APPLIED DATA SCIENCE CAPSTONE

Battle of The Neighborhoods

PROJECT REPORT

Fareed Abdul Hameed
27-JUL-2021

Table of Contents

1. INTRODUCTION

1.1. Background

Riyadh, the heart of the Kingdom of Saudi Arabia has seen exponential growth in the past few years. The government has been injecting massive funds into the city's development with state-of-the-art infrastructure and technological transformation throughout. It is the business hub of the country attracting investors, tourists and jobseekers alike; all looking for their opportunity to make their mark. With this immigration trend, Riyadh has become a diverse city, however, it is not clear how diverse the cuisines are within the city to cater to the diverse masses.

1.2. Problem

Finding a decent restaurant can be a nightmare within the 1972 km2 area of the city. Fast-food and various restaurants are always available in any high street; however, you can never be sure about the cuisine they have to offer and the overall quality or experience to expect unless some research is done over the web.

Hence, the aim of this project is to cluster Riyadh giving a brief idea on cuisine distribution across the city and understanding the cuisine diversity.

1.3. Interest

This project will be of interest mainly for the average foodie looking to find the right district to experience a certain cuisine. It will also help people like me who have recently moved into the city and are unaware of the cuisines around my area. It will also help those considering moving into the city take an informed decision with relevance to availability of restaurants in a specific area.

2. DATA

2.1. Data Sources

FOURSQUARE

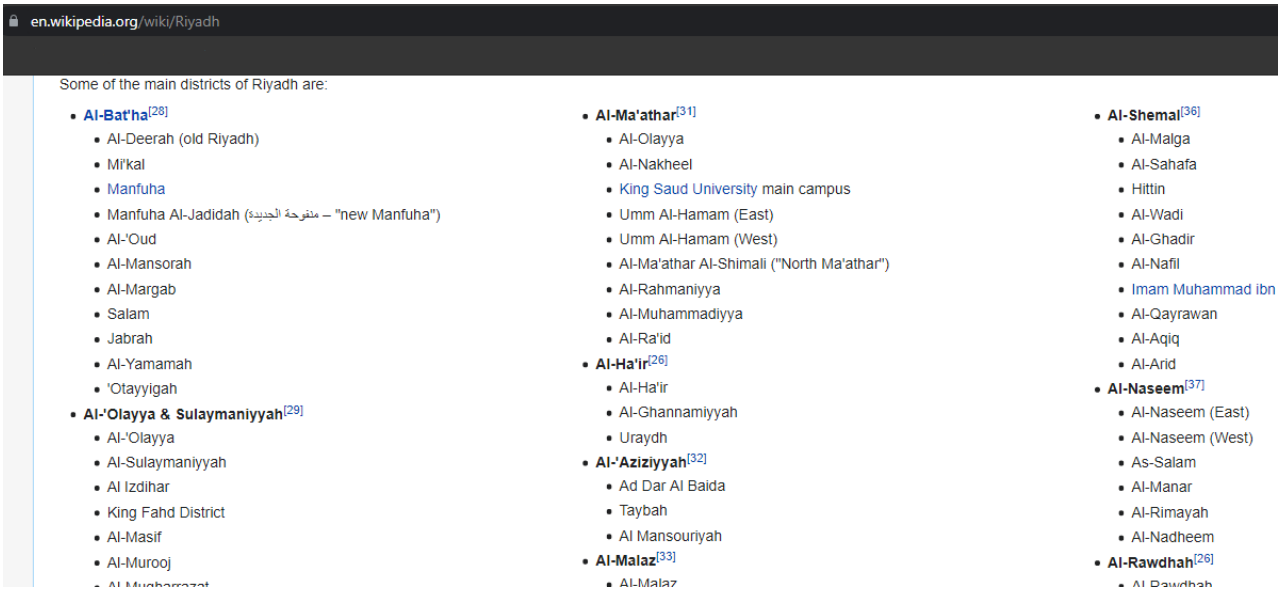
Foursquare is a US tech company from New York focusing on location data. Their technology and data powers apps such as Apple's Maps, Uber, Twitter and many other household names. Hence, will leverage the foursquare API to get the data pertaining to the restaurant name, ID, location and category of food (vegetarian, Italian etc.).

https://developer.foursquare.com/

WIKIPEDIA

In order to divide the city and get the relevant clusters, location data is key, and Wikipedia has this documented to some extent. As Riyadh is a massive city, it will be a massive project to cater to all 130+ districts especially given the lack of data sources (even the districts, coordinates etc.). Hence to make this project feasible, we will only be using the 16 main municipal districts.

https://en.wikipedia.org/wiki/Riyadh



Source Data for Districts

GEOLOCATION

After exploring various APIs for geolocation such as Geopy, Positionstack and Mozilla, it was clear that these APIs were not capable to return the coordinates for the districts of Riyadh. Positionstack came very close, but it expected data in extremely precise way and still would randomly provide data of other countries instead.

