EXPLORING RIYADH CITY NEIGHBORHOODS

HICHEM Dridi

4/5/2021

IBM Developer

SKILLS NETWORK

OUTLINE

- Executive Summary
- Introduction
- Data Acquisition
- Methodology
- Results
  - The city's map
  - Visualizations
- Results Discussion
  - Findings & Recommandations
- Conclusion
- Appendix

EXECUTIVE SUMMARY

Data was collected about the city's neighborhoods

Using Foursquare API, relevant information was obtained about venues in each neighborhood

Computing the frequency of each venue category in every neighborhood helped to identify the list of ten most common venues

Passing the frequency of venues to the DBSCAN clustering algorithm results in grouping neighborhoods based on their similarities

Examining the most common venues in each cluster gives insights about the characteristics of the city

INTRODUCTION

Riyadh city :

- Main Financial hub of the country
- Largest city
- 7.6 M habitants :
  - Most populos city in the country
  - 3rd Most populous in Middle East
  - 38th Most populous in Asia
- 5 M tourists per year :
  - 49th Most visited city in the world
  - 6th in the Middle East

Objective :

Explore the city :

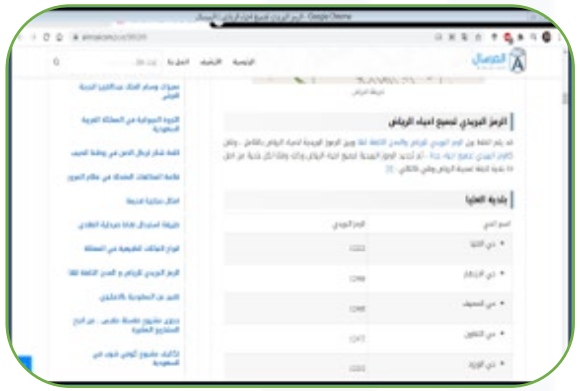
- Identify the most common venues
- Group the neighborhoods based on the similarities of venues

Give insights about the city to :

- Futur visitors and habitants
- Investors

DATA ACQUISITION

Web page



Web scrapping :

- *Selenium*
- *BeautifulSoup*

Translation :

- *Deep-translator*

✓ *185 Neighborhood*

Data Cleaning

Main Features :

- ✓ Municipality name
- ✓ Neighborhood name
- ✓ Postal code

Geocoder API

Neighborhoods' geo-coordination

- Latitude
- Longitude

IBM Developer

SKILLS NETWORK

DATA ACQUISITION

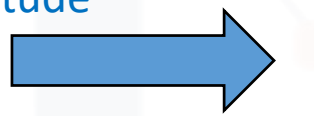
Neighborhoods data

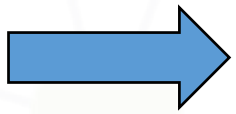
	Municipality_ar	Neighborhood_ar	Postal_code	Municipality_en	Neighborhood_en	Latitude	Longitude
0	بلدية العليا	حي العليا	12222	Olaya	Al Alia	24.69502	46.69004
1	بلدية العليا	حي الازدهار	12486	Olaya	Al-Izdihar	24.77904	46.72446
2	بلدية العليا	حي المصيف	12466	Olaya	Al-Masif	24.76386	46.68849
3	بلدية العليا	حي التعاون	12475	Olaya	Al-Taawon	24.76724	46.69640
4	بلدية العليا	حي الورود	12252	Olaya	Al-Worood	24.72213	46.68584

METHODOLOGY

Data Preprocessing

- ✓ Account credentials
- ✓ Latitude & longitude
- ✓ Radius
- ✓ Limit





- ✓ Venue name
- ✓ Venue geo-coordination
- ✓ Venue category
- ✓ Etc.

