Coursera Capstone Project

Analysis Study

Recommended Riyadh Districts for new Coffee Shop

Daliyah Aljamal

Dalyah_se@outlook.com

Date: January 2020

Table of Contents

Introduction 3
Data Description 3
Data Sources 3
Data Definition 4
Methodology 4
Prepare Riyadh district data 4
Getting Nearby Venues 5
Calculate how busy is each district 5
Calculate Average Café Rating per District 7
Compute the five best districts for a new café 8
Results 9
Discussion 10
Conclusion 10

Introduction

In Saudi Arabia, the coffee industry is expanding, especially in Riyadh which is the capital of the Kingdom. Coffee becomes a daily necessity for most of the people from various social levels. People are seeking a cozy place to have a relaxing time, meet friends or enjoy a family gathering. Also, it is a common behavior among students and workers to grape a cup of coffee from the nearest Café in the mornings. Saudi people like and enjoy the western style coffee shops making this type of business successful.

However, the right place for a café is crucial for business success. Some metrics need to be considered to choose the right place to open the new café. I considered these criteria for this analysis study:

1. Café has to be in Riyadh
2. Café should be in a crowded area (number of population, number of entertainment places)
3. Café should be in an area with low café average rating
4. Café should be in an area with few coffee shops around.

The criteria could be customized based on the café owner's target, budget, and preference.

In this study, I gathered the necessary data from multiple sources to achieve the study's objective. After data cleansing and preprocessing, I did explore and perform some analysis on the data to come up with the best five Riyadh neighborhoods for a new café.

The project will be interesting for anyone interested to open a coffee shop in Riyadh. It will provide the business owner with the top five recommended districts to open the café. Moreover, the project will be helpful for people interested in opening a restaurant having similar criteria to the ones studied here.

Data Description

Data Sources

The data considered in this study comes from these sources.

1. Riyadh district dataset. I found an interesting dataset in GitHub: *https://github.com/homaily/Saudi-Arabia-Regions-Cities-and-Districts.*

   that includes all the districts in Saudi Arabia along with their boundaries coordinates. After filtering and data cleansing, I ended up with riydah_districts dataset which includes Riyadh neighborhoods and their latitude and longitude.

2. Popular Venues details (Categories, ratings and more) from Foursquare API
3. Geopy Python Library to get locations' coordinates for Map visualization.

Data Definition

After the data preprocessing, cleansing and grouping, the data should be similar to the table below filled with dummy data.

District	Latitude	Longitude	Café's avg rating	Number of Café	Number of entertainment places
Alhamra	43.54	73.45	5	10	14

Methodology

Prepare Riyadh district data

The first step was to prepare the Riyadh District dataset. The datasets I got from the GitHub source include all KSA regions, cities and districts with all boundaries geo data in JSON format. I filtered the datasets to extract only Riyadh city districts with their boundaries geographical coordinates. Then, I calculated the median coordinates for each district. I ended up with *riyadh_districts.csv* file. Refer to Fig 1 for a snapshot of the data.

	district	latitude	longitude
0	Al Amal Dist.	24.645703	46.719765
1	Al Namudhajiyah Dist.	24.650184	46.698655
2	Al Jarradiyah Dist.	24.618163	46.694832
3	Al Sinaiyah Dist.	24.648146	46.737433
4	Manfuha Al Jadidah Dist.	24.614977	46.714758

Figure 1: Snapshot of riyadh_district Data frame

Moreover, the Riyadh districts are visualized on the map as shown in Fig[2].