Hence in order to mitigate the situation, the coordinates for the 16 districts were taken from Google Maps in a CSV format.

2.2. Data Cleansing

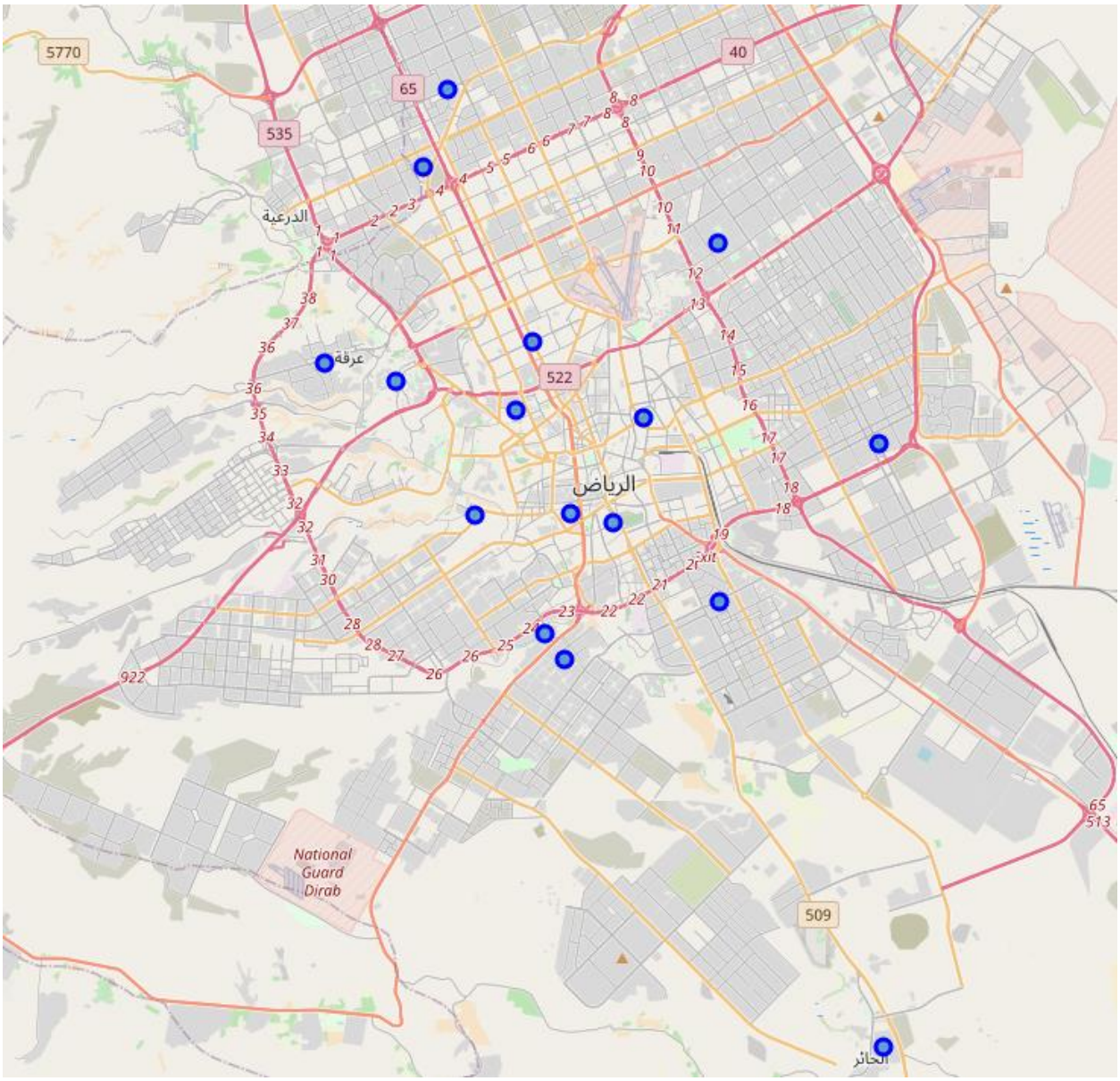
Extensive data cleansing was required wherein Wikipedia was exceptionally challenging. As Wikipedia is an open platform, anyone can modify the district names as they see fit making this project vulnerable to constant issues and making this data source as a weak source.

