

Faisal Alhusayni

Capstone Project - The Battle of the Neighborhoods

06 May 2021

## Visit Madinah City

<b>Visit Madinah City</b>	<b>1</b>
1.Introduction	<b>3</b>
1.1 Background	3
1.2 Problem	3
<b>2. Data</b>	<b>3</b>
2.1 Data sources	3
2.2 Data cleaning	4
<b>3. Data Analysis and visualization</b>	<b>5</b>
<b>4. Modeling</b>	<b>5</b>
<b>5. Discussion:</b>	<b>7</b>
5.1 Explaining the results	7
<b>Conclusion:</b>	<b>8</b>

# 1.Introduction

## 1.1 Background

AL madinah city is one of the holy cities for muslims which have the prophet Mohammed mosque that a lot of muslims visit it every year, so this report is going to help visitors to choose the place that has the most venues nearby by clustering using the machine learning KMeans clustering method.

## 1.2 Problem

Choosing the right place like nearest (Hotels, Restaurants, Coffee shops and Stores ) is sometimes hard for visitors to choose. By using data analyses and machine learning will help the visitors to choose the right place.

# 2. Data

## 2.1 Data sources

Getting the PostalCode, Neighborhood, Latitude and Longitude from [Google Maps](#).

	PostalCode	Neighborhood	Lat	Long
0	42311	Bada'ah	24.472544	39.607427
1	42311	Al Manakhah	24.470527	39.602358
2	42311	Bani Khidrah	24.464277	39.608091
3	42313	Bani Abdul Ashhal	24.478975	39.616919
4	42316	Qurban	24.452750	39.612170
5	42315	Al Mughaisilah	24.456612	39.599381

Table 1. Google maps postal codes data

Getting this data give the ability to get (Venue, Venu Latitude, Venu Longitude, Venu Category) from [foursquare](#)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bada'ah	24.472544	39.607427	Dar Al Iman InterContinental (دار الإيمان إنتر...	24.471807	39.608753	Hotel
1	Bada'ah	24.472544	39.607427	Saja Al-Madinah Hotel	24.473803	39.607655	Hotel
2	Bada'ah	24.472544	39.607427	Anwar Al Madinah Mövenpick (فندق أنوار المدينة)	24.471569	39.607839	Hotel
3	Bada'ah	24.472544	39.607427	I Love Madinah (أنا أحب المدينة)	24.471614	39.609199	Gift Shop
4	Bada'ah	24.472544	39.607427	The Beautiful Names of Allah Exhibition (معرض ...)	24.469822	39.607687	Museum

Table 2. Foursquare venue data

## 2.2 Data cleaning

Data downloaded or scraped from multiple sources must be cleaned and fill the missing values and combine the tables into one table.

First grouping the venue count data by the Neighborhoods to calculate how many venues in every Neighborhood.

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Al Anabis	10	10	10	10	10	10
Al Aridh	11	11	11	11	11	11
Al Fath	17	17	17	17	17	17
Al Iskan	15	15	15	15	15	15
Al Khatim	4	4	4	4	4	4
Al Manakhah	13	13	13	13	13	13

Table 3. Groupby Neighborhood

Then combine the two tables into one table with the maen values of all of the values.

	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arts & Crafts Store	BBQ Joint	Bakery	Big Box Store	Boarding House	...	Supermarket	Swiss Restaurant	Tech Startup	Tou Provide
0	Al Anabis	0.00	0.000000	0.000000	0.000000	0.00	0.0	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
1	Al Aridh	0.00	0.090909	0.000000	0.000000	0.00	0.0	0.090909	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
2	Al Fath	0.00	0.000000	0.000000	0.000000	0.00	0.0	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
3	Al Iskan	0.00	0.000000	0.000000	0.000000	0.00	0.0	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
4	Al Khatim	0.00	0.000000	0.000000	0.000000	0.25	0.0	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
5	Al Manakhah	0.00	0.000000	0.000000	0.000000	0.00	0.0	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
6	Al Mughaisilah	0.25	0.000000	0.000000	0.000000	0.00	0.0	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
7	Al Qiblatayn	0.00	0.000000	0.052632	0.000000	0.00	0.0	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
8	Al Suqya	0.00	0.000000	0.000000	0.000000	0.00	0.0	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
9	As Sih	0.00	0.000000	0.000000	0.333333	0.00	0.0	0.333333	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000

Table 4. combine the two tables

### 3. Data Analysis and visualization

Showing in the bar chart below that the most neighborhood that has more venues is (Bada'ah) neighborhood.

