# CAPSTONE PROPOSAL

**NAME: SIDHARTH KUMAR MOHANTY**
**DATE: 1st June 2020**

**TOPIC: THE BATTLE OF THE NEIGHBOURHOODS**

**Blog Post Link :**

This is my link of the blog post  about this project
https://medium.com/@sidharth.ku178/the-battle-of-the-neighbourhoods-3d90f56dae69

**Domain Background :**

Different cities in the world are filled with numerous kinds of venues that in turn define the cultures of the cities. A city not only differs from another by means of global positioning, what it to showcase to its inhabitants or tourists has put a significant mark on differentiating it from the rest. Despite of having dissimilarities, it is somewhat possible to group together the similar kind of neighborhoods in different cities. It is possible to segment the different venues in a neighborhood according to venue category, and then to group neighborhoods together that incorporate similar kind of neighborhoods. Human migration is the movement by people from one place to another with the intentions of settling, permanently or temporarily in a new location. Having grouped together similar kind of neighborhoods may serve as a variable to help make a decision when people consider moving out of a city to another.

**Problem Statement**

The mission of this project is to use Foursquare location data and regional clustering of venue information to determine what might be the 'best' neighborhood in Dubai to open a restaurant.  As a westerner who has a passion for good Mexican food, I have found that there is not a lot of selection in the region.  This is supported by the fact that a review of Foursquare venues reveals zero Mexican restaurants listed on the site.Due to its central geographic location and generally welcoming climate, Dubai is known for its international atmosphere.  It is home to close to 2.5 Million expatriates who make up 90% of the total population[1] and it hosts close to 15 Million visitors per year.[2]  In addition to its multi-cultural population and capital flows, Dubai is becoming well known as a destination of choice for great food.  It is a place where people can rest and try the best of each culture, either while they work here temporarily, or of they are just passing through.

My proposal, then, is an analysis of the neighborhoods in Dubai for the consideration of opening a new Mexican restaurant.  The objective is to have a location that is within one of the more reasonable rent zones, but also within a close enough range (5km) to a 'high' rent zone.  The assumption that proximity to high rent neighborhoods would result in takeout opportunities or provide residents an option to travel to our restaurant.

## Datasets and Inputs

The data that we will use for this analysis is a combination of a CSV file that has been prepared for the purposes of the analysis from multiple sources (Dubai_neighborhoods.csv) and the location/venue information in foursquare.

For this project we need the following data :

1. Dubai City data that contains list Boroughs, Neighborhoods along with their latitude and longitude.
2. Indian resturants in each neighborhood of Dubai.
3. GeoSpace data

Source 1: Dubai_neighborhoods.csv

We will first determine the most likely neighborhoods for a restaurant based on average rental prices and relative distance to a high rent center. In the case of Dubai, the three highest rent averages can be found in Palm Jumeirah, Jumeirah, and Zabeel. We will then consider the total number of venues and additional criteria such as proximity to a shopping center or offices for midday traffic to make a final determination.

The first step is to etablish the neighborhoods in Dubai and a summary of their average rental prices. The average rental index is published annually and can be referenced by a number of different websites. We then make a calculation of the 'Z-score' to standardize the data and sort from lowest to highest average rent. Using google, I looked up the latitude and longitude of each neighborhood and entered it. This information is used to calculate the distance of each neighborhood from the highest rent regions and will also be entered to the Foursquare database for venue query later.

