**Capstone Project — The Battle of Communities in Dubai: Restaurants**

Dara Sakhnini

Wednesday, July 8, 2020

# Introduction

Dubai is a growing city with many opportunities for investors and working professionals. The city, once a desert, grows and develops into one of the richest cities in the world. Therefore, it becomes increasingly important to examine and understand its communities quantitatively.

Entrepreneurs, investors, and city planners have an interest in identifying opportunities and growing urban footprints in specific communities. Therefore, after the top food venues for each community are presented, this information could be of interest to potential investors looking to open a new restaurant in Dubai.

Also, for tourists visiting Dubai, it would be of interest to cluster similar communities according to their top food venues. Therefore, when visiting Dubai, for the utmost experience, it would be preferable to visit a community from each cluster in order to insure variety.

# Data requirements

For this project, we need following data:

1. List of communities in Dubai from Wikipedia https://en.wikipedia.org/wiki/List\_of\_communities\_in\_Dubai
2. Latitude and Longitude of each communities from Geocoder from Geopy Library of Python.
3. Different types of venues in each community from Foursquare API.

# Methodology

# Data Preparation

# Scrapping data from Wikipedia

# I first make use of ‘List of communities in Dubai’ page from Wikipedia to scrap the table to create a data frame. For this, I’ve used pandas to transform the data in the table on the Wikipedia page into a data frame containing name of the 131 communities, Area, Population, and Population density, as shown in figure 1.

# A screenshot of a cell phone Description automatically generated

# Getting Coordinates

# Next objective is to get the coordinates of these 131 communities using geocoder class of Geopy client, as shown in figure 2.

# A screenshot of a cell phone Description automatically generated

# Data Cleaning

# Next comes cleaning the data by deleting null values, converting all values into floats, removing units, and deleting unrequired columns. The cleaned data is shown in figure 3.

# A screenshot of a cell phone Description automatically generated

# Map

# I used python folium library to visualize geographic details of Dubai and its 131 major communities. I used latitude and longitude values to get the visual shown in figure 4.

# 

# A close up of a map Description automatically generated

# Data Analysis

# Fetching Foursquare Location Data

# Finally, let’s make use of Foursquare API and get the top 100 venues that are within a radius of 500 meters of each community in Dubai. First four venues are shown in the following figures.

# A screenshot of a cell phone Description automatically generated

# Analyzing Venue Data

# In the figure below, it can be seen that in these Dubai communities, there is a total of 163 unique categories of venues. Also, coffee shops are the most common, as it is at the top of the list in the figure below.

# A screenshot of a social media post Description automatically generated

* + 1. **Filtering Restaurant Data**

# Next, I will concentrate on Restaurant category only. I have filtered the restaurant venues as shown in the sample data frame.

# A screenshot of a cell phone Description automatically generated

# Middle Eastern restaurants top the chart as we can see in the figure below.

# A screenshot of a cell phone Description automatically generated

# Al Karma has the highest number of restaurants, as seen below.

A screenshot of a cell phone

Description automatically generated

* + 1. **Top Restaurants in each community**

Let’s analyze each neighborhood to know about the top 5 venues of each one, as shown in figure a.

# So we create a data-frame with pandas one hot encoding for the venue categories, use pandas group by on neighborhood column and calculate the mean of the frequency of occurrence of each venue category, and finally output each neighborhood along with the top 5 most common venues.

# A screenshot of a cell phone Description automatically generated

* + 1. **Clustering**

Finally, we try to cluster these the communities based on the venue categories using K-Means clustering. So, we expect the communities to be clustered based on the similarities of venue categories.

We finally form a comprehensive data frame with cluster labels, latitude and longitude, most common venues, and population and area data.

# A screenshot of a cell phone Description automatically generated

# We can represent these 3 clusters in a leaflet map using Folium library as shown in the following figure. Burj Khalifa Downtown area, which is represented by a purple dot in figure a, has a variety of restaurants different than other communities.

# A close up of a map Description automatically generated

# Results and discussions

# We got a glimpse of the Restaurants in Dubai and were able to find out some interesting insights which might be useful to travelers as well as people with business interests. Let’s summarize our findings:

# · Middle Eastern restaurants top the charts of most common venues in Dubai communities.

# · Al Karama community has maximum number of restaurants.

# · Since the clustering was based only on the category of restaurants on each district, Burj Khalifa Downtown area has a variety of restaurants different than other communities.

# · Al Mamzar, Za'abeel First, and Al Rashidiya has the least number of restaurants.

# The clustering is based on the most common venues obtained from Foursquare data.

# However, in this analysis, we have ignored other factors like distance of the venues from closest stations, range of prices of restaurants, and so on, since we don’t have such data and it would be difficult to obtain it for a small exploratory study like ours. Hence, our analysis only helps to get a rough overview of Restaurants distribution by categories in the 131 communities of Dubai.

# Also, these results also could potentially vary if we use other clustering techniques.

# Conclusion

# There are many real-life scenarios where data can be extremely useful for a wide range of people, from tourists to people with business interests. Like seen in the example above, data was used to cluster neighborhoods in Dubai based on the most common restaurants in its 131 communities. The results can help a traveler to decide about the district that fit the most his needs.

# I have made use of an API library to scrap data from Wikipedia, used Foursquare API to explore the major communities of Dubai, and visualized the results of clustering the communities using Folium leaflet map. Finally, some of the drawbacks and chance for improvements to represent even more realistic pictures are mentioned.