PREDICTING BEST PHARMACY LOCATION

In Dammam City

By: Saleh Alzahrani

TABLE OF CONTENTS

INTRODUCTION

Dammam is the capital of the Eastern Province of Saudi Arabia. The judicial and administrative bodies of the province, plus several government departments, are located in the city. Dammam is the largest city in the Eastern Province, and the sixth largest in Saudi Arabia, after Riyadh, Jeddah, Mecca, Medina and Taif.

Dammam is a major administrative center for the Saudi oil industry. Together with the nearby cities of Dhahran and Al Khobar, Dammam forms part of the Dammam Metropolitan Area, which is commonly known as greater Dammam and has an estimated population of 4,140,000 as of 2012. Dammam and its suburbs form the center of the Dammam metropolitan area, also known as Greater Dammam, which is closely linked to the city through social, economic, and cultural ties. The city is growing at an exceptionally fast rate of 12% a year - the fastest in Saudi Arabia, the Gulf Cooperation Council, and the Arab world. As of 2016 Greater Dammam is the 4th largest area in both size and population in the Gulf Cooperation Council (GCC).

The growth of the Saudi Arabian oil industry into the largest in the world brought about the rapid development of the region. As oil production increased, so did the number of people required to run the industry. The growing population needed more housing and services. First-rate hospitals and schools provided further incentives for people considering a move to the area. Service industries sprouted up to support the oil industry and meet the needs of people living in the Dammam Area. As a result, a region which had several hundred inhabitants some sixty years ago now boasts a population of well over 1.5 million, growing at a pace of over five percent a year.

The key to the success of the Dammam Area is that unlike oil towns in other parts of the world, it has developed in all spheres. It is now a modern urban and industrial center which happens to be the headquarters of the Saudi Arabian oil industry. As this sector was growing in the early years, the Saudi Arabian government took steps to facilitate the evolution of the Dammam Area. New roads and highways connected the area to other urban and industrial centers in the Kingdom. A railway line connected Dammam to the agricultural center of Al-Kharj and on to Riyadh. Dhahran International Airport was established between Dhahran and Al-Khobar to connect the region to other parts of the Kingdom and the world.

Dammam has over 70 neighborhoods and more are developing which make it good city for investments. In this paper we will study the business in Dammam City and the main goal of this paper is to locate the best locations to open a pharmacy in Dammam.

DATA

To solve the problem, we will use following data:

• List of Postal codes in Dammam which will be used to represents Dammam neighborhoods.

• For plotting the locations we need longitudes and latitudes which will be acquired using ArcGIS Geocoder based on Dammam Postal codes numbers.

• Venue data, particularly data related to Pharmacies. We will use this data to perform clustering on the neighborhoods.

• We will generate a list of Postal codes using Numpy by knowing the min and max Postal code in Dammam using google.

After that, we will use Foursquare API to get the venue data for those

Postal codes. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Pharmacy category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis

METHODOLOGY

1) We will generate a list of all Postal Codes in Dammam City. By using numpy arange Function and then convert this list to pandas Dataframe

2) Using geocoder.arcgis we will search for Latitude and longitude for each Postal code in Dammam City

	Postal_Code	Latitude	Longitude
0	32211	26.499826	50.199823
1	32212	26.437088	50.187243
2	32213	26.421794	50.169127
3	32214	26.410584	50.184936
4	32215	26.440410	50.060270
10	32221	26.412785	50.165901
11	32222	26.401352	50.174430
12	32223	26.402796	50.167200
13	32224	26.405882	50.159680

Figure 1: Generated pandas Dataframe

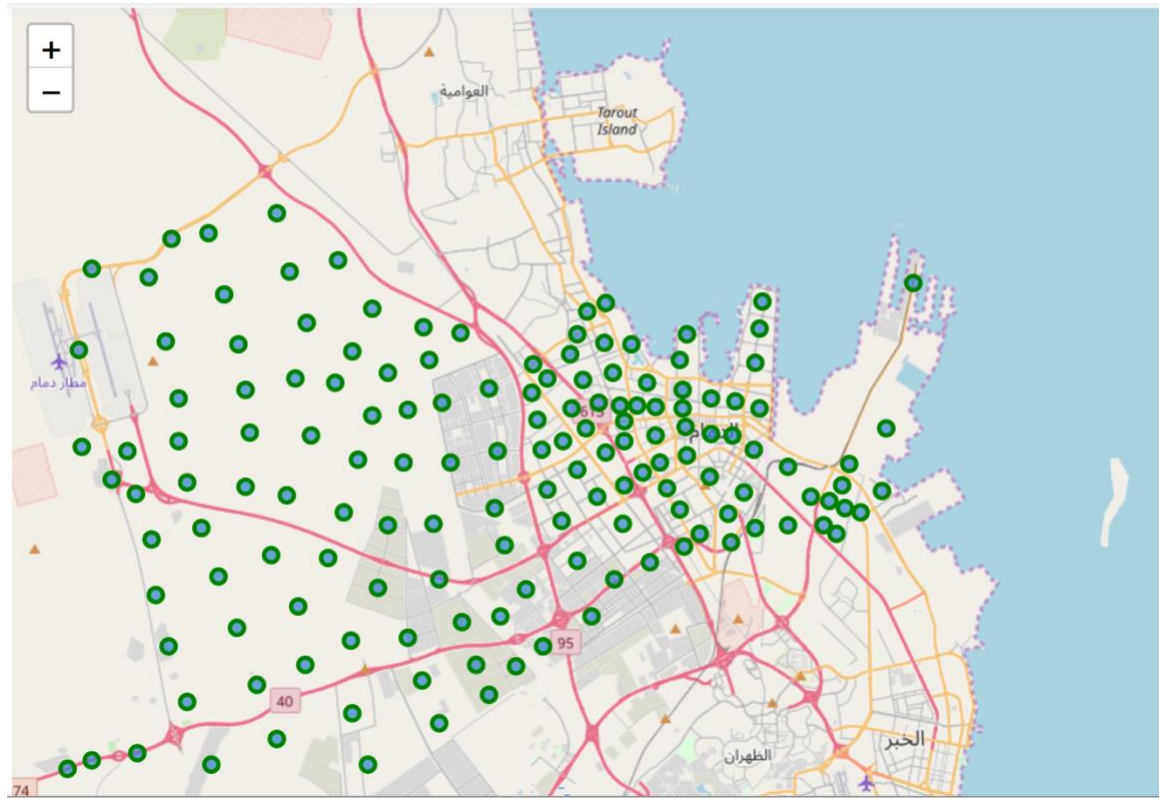3) *Using Folim.Map we will plot a Map of Dammam City with all Postal Codes as Markers*