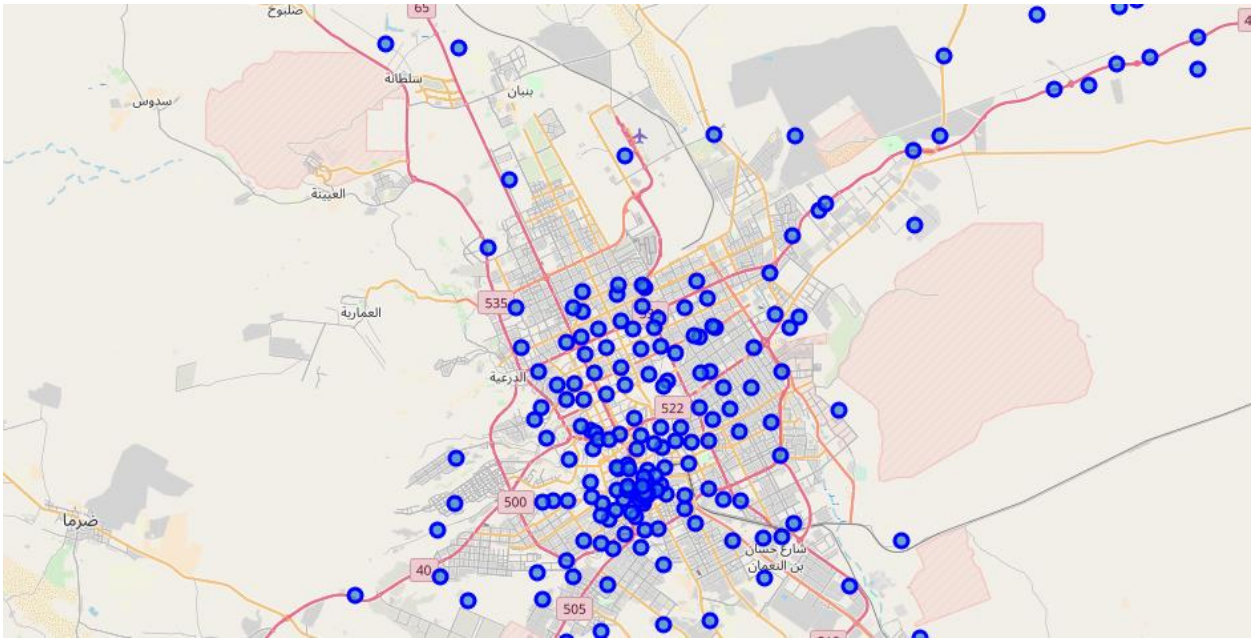

Figure 2: Riyadh Districts Map

Getting Nearby Venues

Using Foursquare API, I got the nearby venues up to 100 venues around each district in a radius of 500 meters using the *explore* endpoint. Getting the nearby venues gives insights about each district and helps for further analysis and calculations. A snapshot of the data after the Foursquare API calls is shown in Fig 3.

```
riyadh_venues.head()
```

	District	District Latitude	District Longitude	Venue Id	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Al Amal Dist.	24.645703	46.719765	4d53844765586ea8efb46cc8	Jollibee (جوليبي)	24.645733	46.717801	Filipino Restaurant
1	Al Amal Dist.	24.645703	46.719765	4dd9493d7d8b3226641ca01d	CITY FOOD COURT	24.645615	46.717704	Fast Food Restaurant
2	Al Amal Dist.	24.645703	46.719765	4ff2a82de4b055dd1c18642d	Al-Garaawi	24.642024	46.717322	Shoe Store
3	Al Amal Dist.	24.645703	46.719765	5abbd263037be133288214d6	Podium	24.646380	46.715550	Boutique
4	Al Amal Dist.	24.645703	46.719765	5145f199e4b016faed62f9e4	Cusina	24.646406	46.716060	Asian Restaurant

Figure 3: Snapshot of Riyadh Venues

Calculate how busy is each district

In this analysis, I considered the district to be busy or crowded looking at three metrics:

1. The number of Restaurants
2. The number of Cafés
3. The number of Stores or Shops

After calculating the required metrics, I got the data frame shown in Fig 4.

	District	Latitude	Longitude	Venue Count	Cafe Count	Restaurant Count	Store Count
0	Al Amal Dist.	24.645703	46.719765	37	1	11	9
1	Al Namudhajiyah Dist.	24.650184	46.698655	8	2	4	4
2	Al Jarradiyah Dist.	24.618163	46.694832	5	2	0	0
3	Al Sinaiyah Dist.	24.648146	46.737433	4	2	1	0
4	Manfuha Al Jadidah Dist.	24.614977	46.714758	4	0	2	0

Figure 4: How Busy is each District?