| Neighborhood | Avg Rent Per Unit | Z-Score | Average Distance from High Rent Areas | | | Latitude | Longitude |
| | | | Distance from Palm | Distance from Zabeel | Distance from Jumeirah | | |
|---|---|---|---|---|---|---|---|
| Discovery Gardens | 44,672 | -1.53 | 8.18 | 26.15 | 20.73 | 25.039 | 55.1445 |
| Dubai Silicon Oasis | 54,417 | -1.30 | 24.96 | 13.31 | 16.39 | 25.1279 | 55.3863 |
| Jumeirah Village Circle | 60,068 | -1.17 | 9.16 | 20.56 | 16.13 | 25.0602 | 55.2094 |
| Dubai Sports City | 62,753 | -1.10 | 11.36 | 22.32 | 18.28 | 25.0391 | 55.2176 |
| Remraam | 67,284 | -0.99 | 16.71 | 25.27 | 22.27 | 25.0014 | 55.2508 |
| Al Furjan | 73,648 | -0.84 | 9.70 | 27.28 | 22.02 | 25.0252 | 55.1459 |
| Jumeirah Village Triangle | 82,014 | -0.64 | 8.87 | 22.78 | 18.04 | 25.0473 | 55.19 |
| Motor City | 83,876 | -0.60 | 12.61 | 20.90 | 17.42 | 25.045 | 55.2397 |
| Damac Hills | 94,630 | -0.34 | 16.40 | 22.41 | 19.37 | 25.0275 | 55.2524 |
| Al Sufouh | 95,804 | -0.31 | 0.70 | 17.88 | 12.02 | 25.1134 | 55.1762 |
| DIFC | 105,183 | -0.09 | 17.86 | 3.02 | 3.57 | 25.2106 | 55.2794 |
| Business Bay | 105,682 | -0.08 | 15.61 | 5.55 | 3.45 | 25.1832 | 55.2729 |
| Jumeirah Lakes Towers | 106,352 | -0.06 | 4.80 | 23.80 | 18.03 | 25.0693 | 55.1417 |
| Barsha Heights | 111,804 | 0.07 | 4.10 | 19.08 | 13.48 | 25.097 | 55.1776 |
| Emirates Living | 114,422 | 0.13 | 7.82 | 23.43 | 18.36 | 25.0496 | 55.174 |
| Dubai Marina | 115,236 | 0.15 | 3.55 | 23.02 | 17.12 | 25.0805 | 55.1403 |
| Dubai Investments Park | 116,379 | 0.18 | 15.30 | 30.13 | 25.71 | 24.979 | 55.1762 |
| Jumeirah Beach Residence | 143,520 | 0.83 | 3.97 | 23.75 | 17.82 | 25.0769 | 55.1341 |
| Dubai Festival City | 151,341 | 1.02 | 25.23 | 5.60 | 11.81 | 25.2171 | 55.3614 |
| Downtown | 153,546 | 1.07 | 16.77 | 4.19 | 3.41 | 25.195 | 55.2784 |
| Zabeel | 176,213 | 1.61 | 20.84 | 0.00 | 6.57 | 25.2231 | 55.3061 |
| Jumeirah | 180,180 | 1.71 | 14.59 | 6.57 | 0.00 | 25.2016 | 55.2453 |
| Palm Jumeirah | 204,430 | 2.29 | 0.00 | 20.84 | 14.59 | 25.1124 | 55.139 |

**Target =** Neighborhood with reasonable rent and close proximity to high rent area: Zabeel, Jumeirah or Palm Jumeirah

Low Rent Avg

High Rent Avg

Source 2: Venue data via Foursquare:

Using clustering techniques developed in prior exercises in the capstone project, we will examine the most common venues by neighborhood listed in Foursquare. When we cluster the data together and rank by 'most common', it appears that Mexican venues not registered in any of the neighborhoods, so it is safe to assume there would be minimum competition.
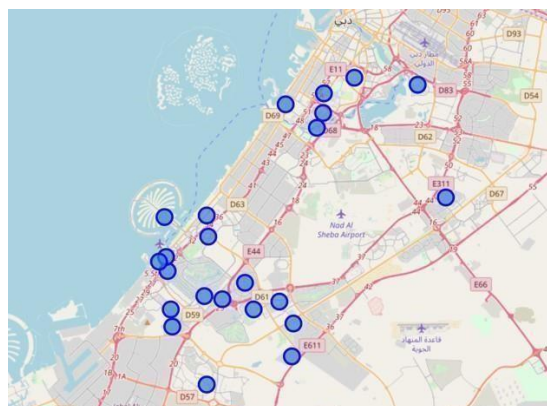