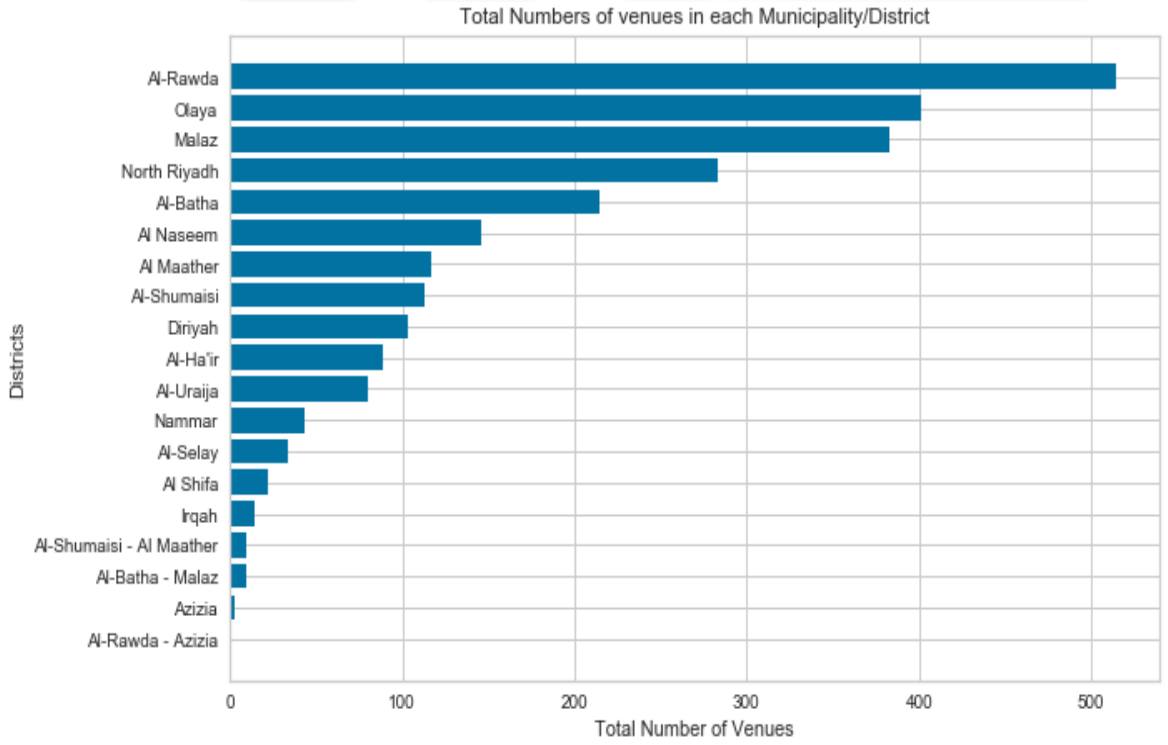
METHODOLOGY

Data Preprocessing Results

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Al Alia	24.69502	46.69004	Paskalia (باسكاليا)	24.697674	46.687086	Restaurant
1	Al Alia	24.69502	46.69004	Cacti Cafe	24.698523	46.688123	Coffee Shop
2	Al Alia	24.69502	46.69004	Starbucks (ستاربكس)	24.699015	46.689939	Coffee Shop
3	Al Alia	24.69502	46.69004	Tutti Cafè (أنواع القهوة)	24.697852	46.687181	Café
4	Al Alia	24.69502	46.69004	Adidas Originals	24.698398	46.687715	Sporting Goods Shop
5	Al Alia	24.69502	46.69004	Tim Hortons (تيم هورتنز)	24.698790	46.689771	Coffee Shop
6	Al Alia	24.69502	46.69004	Le Gourmet (الذواق)	24.691527	46.691695	Bakery
7	Al Alia	24.69502	46.69004	Zara Man (زارا مان)	24.698963	46.691498	Clothing Store
8	Al Alia	24.69502	46.69004	Fitness First Ladies	24.698749	46.691229	Gym / Fitness Center
9	Al Alia	24.69502	46.69004	Labouchee (لابوشيه)	24.691923	46.692852	Cupcake Shop
10	Al Alia	24.69502	46.69004	الرماح للأقمشة الرجالية	24.690697	46.689549	Men's Store
11	Al Alia	24.69502	46.69004	Dunkin' (دانكن)	24.698873	46.689581	Donut Shop
12	Al Alia	24.69502	46.69004	Gardenia	24.698984	46.689034	Flower Shop
13	Al Alia	24.69502	46.69004	PAUL (بول)	24.697306	46.686645	French Restaurant
14	Al Alia	24.69502	46.69004	Fire Grill (فاير جريل)	24.698120	46.687036	Mexican Restaurant