To judge if a certain district is busy, I calculated *howBusy* feature which is the sum of *Café Count*, *Restaurant Count* and *Store Count*. The bar chart in Fig 5 shows the 20 busiest districts in Riyadh that were considered in the further analysis of this study.

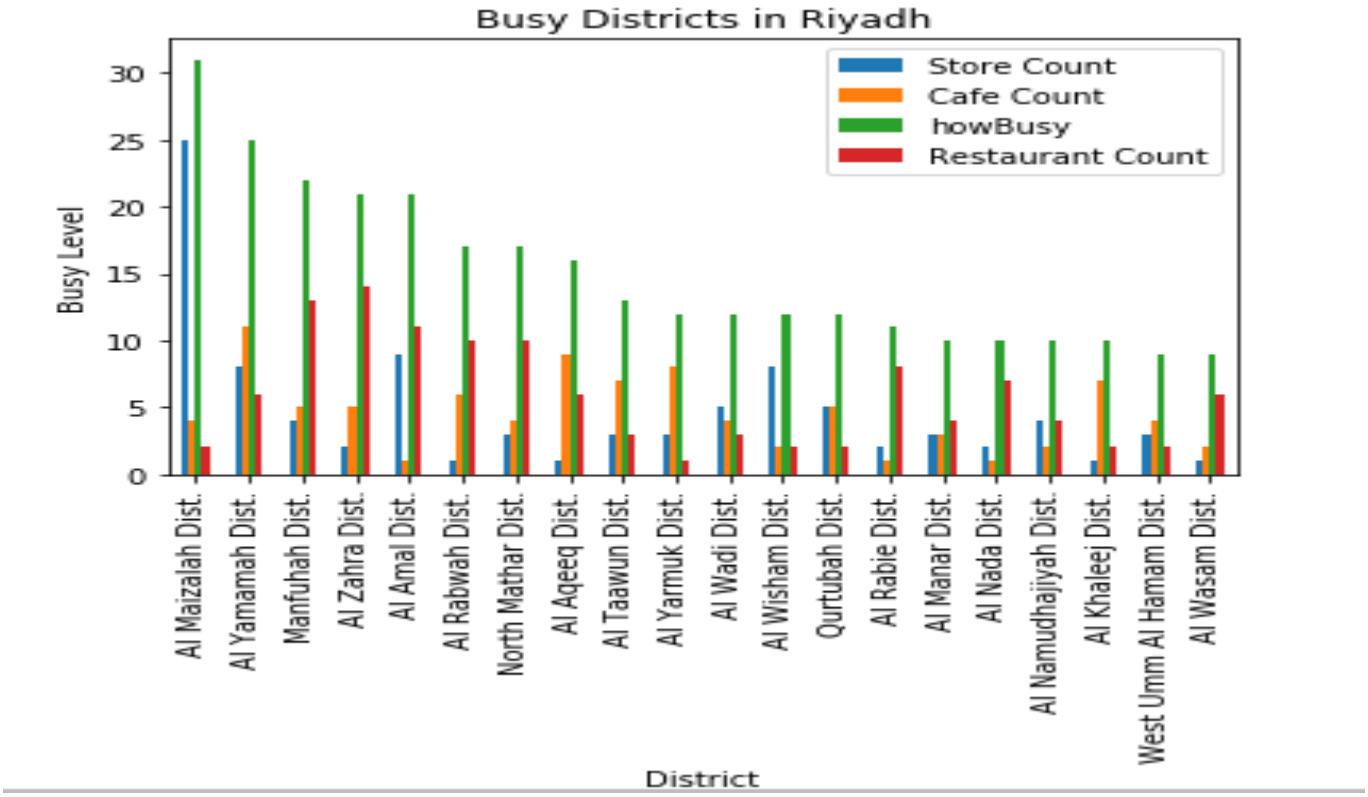

Figure 5: the 20 Busiest Districts in Riyadh

Calculate Average Café Rating per District

To consider having the new café in a district with low average café rating, Foursquare API premium calls were made using *venues* endpoint to get the rating for each café in Riyadh. First, I filtered riyadh_venues data frame to include only Café and coffee shop venues as shown in Fig 6.