*Map of Dubai neighborhoods using Folium*

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|
| 0 | Business Bay | Restaurant | Middle Eastern Restaurant | Italian Restaurant | Tapas Restaurant | Chinese Restaurant |
| 1 | Deira | Asian Restaurant | Indian Restaurant | Vietnamese Restaurant | Japanese Restaurant | Brazilian Restaurant |
| 2 | Downtown | Middle Eastern Restaurant | American Restaurant | Restaurant | Asian Restaurant | Turkish Restaurant |
| 3 | Dubai Marina | Middle Eastern Restaurant | Italian Restaurant | Asian Restaurant | Restaurant | French Restaurant |
| 4 | Jebel Ali | Italian Restaurant | Seafood Restaurant | Ethiopian Restaurant | Indonesian Restaurant | Indian Restaurant |
| 5 | Jumeirah Beach | Restaurant | Thai Restaurant | Sushi Restaurant | Moroccan Restaurant | English Restaurant |
| 6 | Jumeirah Lake Towers | Italian Restaurant | Vietnamese Restaurant | Modern European Restaurant | Greek Restaurant | Indian Restaurant |
| 7 | Media City | Middle Eastern Restaurant | Italian Restaurant | French Restaurant | Fast Food Restaurant | Indonesian Restaurant |
| 8 | Old Dubai | Middle Eastern Restaurant | Vietnamese Restaurant | Fast Food Restaurant | Indonesian Restaurant | Indian Restaurant |
| 9 | Palm Jumeirah | Restaurant | Seafood Restaurant | Indian Restaurant | Brazilian Restaurant | English Restaurant |

The remaining steps in the capstone project will be to evaluate which neighborhood is most suited using the criteria shown above: the rent index of the neighborhood is close to the mean average rent in Dubai, the neighborhood's is located close to a 'high rent' neighborhood, the general restaurant frequency in the neighborhood is reasonable, and proximity to 'other' venues such as business centers or malls is maximized.

**Data Exploration**

As shown above, a quick sort of average rent by neighborhood and geographic proximity to the high rent regions narrows our search to a few neighborhoods. Although other neighborhoods could be selected due to low rent, they may be too far away from an area that would have more propensity for residents to dine out. The regions shown in green are also high traffic areas for tourists due to their proximity to shopping centers and high density of hotel accommodation.

Therefore, further analysis will focus on (6) neighborhoods: **Al Sufouh, DIFC, Business Bay, Jumeirah Lake Towers, Barsha Heights, and Dubai Marina**.

Each is within 5 Km of a high rent district but is in the middle of the range of the Dubai rent index. Travel to the restaurant would give families an opportunity to 'get out' without going too far, and delivery is also an option.



So then the question becomes, which of these neighborhoods is preferred? Aside from proximity to a high rent neighborhood, we also want to take into consideration the amount of commercial activity in the selected neighborhood, tourist traffic and relative competition from other restaurants. We have already established that there would not be many 'Mexican' restaurants available, but we also do not want to enter a region that is over-crowded with options.

To begin this analysis, I first made a master data frame of all venues in consideration, and then flagged any that had the word 'Restaurant' or 'Hotel' as 'Venue Type', dropping remaining.
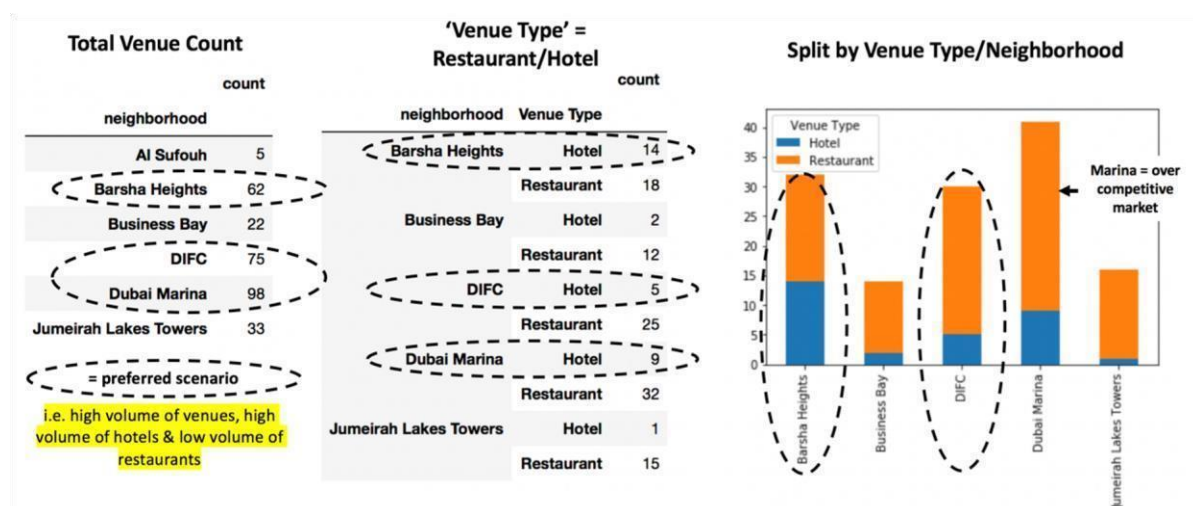


