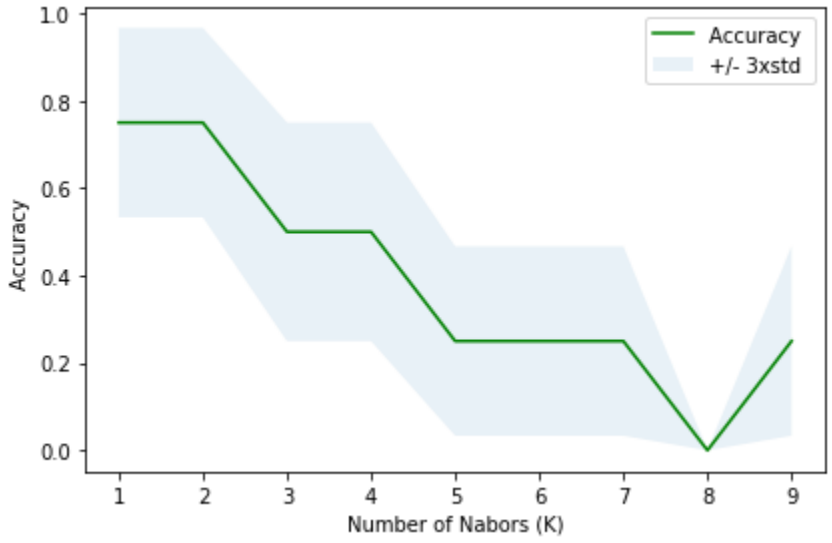


Above is the data of categories of cuisines we have in the area. There are more Vegetarian & Asian Restaurants as compared to Italian & specialty cuisine restaurants.

Data frame showing Restaurants based on Distance

	name	categories	lat	lng	distance	cat_restaurants
13	Olive Gourmet	Middle Eastern Restaurant	25.254631	55.303768	54	3
24	Mr Chaat	Indian Restaurant	25.254322	55.304618	69	5
11	Aryaas Veg Restaurant	Vegetarian / Vegan Restaurant	25.255373	55.303743	77	1
63	Ravi Darbar	Asian Restaurant	25.254391	55.304898	83	5
46	Pita Restaurant	Asian Restaurant	25.254432	55.305111	98	5
39	GRK Fresh Greek Burjuman	Greek Restaurant	25.254017	55.303518	119	1
41	Al Ustad Special Kabab	Persian Restaurant	25.255458	55.303137	130	1
44	YO! Sushi	Sushi Restaurant	25.253792	55.303534	138	1
56	KFC	Fast Food Restaurant	25.253985	55.303098	151	3
53	Nando's	African Restaurant	25.253682	55.303324	160	1
28	McDonald's	Fast Food Restaurant	25.253920	55.302908	170	3
40	Teriyaki Boy	Japanese Restaurant	25.253289	55.303350	196	1
27	Rocky's Cafe	Asian Restaurant	25.253096	55.303884	199	5
1	Jollibee	Fast Food Restaurant	25.253250	55.303333	201	3
18	Nanking	Asian Restaurant	25.252908	55.303574	227	5
14	Chinchaaus Restaurant	Chinese Restaurant	25.252730	55.304966	247	2

Result of KNN Neighborhood Classification Accuracy



As per KNN classification model, we considered restaurants, their distance as X (dependent variables) category_id i.e., type of restaurant as Y (Independent Variable). From the selected model and the dataset of the nighborhood used for test & training the data, we found it to be accurate at k=1.

KNN Classification's Train & Test Set Accuracy of our Model

```
Train set Accuracy:  1.0
Test set Accuracy:  0.75
```

```
The best accuracy was with 0.75 with k= 1
```

Hence, we can consider the Dataset to be accurate to find the location of a restaurant in Dubai.

