IBM Applied Data Science Capstone Report

Help you to determine your new business location in Riyadh, Saudi Arabia

Introduction

Riyadh city is the capital of Saudi Arabia which is full of business and exciting places such as fancy restaurants shopping mall est. Recently, Riyadh started attracting many tourists due to its international activities and events. This has led many businessmen to consider opening new businesses to increase their investments and attract more tourists.

Problem description

Any new businessmen who would like to open a new business place whatever the new business would be a restaurant, hotel, or movie theater. Some business either needs to check whether this area has similar business exist around or not to avoid duplicate. In this report, I took an example to help businessmen to open a new restaurant in Riyadh and where.

Data

To solve the problem, we will need the following data:

• List of neighborhoods of Riyadh. This defines the scope of this project which is confined to the city of Riyadh, the capital city of the country of Saudi Arabia.

• Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and to get the venue data.

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

For extracting previous data, I used this link (https://en.wikipedia.org/wiki/Riyadh). There are 97 neighborhoods. Then, I used beautiful soup package in Python for parsing data from HTML to Python lists. In the other hand, we used Foursquare API to provide all venue cross all neighborhoods were provided before. Foursquare data contains 293 categories such as coffee shops, Market, Art gallery est. Then, I used Python package called Folium to show maps. For Methodologies part, we used machine learning method K-mean clustering.

Methodology

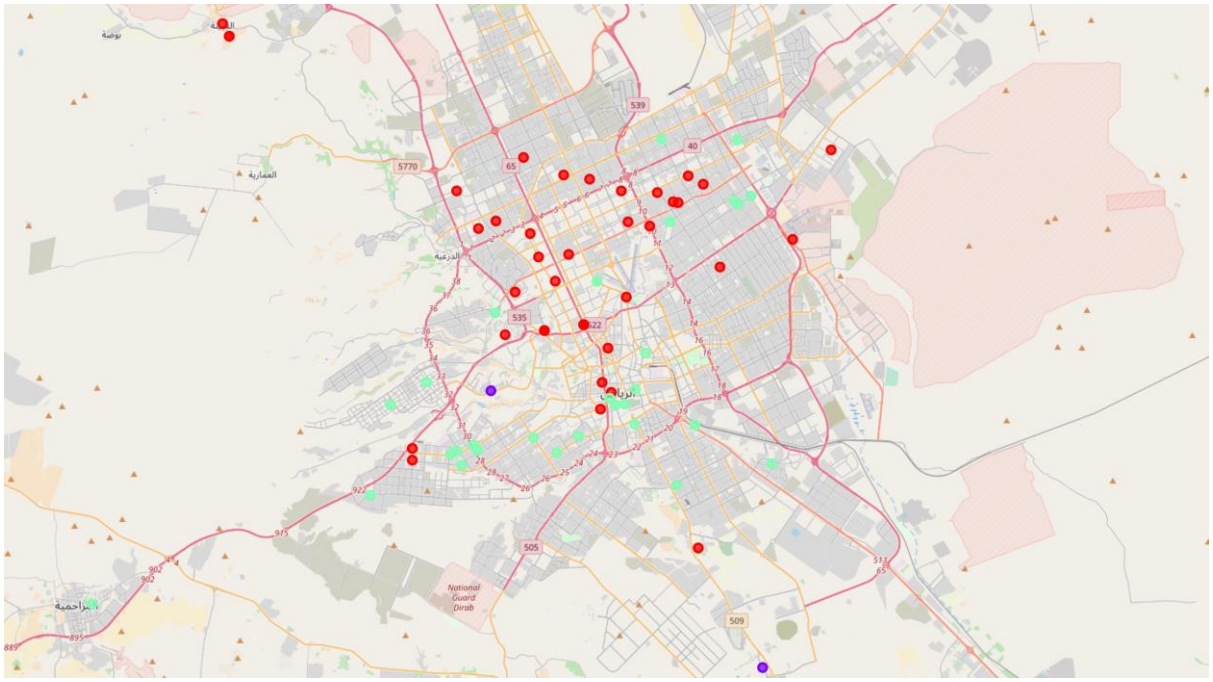
I will conclude the methodology in steps:

- Provide list of Riyadh neighborhood from (https://en.wikipedia.org/wiki/Riyadh) using beautiful soup package in Python.
- Provide list of geographical coordinates using Geocoder package in Python.
- Use Foursquare API to parse all venue in Riyadh by providing their latitude and longitude.
- Using Python Panda to generate a dataframe merging all the above.
- Applying Machine learning method (K-mean clustering). We will cluster neighborhoods into 6 clusters.

Result

The result was distributed by cluster as below:

- Cluster 0: high concentration of restaurants.
- Cluster 1: low concentration of restaurants.
- Cluster 2: High concentration of restaurants.
- Cluster 3: low concentration of restaurants.
- Cluster 4: low concentration of restaurants.



Discussion

In my opinion, I think if we do the with crowded neighborhood in our consideration. This factor will impact our result positively. Crowded neighborhood will attract more customer to our business>

Conclusion

As we can see most of categorize restaurants are distributed in the Riyadh downtown area. The highest number in cluster 0 then cluster 2. On the other hand, cluster 1,3,4 has very low number to totally no restaurants in the neighbourhoods. Due to that cluster 0 and cluster 2 represent a great opportunity and high potential areas to open new restaurants. From another perspective, if you are looking to open your business in a new area which is not crowded with other restaurants you can choose clusters 1,3 and 4.