The other sources were pure with minimal data cleansing activities required with key focus on transforming data to make specific analysis.

3. Methodology

3.1. Plotting the districts of Riyadh

The first dataset was scraped off Wikipedia (Refer 2.1). This source gave the foundation master data. Next dataset was the Geolocation Coordinates in CSV Format. The merged data sets would provide as the source to allow plotting the centers of the main districts in the city of Riyadh. Folium package was used to generate the below.

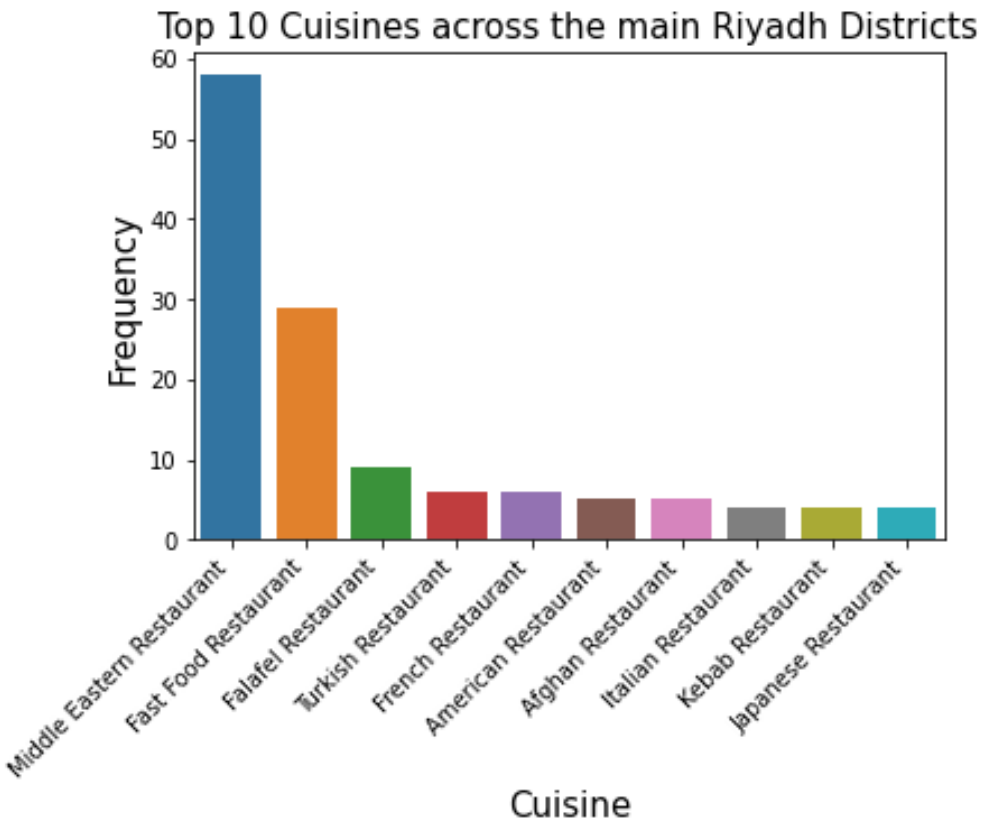


Riyadh Clustered by Main Districts

3.2. Restaurant Data from Foursquare API

The next step in the process was to fetch in the data from Foursquare. The assumption is that a person will be willing to travel within a 6km radius from the center of the main district allowing for a maximum 12km commute from the centroid of their district's cluster.

Keeping this in view, a function was built allowing to fetch data from the API as required. The data received provided the below details indicating that Middle Eastern Cuisine was the most prominent followed by fast food chains.



3.3. The Data Story

All the data required for clustering is now available within various data frames. Hence, the next step was to unify everything into a specific format which would aid in data analysis and readiness for machine learning unsupervised clustering algorithm. This required various transformations including merging data sets, performing one-hot encoding, grouping by districts and the cuisine type etc. During the process, a district AlHair was found to have no data from Foursquare. Looking int up on the map, it was interesting to see that this district was a few kilometers out of Riyadh and does not have restaurants. Hence the results below will no longer have that district.