METHODOLOGY

Initial Analysis

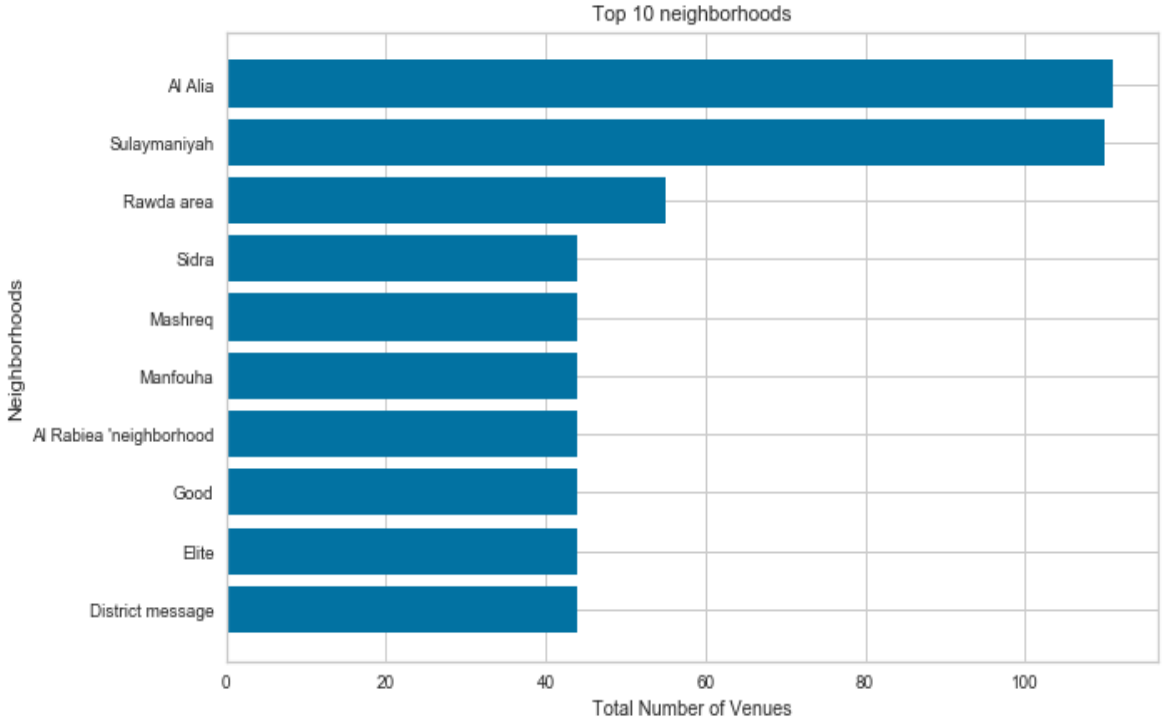


We can divide municipalities to three sections:

- Municipalities with crowded neighborhoods (more than 200 venues).
- Medium size municipalities: more than 80 and less than 200 venues.
- Small districts with less than 50 venues.

METHODOLOGY

Initial Analysis



- *'Al Alia'* and *'Sulaymaniyah'* differ a lot from the rest. They have more than 100 venues each one while the rest have less than 60.
- These two neighborhoods are very crowded with habitants and workers.

METHODOLOGY

Compute The Venues' Frequency In Each Neighborhoods

	Neighborhood	ATM	Accessories Store	Afghan Restaurant	African Restaurant	Airport Lounge	Airport Terminal	American Restaurant	Amphitheater	Antique Shop	...	Used Bookstore	Vacation Rental	Vegetarian / Vegan Restaurant	Vide Gam Stor
0	Ad-Durahimiyah	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.
1	Al Alia	0.0	0.0	0.0	0.0	0.0	0.0	0.009174	0.0	0.0	...	0.0	0.0	0.0	0.
2	Al Aqiq	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.
3	Al Faisaliah	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.
4	Al Fakhriyah	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.