The outcome is to consider 'Barsha Heights' or 'DIFC' as a final consideration for launch of a new Mexican restaurant. Both list 'Hotel' as the most common venue. Dubai Marina is a close third given the volume of venues and hotels, but the number of restaurants is already very high.

| neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|
| 0 | Barsha Heights | Hotel | Middle Eastern Restaurant | Italian Restaurant | Thai Restaurant | Hotel Bar |
| 1 | Business Bay | Restaurant | Italian Restaurant | Middle Eastern Restaurant | Hotel | Tapas Restaurant |
| 2 | DIFC | Hotel | Italian Restaurant | Restaurant | Indian Restaurant | Asian Restaurant |
| 3 | Dubai Marina | Hotel | Middle Eastern Restaurant | Italian Restaurant | Asian Restaurant | Restaurant |
| 4 | Jumeirah Lakes Towers | Italian Restaurant | Vietnamese Restaurant | Theme Restaurant | Modern European Restaurant | American Restaurant |

**Final Recommendation**

At first glance, Barsha Heights seems preferable due to the fact that it is one of the top three neighborhoods for volume of venues, and the ratio of restaurants to hotels is relatively less. However, after further consideration, Barsha Heights is only close to one high rent center, while DIFC has the added benefit of being close to three high rent centers.

In terms of relative 'popularity', DIFC would be the preferred location. The acronym stands for 'Dubai International Finance Center'. It well known as the financial capital of the Middle East, attracting investors from all across the region looking to impress clients. It is also home to Emirates Towers (seen in the picture below), which is known be the location of the office of HH Sheikh Mohammed bin Rashid Al Maktoum, ruler of Dubai and Vice President of the UAE.
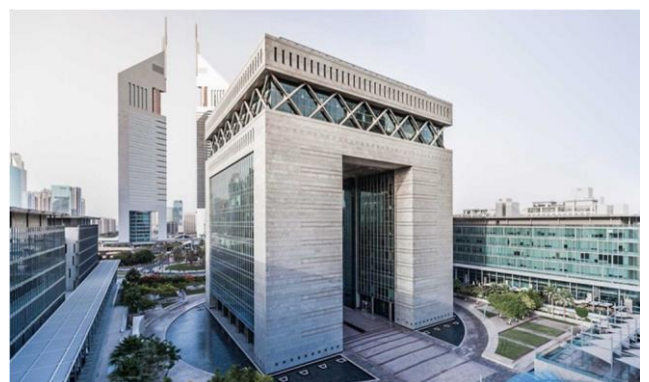


Known as the investment center of the Middle East, DIFC is high traffic, medium rent & close to 3 'high rent' centers.

Finally, in terms of culinary reputation, DIFC known for some of Dubai's most prestigious restaurants like 'Zuma' and 'Hakkasan' – international chains known for high value and sought-after cuisine. There is a popular gallery section for art collectors that is home to 'Christies' the high value auction house known for high end trading.

## Solution Statement:

This capstone project deals with the process of leveraging location data acquired from data providers such as Foursquare to explore the neighborhoods within a targeted city and create clustering models. Using K-means cluster, similar locations with minimum distance shall be grouped into clusters. It is the simplest form of unsupervised machine learning algorithm and it helps in grouping similar data points. Utilizing this model, I intend to create a solution to use Foursquare location data and regional clustering of venue information to determine what might be the 'best' neighborhood in Dubai to open a restaurant.

