Exploring the neighborhoods in Doha,Qatar  
For starting a crepe shop.

Applied Data science Capstone project .

By : Ahmad Karasneh

Date 09th -June-2021

Table of Contents

**Introduction ......................................................................... 1**

**Business problem ..................................................................... 2**

**Data Required ……………........................................................... 3**

**Data collection ………………………....................................................4**

**Methodology….............................................................................7**

**Machine learning .........................................................................12**

**Results and discussions ………………………………………………………………13**

**Conclusion ………………………………………………………………………………….14**

Introduction



Doha is the capital and most populous city of Qatar. It has a population of **956,460** .

It is Qatar's fastest growing city, with over 80% of the nation's population living in Doha or its surrounding suburbs, and it is the economic centre of the country.

Having said the above , an existent business owner is interested in the idea of opening a new dessert shop ‘**Crepeaholic’** in Qatar ,Doha.

A picture containing icon

Description automatically generated

Crepeaholic, an upscale creperie, was founded in Amman, Jordan in the year 2013. Because of its booming success, it has expanded into an international franchise, and is now also in the glistening capitol of the UAE, Abu Dhabi.

Crepeaholic is known for its hot, freshly-baked savory and sweet crepes and waffles, which are crafted with top quality flour and naturalx homemade ingredients.

Business Problem

In this project we aim to explore the idea of opening a new Crepes&Waffles shop in Doha,Qatar .Followed by recommendation of the best location for it .

First , we would like to research the the demographics of Qatar and Doha in particular . IF possible ,pursued by analysis of market growth (Resturants/cafes count during time) .with intention to investigate the question of how many resturants/cafes were opened and closed in a short period of time .

Since there are lots of restaurants&cafes in Doha, we will try to detect locations that are not already crowded with restaurants/cafes but crowded with people. and as per the owner ; the focus shall be on family oriented areas and with a preference of a community aged between 16-35 years old . We are also particularly interested in areas with no or few desserts&Cafe shops in vicinity. We would also prefer locations as close to city center as possible, considering all other conditions are met .

Data Required

Based on definition of our business case, factors that will influence our decision are:

* Number of current restaurants/cafes in Doha and each Neighbourhood.
* A time series that shows the growth of restaurants/cafes In Doha and specifically the targeted kind of shops.
* Demographics of each neighbourhood .
* If available ,investigate the number of restaurants and cafes that were opened and closed in the period of 2017-2021.
* Distance of neighbourhood from city centre.
* Top 5 venue categories in each neighbourhood and its corresponded average customer review rates.

Data Needed :

* List of Doha's Neighbourhoods.
* Demographics tables about Doha .
* Foursquare API location data.
* If available , customer reviews and rates for each venue in Foursquare data.
* If available , foot traffic and check-ins details .

Data collection

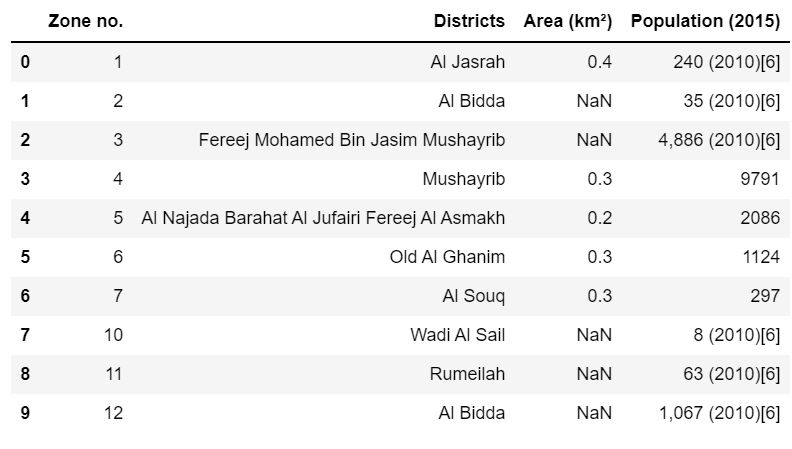
**Neighbourhoods ,Area and population Data**

The data of the neighbourhoods and population in Doha was scraped from

<https://en.wikipedia.org/wiki/Zones_of_Qatar#Ad_Dawhah_Municipality>.

The data is read into a pandas data frame using the read\_html() method.

The top 10 rows of the Data-frame are shown in **Figure 1**



**Geographical coordinates Data**

The data of geographical coordinates can be obtained from the **Geocoder Library**.