5 rows × 240 columns

This table data will be used in the clustering exercise

METHODOLOGY

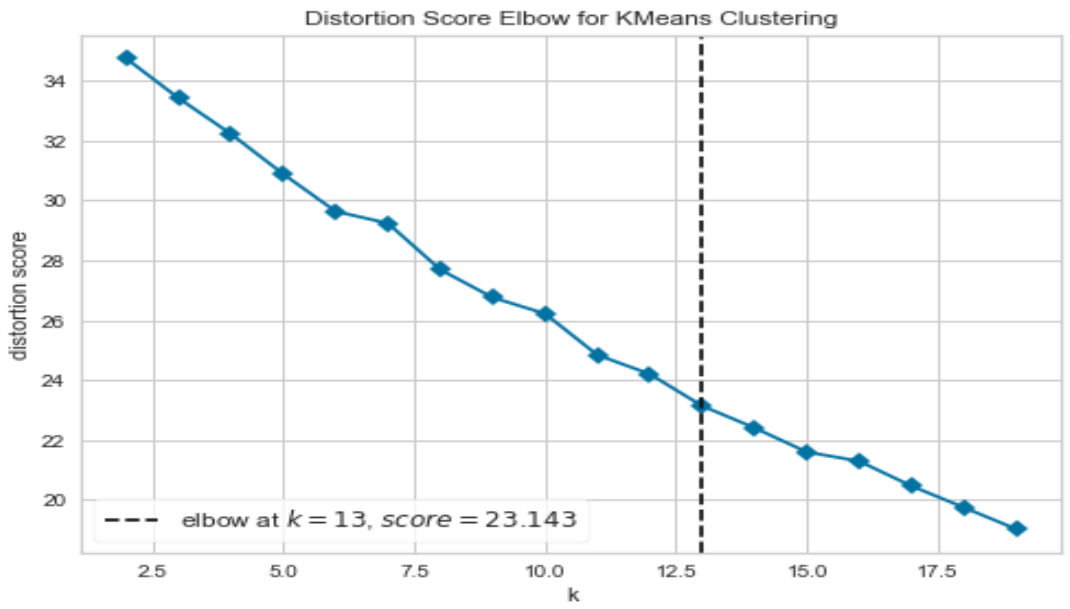
Determine The 10 Most Common Venues For Each Neighborhood

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Ad-Durahimiyah	Food Truck	Sandwich Place	Arts & Crafts Store	Yoga Studio	Film Studio	Filipino Restaurant	Fast Food Restaurant	Farm	Falafel Restaurant	Fabric Shop
1	Al Alia	Coffee Shop	Café	Cupcake Shop	Dessert Shop	Sporting Goods Shop	Men's Store	Clothing Store	Donut Shop	Bakery	Ice Cream Shop
2	Al Aqiq	Juice Bar	Mobile Phone Shop	Breakfast Spot	Fast Food Restaurant	Sporting Goods Shop	Burger Joint	Food Truck	Cafeteria	Café	Restaurant
3	Al Faisaliah	Coffee Roaster	Department Store	Furniture / Home Store	Auto Garage	Yoga Studio	Electronics Store	Flea Market	Film Studio	Filipino Restaurant	Fast Food Restaurant
4	Al Fakhriyah	Hookah Bar	Asian Restaurant	Coffee Shop	Pedestrian Plaza	Campground	Sports Bar	Health Food Store	Health & Beauty Service	Fast Food Restaurant	Farm

METHODOLOGY

K-Means Clustering

- First choice was to use KMeans algorithm
- Using the Elbow Method to find the optimal value of K (number of clusters)