5. Discussion

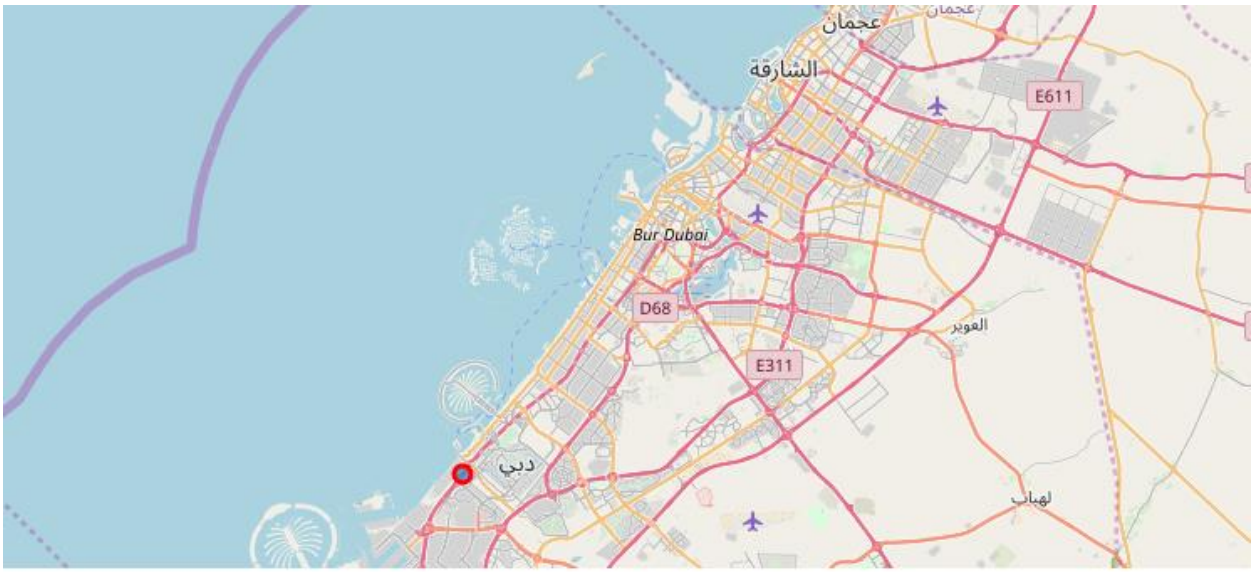
5.1 Observing least used metro stations

While considering Metro stations, we explored the entire neighborhoods of the least used metro stations in Dubai. The city is spread across two major metro lines – Red & Green. We also explored the metro stations - Nakheel Harbour on Red Line & Dubai Creek.

When observed the both metro stations, we noticed many similarities as well as differences in both.

1. Both metro stations are away from the Centre (From the point of intersection of Red & Green Line i.e., Burjuman metro station)
2. Both the metro stations have fewer trips & have less number of nearby places.
3. Dubai Creek Harbour is at one of the end points of the metro track. Hence, it is used more out of the two because it has some Marina viewpoints & amusement parks nearby.
4.

Map showing Nakheel Harbour Metro Station from Bur Dubai



Map showing Dubai Creek Metro Station from Bur Dubai

Thus, out of the two least used metro stations, we can consider Dubai Creek Harbour as good option to open restaurant. However, we would also require the data of prices in the region. When we consider investing into a food business, the utility and overall expense & capital should be considered well in advance. Hence, it is good to have some information on the same.

5.2 Observing most used metro station

While we consider utility & competition, the metro station having most of trips of Dubai is Burjuman Metro Station. When we explored the metro map, we also noticed that this metro station is an interchange station between the green & red line routes. It is very obvious that this metro station will have many users commuting to and fro.

As it has more number of metro users, a future prediction of having restaurant here can be considered. We can also notice that the Vegetarian and Asian cuisines are located at many places. We can consider a different or a multi-cuisine in this area. As there are many restaurants, the rent/costing of hotel should be less. Again, this is a presumption, and we definitely need some data related to this to understand the specs more and find an exact location for our as well as customers' benefits.

6. Conclusion

With the Scrapping, Preprocessing, Analysis, Modeling & Results we conclude that:

- A restaurant near metro station will be beneficial for customers & metro commuters.
- We can open a restaurant where there is more number of metro users.
- While we conclude the above statements, we also have evidences which prove that there are fewer restaurants where metro users are less.
- However, we do require additional data to support this argument. We do have enough restaurants in the crowd area of Dubai which is Burjuman.
- We are able to predict cuisines which will be most suitable for the Restaurant.
- We can have a multi-cuisine restaurant at the heart of Dubai which is expected to have less setup cost, more utility and more profit as well.

7. Drawbacks

The data that we used has been provided by https://bayanat.ae. It is a government organization in Dubai, which deals with statistical & relative data of Dubai. We have the data of Red & Green lines metro station, but there is a newline getting constructed for Expo 2020. This new line shall be operable temporarily. We didn't found any data regarding the same.

Moreover, we need to gather data about the tariff rates of commercial places in Dubai, area wise. That will give us good insights and we will be able to conclude exact location, distance from metro and proper cuisine which will be feasible for the restaurant in Dubai.