Using Geocoder ,we will bring the latitude and longitude for neighbourhood as below table for the head 20 neighbourhoods before data cleansing :

Table

Description automatically generated

**Venues’ data**

The venue data has been extracted using the **Foursquare API**. This data contains

venue recommendations for all neighbourhoods in Doha and is used to study

the popular venues of different neighbourhoods as well as build the unsupervised

learning model to cluster neighbourhoods. The venue recommendations of all

neighbourhoods were obtained with a limit of 200, that is, maximum of 200 venue

recommendations per neighbourhood and a radius of 1 km around the

neighbourhood’s geographical coordinates. Figure 5 shows the top 5 rows

depicting the results obtained after cleaning the data from Foursquare API

Table

Description automatically generated with low confidence

Methodology

This section provides details for the methodology used in the project.

As we are constrained with the data provided and its limits . our aim in this study would be understanding the similarities between neighbourhoods based on # of venues , type of venues , number of dessert shops , population and population densities .we will be running an unsupervised machine learning technique   
called K means clustering to group the neighboorhoods based on the similarities .   
with conjunction with data visualization in graphs and maps , we will be able to reach our goal of this study to determine a location for our new dessert shop that we think will have more potential to succeed .

* In order to analyse the neighbourhoods and venues’ surrounding , we need to cleanse the data and make sure our data is correct ,logical , readable with right formats and doesn’t have duplicates .

**Neighbourhoods**

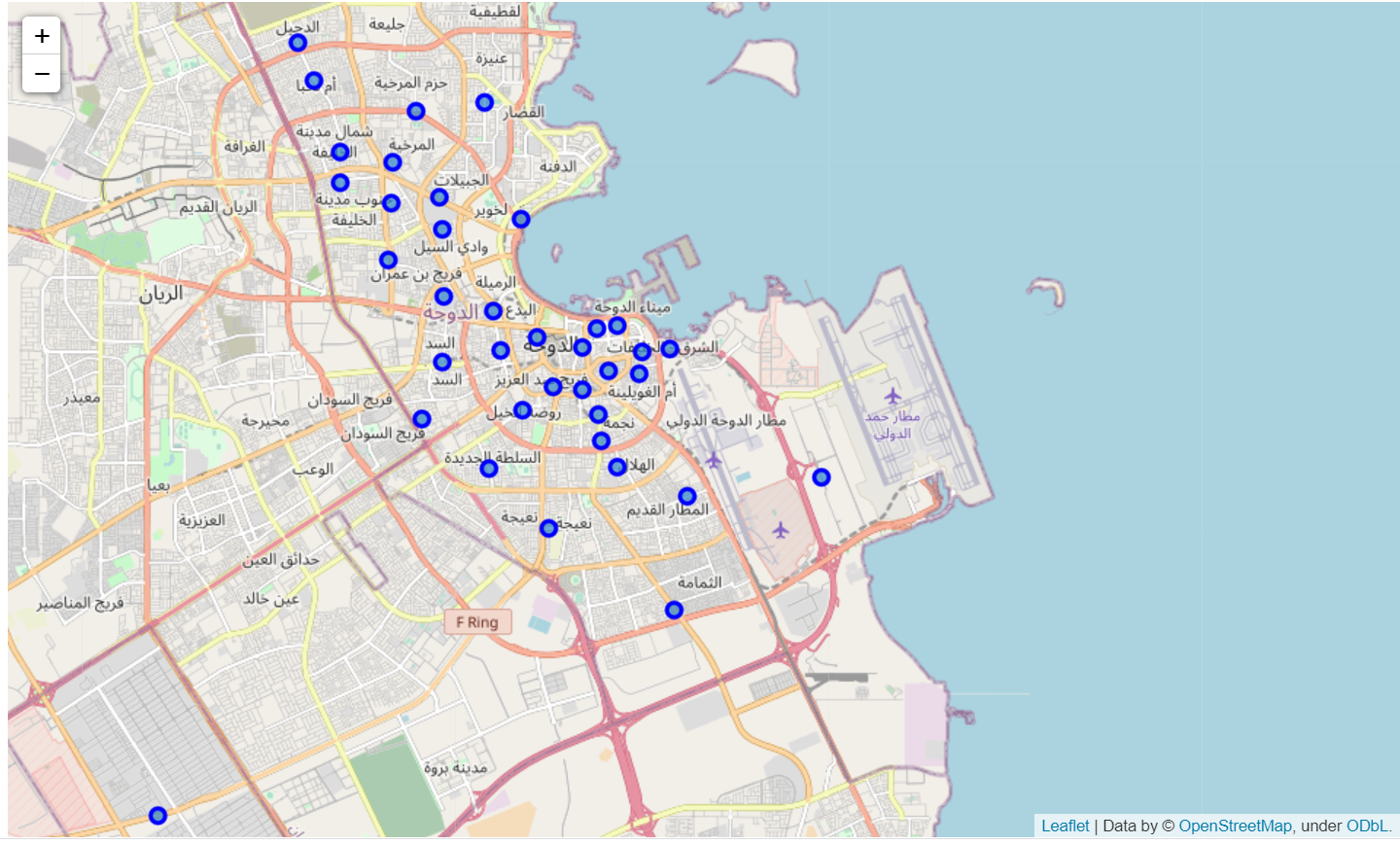
* After Examining the available raw data , these points were noticed :
* **The neighbourhoods data** consists of columns : Zone #, District or Neighbourhood, Area ,and corresponded Population.
* Data extracted from Wikipedia but the reference is a report from Ministry of development planning and statistics in Qatar.
* Data originally had 56 neighbourhoods/districts.
* Some population values are based on old report in 2010 .by observing these values , we can see they represent a very minimal out of total .Thus , we may remove these rows for the sake of our study .
* Last row represents the total for all Doha and we may drop it .
* some districts columns have more than one neighbourhood inside it.
* Population data type is String, we needed to change it to Integer.
* Population density was calculated by dividing the population by the area of neighbourhood .
* It was noticed that there are duplicates in the Neighbourhood column but it has different population and areas. We will assume that each row of them is a part of total , therefore we will group them based on Neighbourhoods .
* After making the amendments on the neighbourhoods’ data , the resultant table consists of 41 neighbourhood in Doha , Qatar .
* Plotting the population density Vs Neighbourhoods .