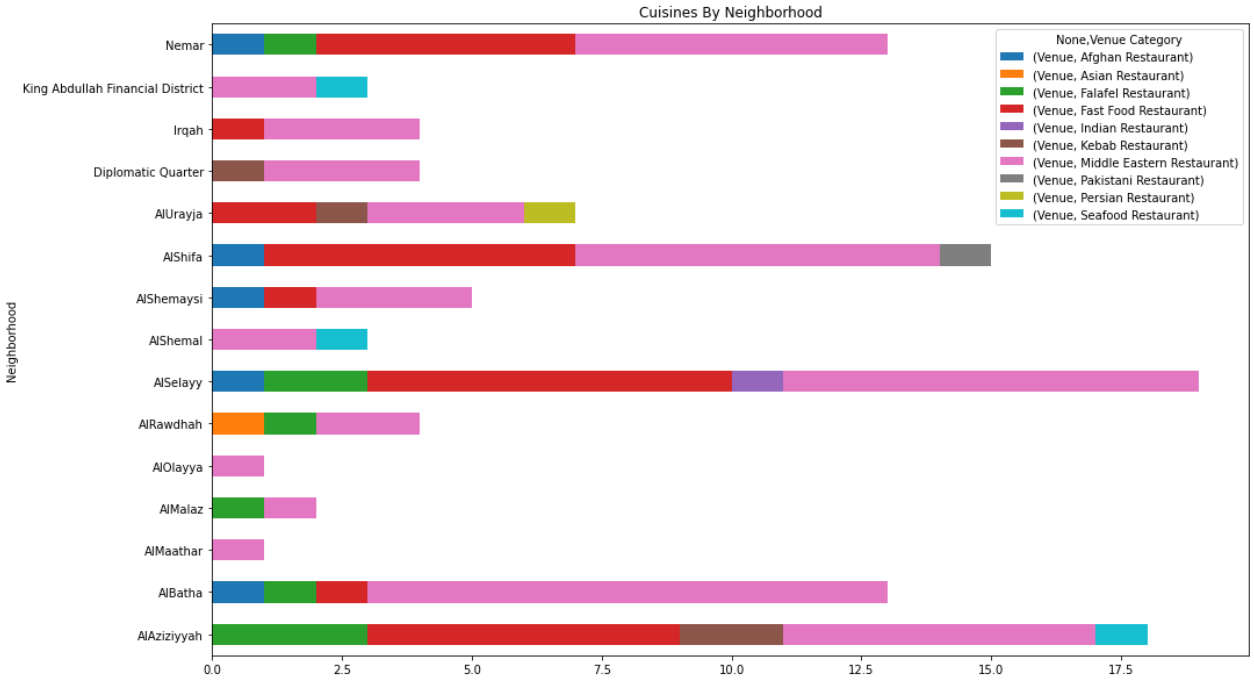
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	AlAziziyyah	Fast Food Restaurant	Middle Eastern Restaurant	Turkish Restaurant	Falafel Restaurant	Kebab Restaurant
1	AlBatha	Middle Eastern Restaurant	Afghan Restaurant	Arepa Restaurant	Falafel Restaurant	Fast Food Restaurant
2	AlMaathar	Middle Eastern Restaurant	Eastern European Restaurant	French Restaurant	Yemeni Restaurant	Indian Restaurant
3	AlMalaz	Middle Eastern Restaurant	Falafel Restaurant	Yemeni Restaurant	Indian Restaurant	American Restaurant
4	AlOlayya	Italian Restaurant	Armenian Restaurant	Middle Eastern Restaurant	Eastern European Restaurant	French Restaurant
5	AlRawdhah	Middle Eastern Restaurant	Asian Restaurant	Eastern European Restaurant	Falafel Restaurant	Japanese Restaurant
6	AlSelayy	Middle Eastern Restaurant	Fast Food Restaurant	Falafel Restaurant	Afghan Restaurant	Turkish Restaurant
7	AlShemal	Middle Eastern Restaurant	Japanese Restaurant	Seafood Restaurant	Yemeni Restaurant	French Restaurant
8	AlShemaysi	Middle Eastern Restaurant	Afghan Restaurant	Sushi Restaurant	Fast Food Restaurant	French Restaurant
9	AlShifa	Middle Eastern Restaurant	Fast Food Restaurant	Afghan Restaurant	Pakistani Restaurant	French Restaurant
10	AlUrayja	Middle Eastern Restaurant	Fast Food Restaurant	American Restaurant	Italian Restaurant	Sushi Restaurant
11	Diplomatic Quarter	Middle Eastern Restaurant	French Restaurant	Sushi Restaurant	Kebab Restaurant	Yemeni Restaurant
12	Irqah	Middle Eastern Restaurant	Italian Restaurant	French Restaurant	Mexican Restaurant	Fast Food Restaurant
13	King Abdullah Financial District	French Restaurant	Middle Eastern Restaurant	American Restaurant	Seafood Restaurant	Japanese Restaurant
14	Nemar	Middle Eastern Restaurant	Fast Food Restaurant	American Restaurant	Yemeni Restaurant	Falafel Restaurant