**Benchmark Model:**

Using location as a strategy, businesses analyze seemingly different data for patterns and trends in their occurrence within a geographic region. Moreover, by visualizing the data, relationships can be identified that are crucial for strategic. So in light of this, I built my first capstone project in machine learning to use Foursquare location data and regional clustering of venue information to determine what might be the 'best' neighborhood in Dubai to open a restaurant.

## Pearson correlation matrix for neighbourhood scale items :

This matrix shows the pearson correlation matrix for neighbourhood scale items

| | Pleasantness for walking | Attractiveness | Proximity to park | Green space | Public transport | Proximity to shops | Routes for cycling | Routes for walking | Safety walking after dark | Likelihood of attack | Traffic volume | Traffic noise | Safety crossing the road |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Attractiveness | 0.48 | | | | | | | | | | | | |
| Proximity to park | 0.23 | 0.16 | | | | | | | | | | | |
| Green space | 0.18 | 0.26 | 0.26 | | | | | | | | | | |
| Public transport | 0.19 | 0.10 | 0.16 | 0.01 | | | | | | | | | |
| Proximity to shops | 0.06 | 0.09 | 0.19 | 0.16 | 0.16 | | | | | | | | |
| Routes for cycling | 0.24 | 0.15 | 0.14 | 0.09 | 0.04 | -0.06 | | | | | | | |
| Routes for walking | 0.29 | 0.33 | 0.23 | 0.29 | 0.13 | 0.29 | 0.17 | | | | | | |
| Safety walking after dark | 0.37 | 0.27 | 0.18 | 0.10 | 0.15 | 0.05 | 0.22 | 0.13 | | | | | |
| Likelihood of attack | 0.33 | 0.38 | 0.06 | 0.21 | 0.04 | 0.06 | 0.15 | 0.24 | 0.42 | | | | |
| Traffic volume | 0.18 | 0.08 | -0.05 | 0.06 | -0.14 | -0.16 | 0.19 | -0.03 | 0.16 | 0.13 | | | |
| Traffic noise | 0.11 | 0.18 | -0.02 | 0.13 | -0.17 | -0.03 | 0.08 | 0.08 | 0.08 | 0.20 | 0.50 | | |
| Safety crossing the road | 0.21 | 0.15 | 0.19 | 0.11 | 0.12 | 0.03 | 0.18 | 0.13 | 0.25 | 0.18 | 0.29 | 0.22 | |
| Road safety for cyclists | 0.14 | 0.18 | -0.02 | 0.12 | -0.08 | -0.01 | 0.22 | 0.15 | 0.12 | 0.27 | 0.28 | 0.40 | 0.27 |

n = 125

## Project design:

For this project, I intended to analyze the neighborhoods of Toronto city to identify ideal locations for starting out a small local cafe business. Now post data extraction and transformation, I connected to the third party APIs providing location data such as Foursquare to extract information on the most popular venues that are located within a radius of one kilometer. The list of nearby venues were extracted using the location data provider's API. K-means Clustering is a machine learning algorithm for unsupervised classification. The algorithm works iteratively to find data groups with similar features that are defined by the variable K. In this method the groups are formed based on the group's centroid feature which is the measure of the feature values of the data points. Using the machine learning algorithm of k-means cluster, I was able to spot similar neighborhoods within a radius of one kilometer that lacked proper shops. The aim of this project is to use Foursquare location data and regional clustering of venue information to determine what might be the 'best' neighborhood in Dubai to open a restaurant.

## Conclusion:

In the end, DIFC would be the best choice considering all the factors: medium rental index, relative high volume of things to do for tourists and locals, it is a business center, it is close in proximity to other high rent neighborhoods, and the ratio of hotels to restaurants is not too high. Other neighborhoods that were considered were Barsha Heights and the Dubai Marina.  Although Barsha Heights seems preferable given the low ratio of hotel venues to restaurants, it is not a particularly well know neighborhood and people would likely not travel there specifically to eat.  It is also only close to one of the 'high rent' neighborhoods.  Finally, Dubai Marina was not a preference as it is already saturated with restaurants.

Thank you