

PREDICTING BEST PHARMACY LOCATION

IN DAMMAM CITY

BY: SALEH ALZahrani

OUTLINES

- Problem
- Data
- Methodology
- Results
- Discussion
- Conclusion

PROBLEM

- Find the perfect location in Dammam to open pharmacy.
- Make decisions about the right choices with a focus on the influencing factors.
- Extract useful insights using the online data set.

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.

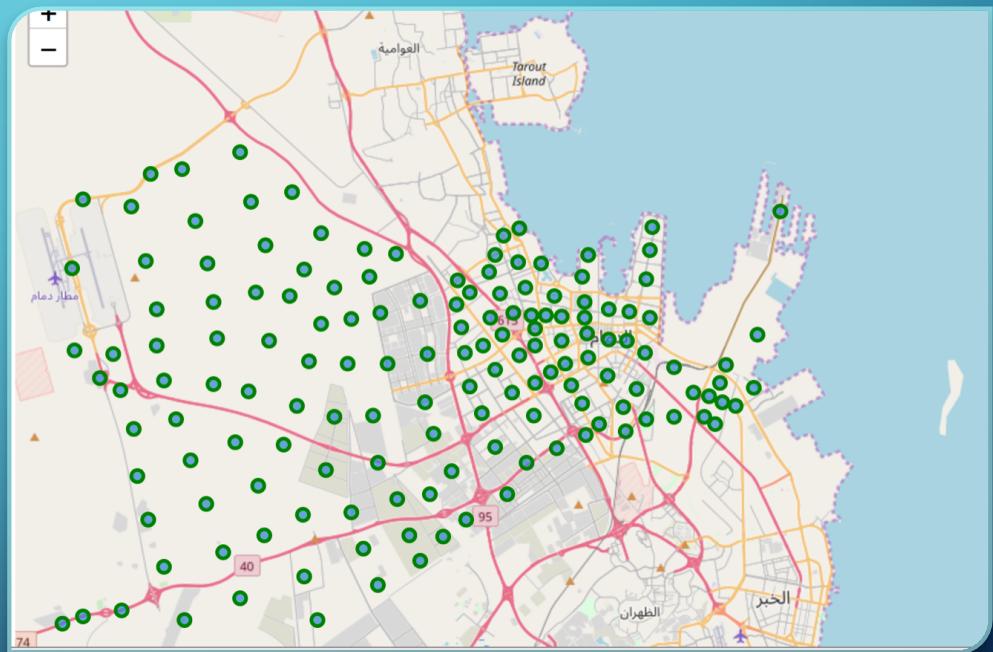
METHODOLOGY

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

METHODOLOGY

- Using Folium.Map we will plot a Map of Dammam City with all Postal Codes as Markers



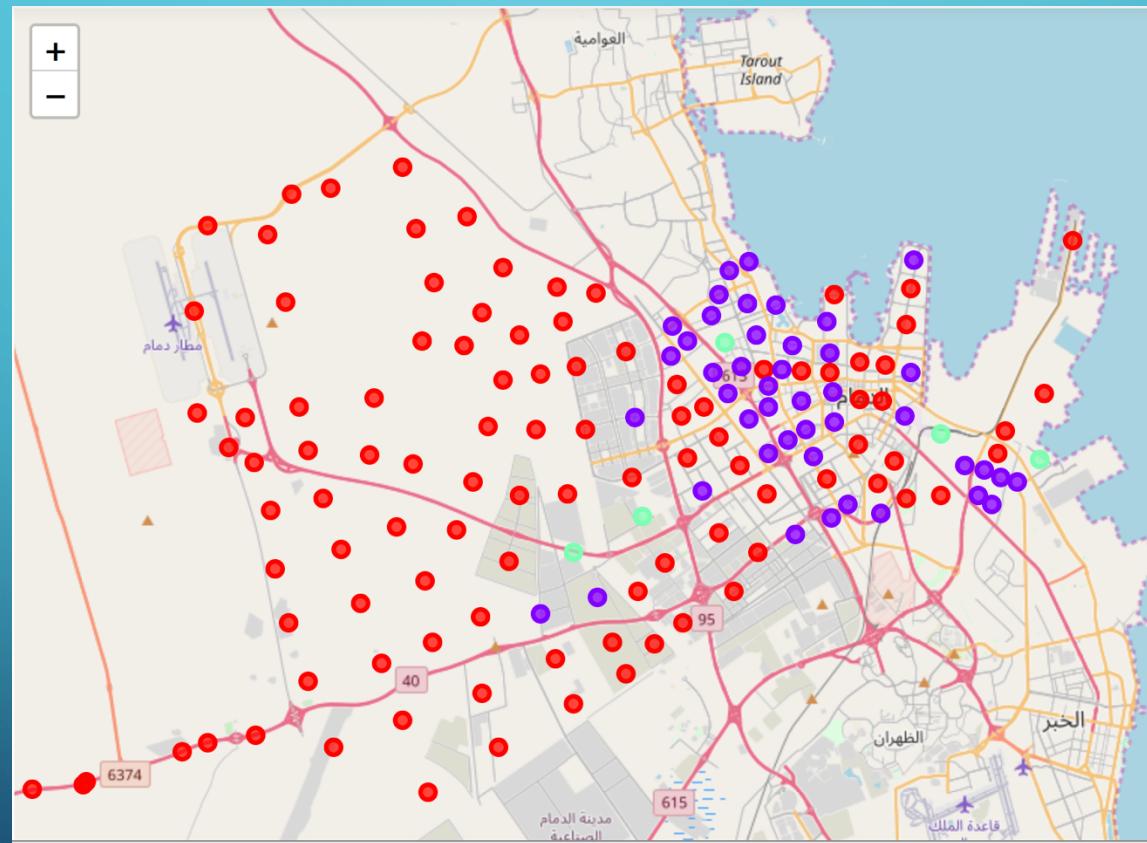
METHODOLOGY

- Using Foursquare API we can get information about all venues around each location to study
- 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.
- 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 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

RESULTS



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:
 - *Population of each location*
 - *Distance from city center*
 - *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.



THANK YOU