	District	District Latitude	District Longitude	Venue Id	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue Rating	Venue Price
0	Al Namudhajiyah Dist.	24.650184	46.698655	573c5330cd10c16feed7f972	Java cafe Al Washm (جافا كافيه الوشم)	24.650939	46.703281	Coffee Shop	7.1	1
1	Al Sinaiyah Dist.	24.648146	46.737433	5719f5fe498ecff1fb63b781	Starbucks (ستاربكس)	24.650643	46.740358	Coffee Shop	8.0	1
2	Al Dirah Dist.	24.631157	46.714473	5c6976a17c891c003a4d16f2	Elixir Bunn Coffee Roasters (محمصة اكسير البن)	24.630936	46.712279	Coffee Shop	8.8	1
3	Al Dirah Dist.	24.631157	46.714473	5294ec4011d2ce5d6a3373ba	Bread And Coffee (رغيف وكيف)	24.630363	46.711909	Coffee Shop	8.4	1
4	Al Dirah Dist.	24.631157	46.714473	5daa086c6bfe6b00080bd40a	(Zad) زاد	24.630958	46.712257	Coffee Shop	8.4	1

Figure 6: Coffee Shops in Riyadh

Then, I made Foursquare API premium calls to get the rating for each café. Finally, the average café ratings per district were computed. Fig 7 shows a snapshot of the result.

District	Venue Rating
Al Amal Dist.	7.8
Al Andalus Dist.	8.0
Al Aqeeq Dist.	6.3
Al Awaly Dist.	8.1
Al Badeah Dist.	7.9

Figure 7: Café Rating per District

Compute the five best districts for a new café

Now, after the data has been gathered, grouped and calculated, there exist three features for each district (*Howbusy*, *Café Count* and *Average Café Rating*) to decide where to locate a new café. Based on the requirements stated earlier, the café district should be busy, have few coffee shops around for less competition and low average café rating. To best achieve these requirements, I used a scoring method in which I assigned a specific scoring rate for each feature, as follows:

1. How Busy feature: **50%**
2. Café Count: **30%**
3. Average Cafe Rating: **20%**

Note: The scoring rates could be adjusted based on the requirements and the target of the new coffee shop owner.

Before computing the score for each district, I normalized all the three features to have them all in a standard scale. Then, I computed the score using this formula:

$$df['howBusy'] * 0.5 + df['Cafe\ Count'] * -0.3 + df['Avg\ Cafe\ Rating'] * -0.2$$

The results were stored for each district and sorted by *Score* in descending order to give the resulted data frame as shown in Fig 8.

	District	howBusy	Avg Cafe Rating	Cafe Count	Score	Latitude	Longitude
0	Al Maizalah Dist.	2.677848	-1.797147	-0.201538	1.758815	24.791527	46.850795
1	Al Amal Dist.	1.004193	0.581558	-1.300838	0.776036	24.645703	46.719765
2	Manfuhah Dist.	1.171558	-0.739945	0.164895	0.684300	24.591191	46.730304
3	Al Yamamah Dist.	1.673655	-1.268546	2.363494	0.381488	24.589865	46.716110
4	Al Wisham Dist.	-0.502096	-1.114371	-0.934405	0.252147	24.646955	46.700661

Figure 8: Districts Café Score

Results

After this deep analysis, five districts are recommended as they are the best for achieving the requirements of this study. The districts are sorted in order:

1. Al Maizalah District.
2. Al Amal District.
3. Manfuhah District.
4. Al Yamamah District.
5. Al Wisham District.

The recommended districts are visualized in Riyadh map in Fig 9.