Figure 2: Dammam Map with Postal Codes Locations

4) *Using Foursquare API we can get information about all venues around each location to study*

5) *With the data, we can check how many venues were returned for each Location and examine how many unique categories can be curated from all the returned venues. Then, we will analyses each Location by grouping the rows by Postal Code and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering. Since we are analyzing the "Pharmacy" data, we will filter the "Pharmacy" as venue category for the Postal Code.*

6) *Lastly, we will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is*

5

one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. We will cluster the locations into 3 clusters based on their frequency of occurrence for "Pharmacy".

RESULTS

The data will be divided into three groups representing the different density of pharmacies located in the area

CLUSTER 0

Low density of Pharmacies. It is represented by red color.

CLUSTER 1

Moderate density of the Pharmacies. It is represented by purple color.

CLUSTER 2

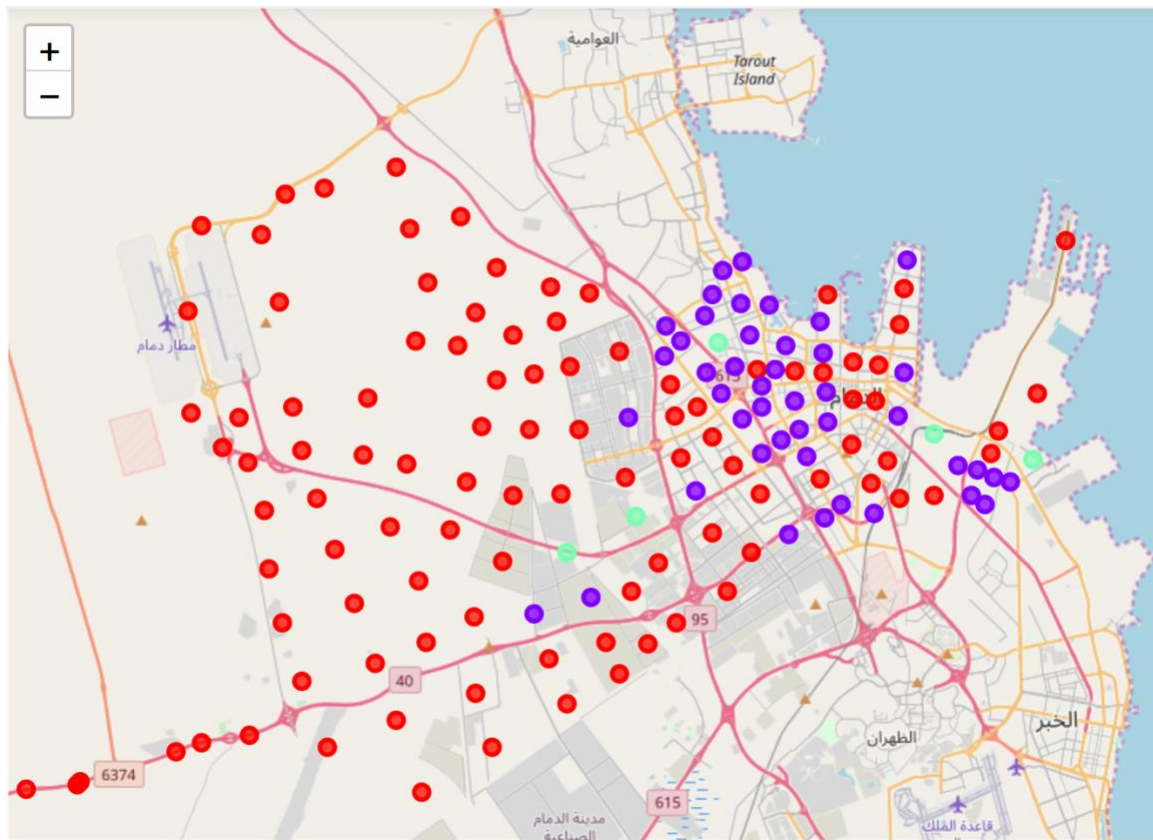High density of Pharmacies. It is represented by mint green color

Figure 3: Dammam Map with Clustered Postal Codes Locations

DISCUSSION

Using only one category we manage to divide Dammam into three categories based on the density of pharmacies in each location. Adding more features will improve the accuracy for our clustering algorithms.

Our suggestion for the features that needed to be take into consideration are as the following:

1) Population of each location
2) Distance from city center
3) Number of customers in nearby venues

And any other relevant features.

CONCLUSION

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 3 clusters based on their similarities.

Using limited source of data and simple Clustering algorithm we manage to cluster Dammam City into 3 categories. Which can give us a good start where to look in Dammam City to open a new Pharmacy.