Top 10 prominent cuisines by District

Visualization of districts and cuisines

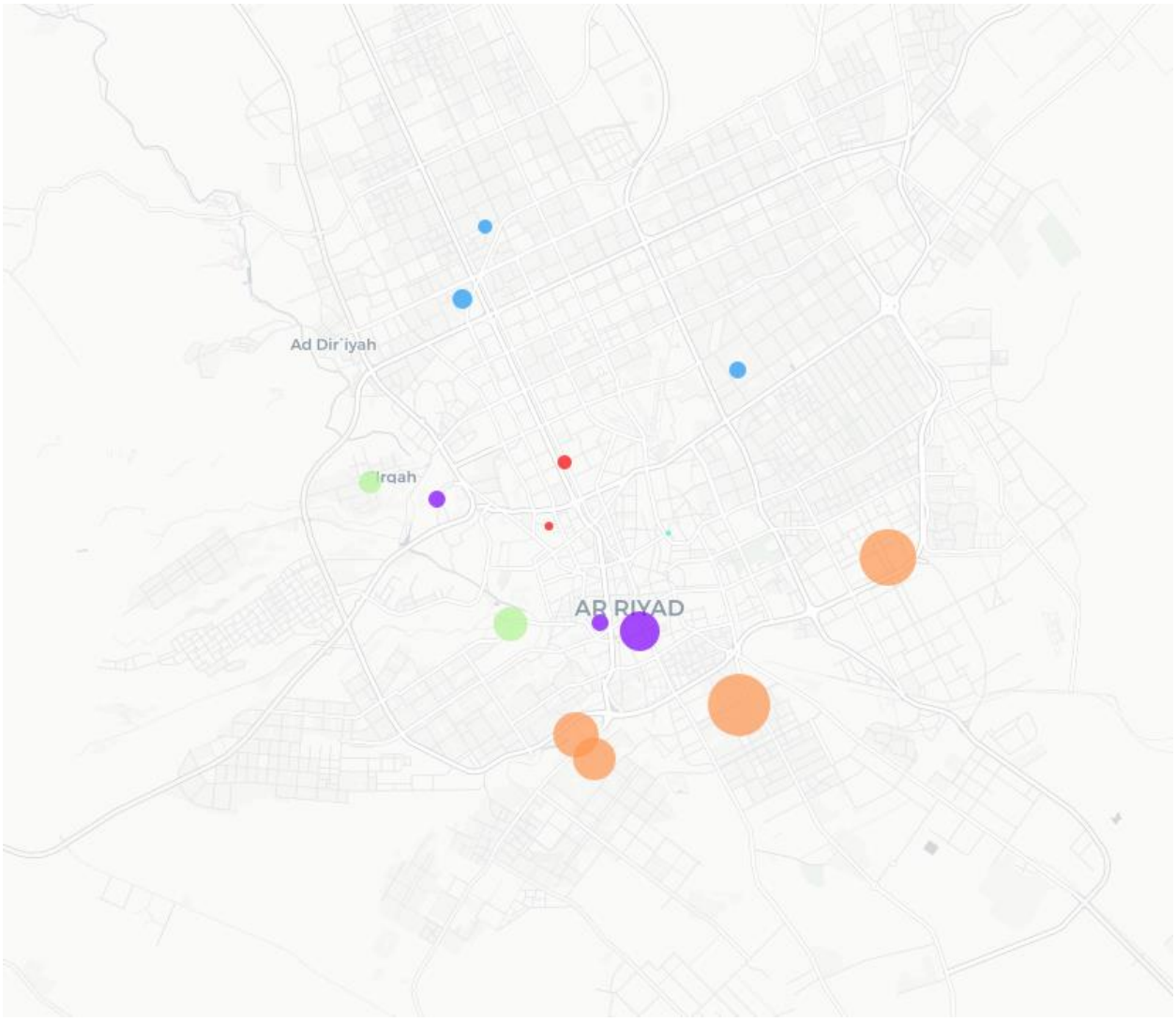
It seemed quite interesting to see the breakdown of the cuisines associated with the short-listed top 10 cuisines across the city. Is Middle Eastern food the sole winner across all the districts? Or do we have some cuisines dominating certain districts?