Chart, histogram

Description automatically generated

**Geographical coordinates .**

* After extracting the longitude and latitude for each neighbourhood , we Visualized and plotted them in maps :



* It was noticed that two neighbourhoods are having the same geographical coordinates (zone 50 and zone 58 ), the one with least population were eliminated from the list .

**Venues**

* One of the objectives included in the study was to determine how the dessert shops and restaurants/cafes in general were growing during time . Another objective was to detect how many dessert shops were opened and closed during the time between 2015 till today. Both information could have been used as a signal of venues growth and good /bad investing trend in each neighbourhood .

Unfortunately , the data from the FOURSQUARE API doesn’t provide historical info . thus , these questions were skipped in the study .

* As we have taken the exploration radius as 1 KM ,some neighbourhoods might interfere in each other area boundaries .

Therefore , we have to remove the duplicates venues that appear in various neighbourhoods .

* We would like to examine the number of venues for each category , as below sample from the table :

Table

Description automatically generated

* And more interestingly for the purpose of this study , we would like to see the number of desserts shops in each neighbourhood as below :

Table

Description automatically generated

Machine learning

* K-means unsupervised learning technique was used to cluster neighbourhoods based on the category of venues near neighbourhoods.
* We grouped the data obtained from the FOURSQUARE API to be based on neighbourhoods and converted to the numerical variables by using   
  get\_dummies method in pandas .
* In later stage , we also added the columns of ( population , density , # of venues ) as to be considered while clustering .
* As data has different scales , we have used a standardization of the X’s before running the clustering process . we have used StandardScaler Library .
* The number of clusters were chosen based on Silhouette scores.

And K equals to 3 .

Results

* An output was obtained after fitting a clustering model using   
  SK-Learn library and was plotted in map to have wider view in order to take the right Decision about the next shop’s location in Doha .

The map as below :

Map

Description automatically generated

Conclusions

* We have successfully analysed the neighbourhoods in Doha, Qatar for determining which would be the best neighbourhoods for opening a new dessert shop. We have considered in our decision the location data and some demographics about the city . We have plotted the neighbourhoods and its corresponded info in the map above . By looking into map in Conjunction with the current dessert\_shops list , It was noticed that most Dessert shops are located in the north and west of Doha . interestingly ,these locations aren't the most populated ones . Therefore , I see a great opportunity to locate a new dessert shop near the port in the city centre . This area is highly dense populated area having a lot venues and malls(which's a sign of good traffic ) but not dessert (crepe and waffles) shops . The Industrial area is highly populated and has no dessert shops but I eliminated it as it's unlikely for dessert shops to work successfully in such areas .
* The optimal neighbourhoods would be :

1. Ad Dawhah al Jadidah --- as it's the centroid of other populated neighbourhoods and it seems easier to reach .
2. Old Al Ghanim