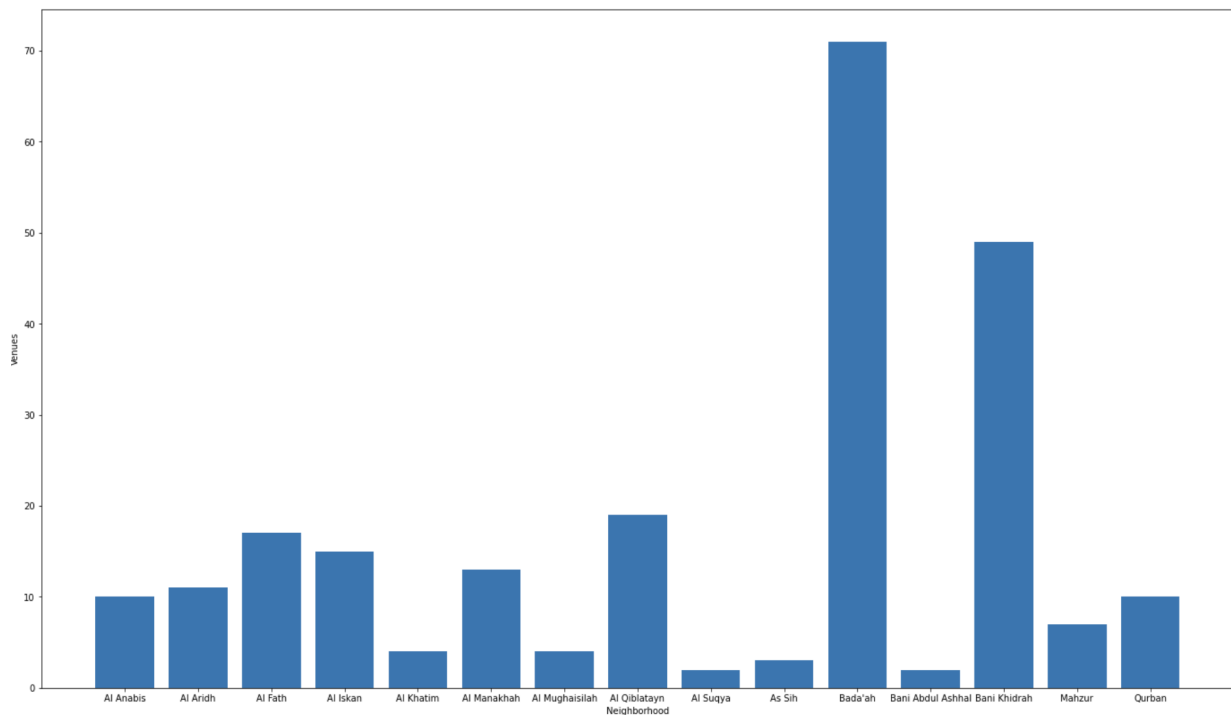


Figure 1. bar chart

### 4. Modeling

In this project used K-Means clustering which can group data only unsupervised based on the similarity of the data.

Using this methodology to group the venues of Madinah city neighborhood data to different clusters, the tools that used to calculate the K-Means is Sklearn with Python.

- Adding the cluster labels to the dataframe

	PostalCode	Neighborhood	Lat	Long	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
0	42311	Bada'ah	24.472544	39.607427	3	Hotel	Shopping Mall	Pizza Place	Ice Cream Shop	Museum	Gift Shop	Donut Shop	Fast Food Restaurant	Boarding House
1	42311	Al Manakhah	24.470527	39.602358	3	Hotel	Breakfast Spot	Pharmacy	Middle Eastern Restaurant	Plaza	Grocery Store	Gift Shop	Garden	Furniture Store
2	42311	Bani Khidrah	24.464277	39.608091	3	Hotel	Coffee Shop	History Museum	Ice Cream Shop	Museum	Plaza	Café	Pakistani Restaurant	Middle Eastern Restaurant
3	42313	Bani Abdul Ashhal	24.478975	39.616919	1	Hotel	Donut Shop	Falafel Restaurant	Farmers Market	Fast Food Restaurant	Food & Drink Shop	Food Court	Food Truck	French Restaurant
4	42316	Qurban	24.452750	39.612170	4	Juice Bar	Afghan Restaurant	BBQ Joint	Restaurant	Shopping Mall	Kids Store	Fried Chicken Joint	Ice Cream Shop	Pharmacy