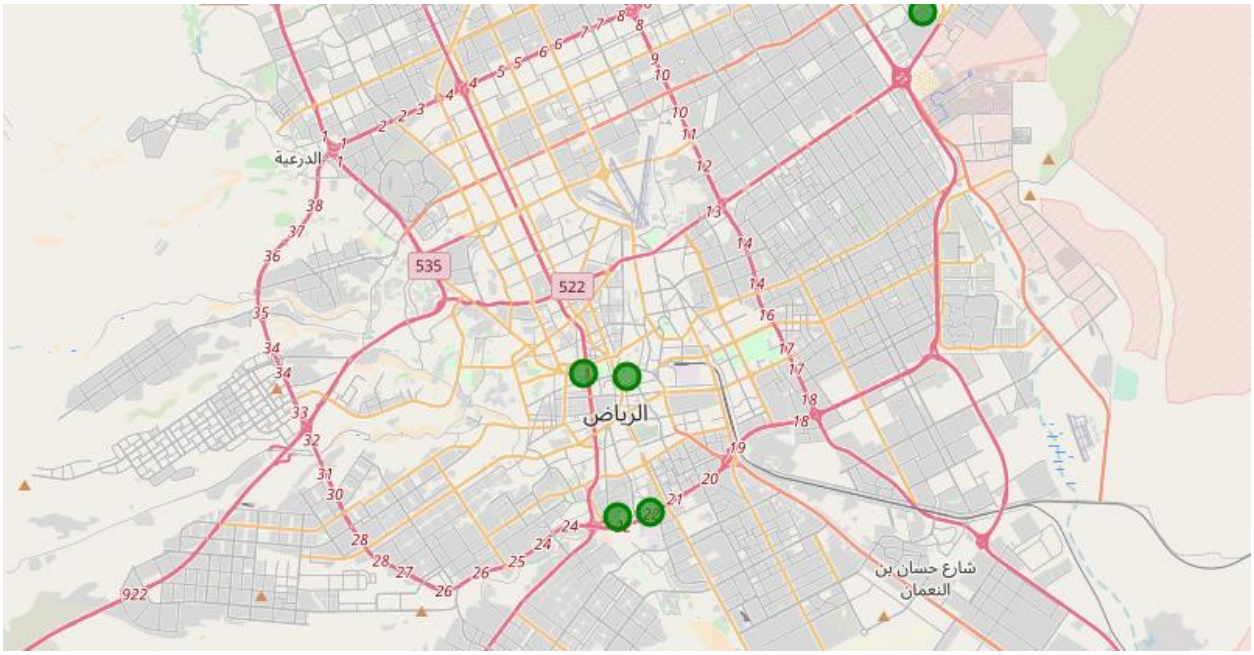

Figure 9: Best Riyadh Districts for new Café

The five recommended districts are shown in Fig 10 along with their scores.

	District	Score
0	Al Maizalah Dist.	1.758815
1	Al Amal Dist.	0.776036
2	Manfuhah Dist.	0.684300
3	Al Yamamah Dist.	0.381488
4	Al Wisham Dist.	0.252147

Figure 10: Recommended Districts for the new Café

Discussion

Looking at the results, it is clear that *Al Maizalah District* is recommended with a high score as a location for a new café in Riyadh. It is an interesting result; people usually don't think of this district when deciding where to locate their coffee shop or restaurant, although the district has huge potential for success. It is important to mention that these recommendations were made based on the requirements stated earlier and the rate or the level of importance for each. Various recommendations could result in different requirements or different requirement rates or scores.

Conclusion

Data is gold when it is used and analyzed in a useful way. This analysis shows how data helps in drawing a decision or a recommendation for the best places to locate a new café based on specific requirements from the business owner. This Data-driven analysis will act as a guide for any business owner looking for a place for a new café or restaurant with the same or similar requirements. Further analysis could be performed with more data, different criteria or perhaps different methodologies.

Thanks for your interest in this analysis!