Hence to answer these questions, a few data transformations was all that it took to build a horizontal bar plot that aided with the questions.



Machine Learning Unsupervised (k-means clustering)

The last step within the project was to plug in the k-means clustering, an unsupervised machine learning algorithm. Dividing the data into 6 clusters (k=6), the below cluster map was formed.



The clusters outputs are provided below aiding in understanding the cuisine distribution across the city. They are in descending order based on their prominence.

CLUSTER ORANGE – Middle Eastern and Fast Food Cuisine

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2	Nemar	Middle Eastern Restaurant	Fast Food Restaurant	American Restaurant	Yemeni Restaurant	Falafel Restaurant
8	AlAziziyyah	Fast Food Restaurant	Middle Eastern Restaurant	Turkish Restaurant	Falafel Restaurant	Kebab Restaurant
10	AlShifa	Middle Eastern Restaurant	Fast Food Restaurant	Afghan Restaurant	Pakistani Restaurant	French Restaurant
14	AlSelayy	Middle Eastern Restaurant	Fast Food Restaurant	Falafel Restaurant	Afghan Restaurant	Turkish Restaurant

CLUSTER PURPLE – Middle Eastern and Asian Minority Cuisine

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	AlBatha	Middle Eastern Restaurant	Afghan Restaurant	Arepa Restaurant	Falafel Restaurant	Fast Food Restaurant
4	Diplomatic Quarter	Middle Eastern Restaurant	French Restaurant	Sushi Restaurant	Kebab Restaurant	Yemeni Restaurant
5	AlShemaysi	Middle Eastern Restaurant	Afghan Restaurant	Sushi Restaurant	Fast Food Restaurant	French Restaurant

CLUSTER GREEN – Middle Eastern and European Cuisine

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
3	Irqah	Middle Eastern Restaurant	Italian Restaurant	French Restaurant	Mexican Restaurant	Fast Food Restaurant
11	AlUrayja	Middle Eastern Restaurant	Fast Food Restaurant	American Restaurant	Italian Restaurant	Sushi Restaurant

CLUSTER BLUE – Middle Eastern and Asian Cuisine

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
12	AlShemal	Middle Eastern Restaurant	Japanese Restaurant	Seafood Restaurant	Yemeni Restaurant	French Restaurant
13	AlRawdhah	Middle Eastern Restaurant	Asian Restaurant	Eastern European Restaurant	Falafel Restaurant	Japanese Restaurant
15	King Abdullah Financial District	French Restaurant	Middle Eastern Restaurant	American Restaurant	Seafood Restaurant	Japanese Restaurant

CLUSTER RED – European Cuisine

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	AlOlayya	Italian Restaurant	Armenian Restaurant	Middle Eastern Restaurant	Eastern European Restaurant	French Restaurant
6	AlMaathar	Middle Eastern Restaurant	Eastern European Restaurant	French Restaurant	Yemeni Restaurant	Indian Restaurant

CLUSTER AQUA – Middle Eastern Cuisine

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
9	AlMalaz	Middle Eastern Restaurant	Falafel Restaurant	Yemeni Restaurant	Indian Restaurant	American Restaurant

4. Results

Based on the project, the following can be deduced:

1. The most prominent restaurants in Riyadh are Middle Eastern and Fast-Food Joints.
2. The above statement has negligible change across all the main districts evaluated.

5. Conclusion

With this, we can meet the initial objectives of understanding the restaurant distribution across the main districts of the city. However, this analysis can be more powerful if the data sources for population distribution, spending power, all districts and their coordinates, top nationality in those districts etc. were available as this will allow further strength into the project such as:

1. Convert this project into a recommendation engine for potential investors looking to open a restaurant in the city based on:
    a. Spending Power (Will be able to determine the standard required)
    b. Top Nationality (Cuisine advisor)

Moreover, this analysis just opened a door or new questions such as:

1. The district of AlMalaz is predominantly a hotspot for Indian and Pakistani communities. As such the expectation is to have Indian or Pakistani cuisine in the top 3. This scenario could be either:
    a. Lack of data from foursquare.
    b. A possible investment goldmine for Indian and Pakistani restaurant franchises and brands.
2. The district AlOlaya returned very few results overall. This scenario could most likely point to lack of data.

In addition, the analysis with 16 districts was strong enough to bring light to the opportunities in the city, doing for all 130 districts will enable us to insights that will open further questions and massive opportunities.