Table 5. clustering

- Using the map to visualize the clusters

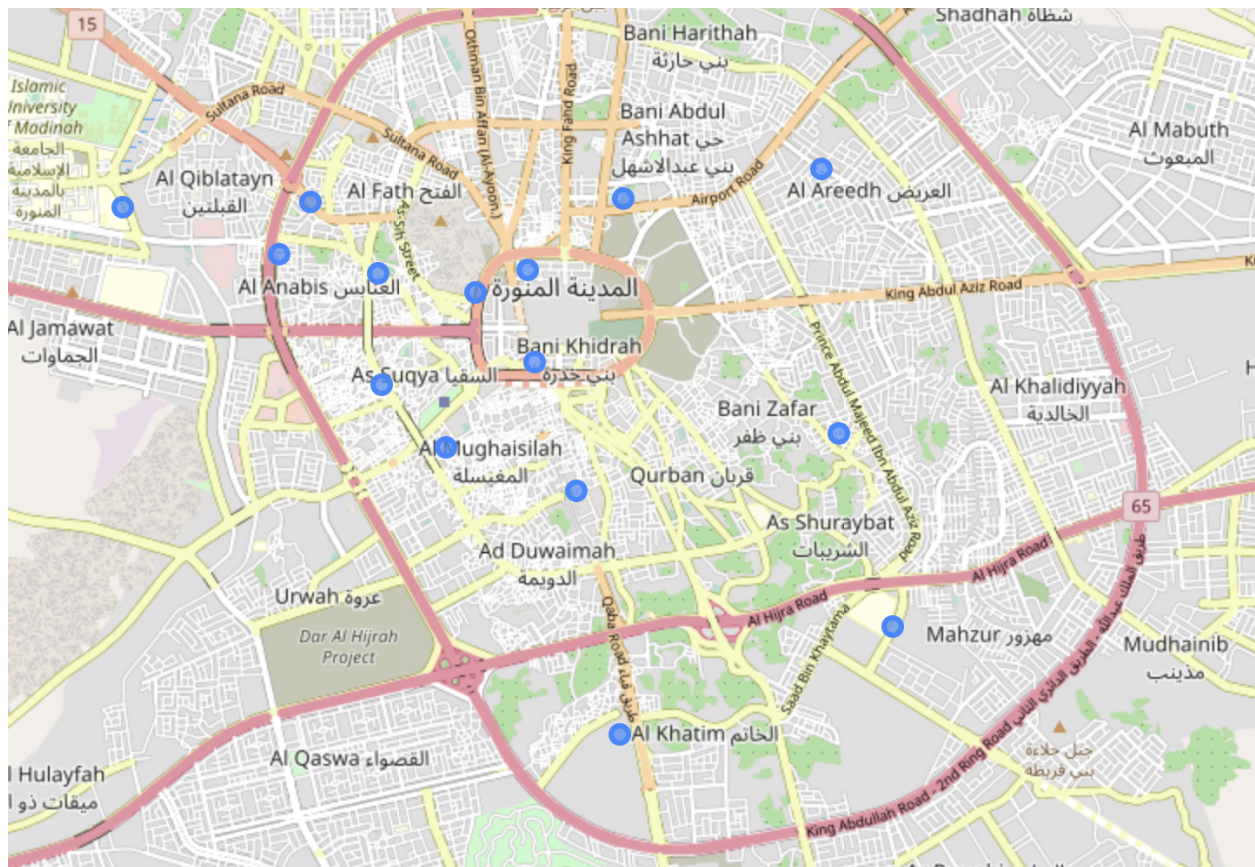


Figure 2. The map of madinah city with clustering

## Conclusion:

As we built our list of neighborhoods with Restaurant venues exclusively we discovered most neighborhoods were similar and the greatest concentration of restaurants was in Madinah.

This helps visitors to choose the right neighborhood after clustering results.