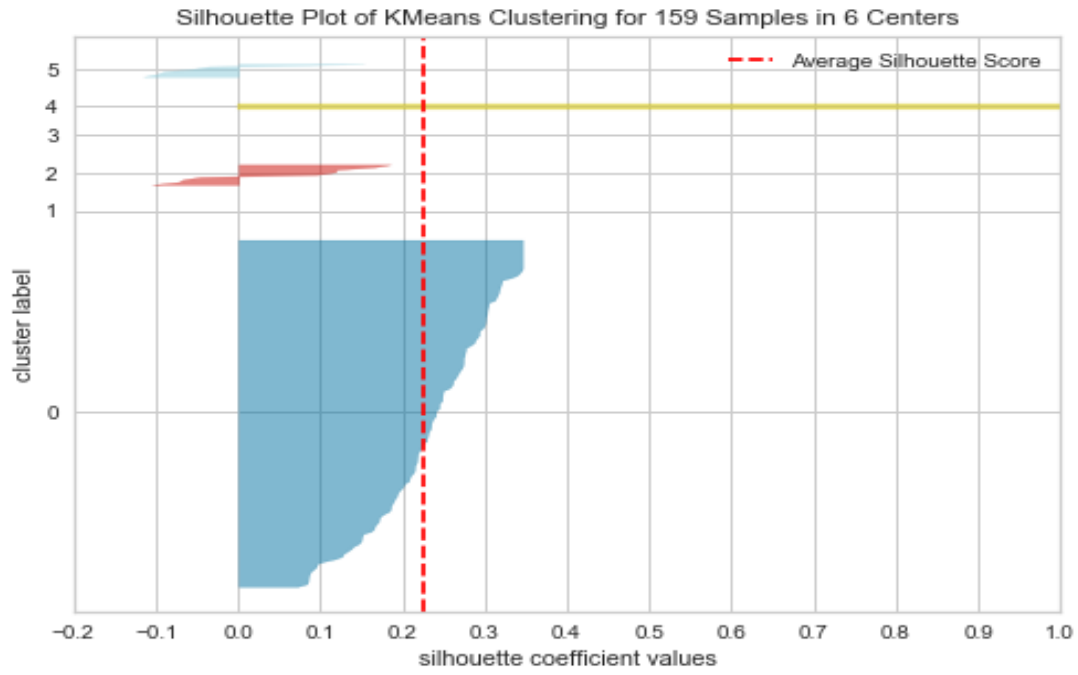


- The distortion score is decreasing gradually as the value of k increases
- The plot doesn't contain any sharp decrease
- The elbow method isn't decisive

METHODOLOGY

K-Means Clustering

- Using the silhouette score to determine the optimal value of k



- The best average score is about 0.22, not a good score
- Clusters 1 and 3 are so tiny
- Clusters 2 and 5 contain neighborhoods incorrectly classified

Decision :

Use DBSCAN algorithm

METHODOLOGY

DBSCAN Clustering

Pros :

- Minimal domain knowledge to determine the input parameters
- Discovery of clusters with arbitrary shapes.
- Good efficiency on large databases.

Algorithm :

- Computing the distance between every point and all other points.
- Place the points into one of three categories.

Inputs :

- **Epsilon** : radius that if includes enough number of points within, we call it dense area.
- **Minimum_Samples** : minimum number of data points in a neighborhood to define a cluster

Outputs :

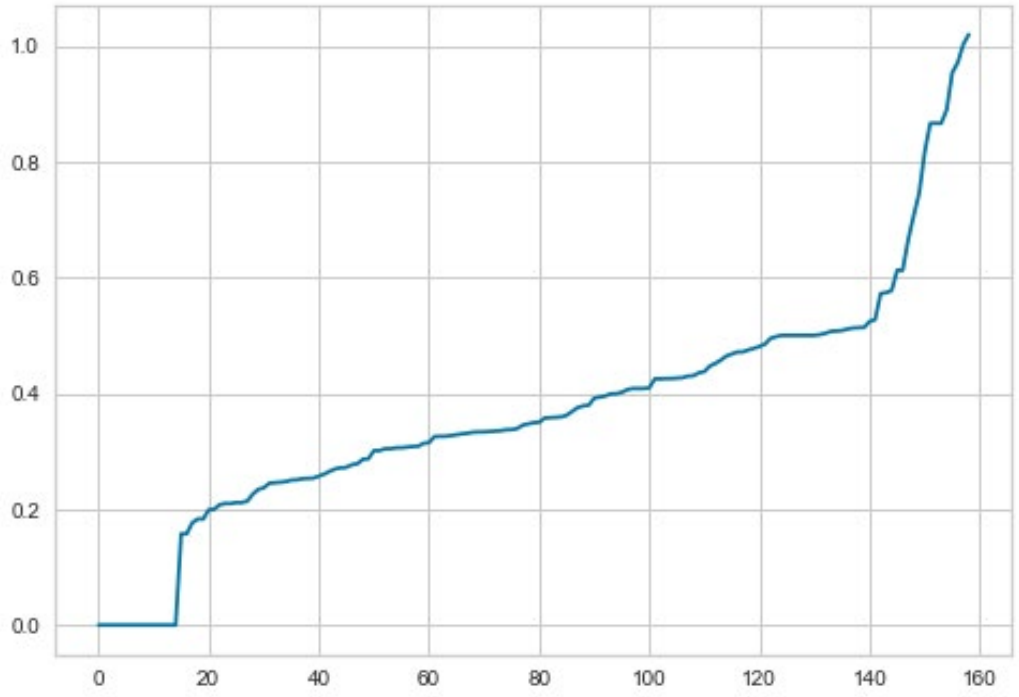
- **Core point** : has at least min_samples points whose distance with respect to the point is below the threshold defined by epsilon.
- **Border point** : A point that isn't in close proximity to at least min_samples points but is close enough to one or more core point.
- **Noise point** : Points that aren't part of any cluster.

METHODOLOGY

DBSCAN Clustering

Find the best value for *Epsilon :*

1. Calculate the distance from each point to its closest neighbor.
2. Sort and plot results.
3. The optimal value for epsilon will be found at the point of maximum curvature.



The best value for Epsilon is 0.5

METHODOLOGY

DBSCAN Clustering Results

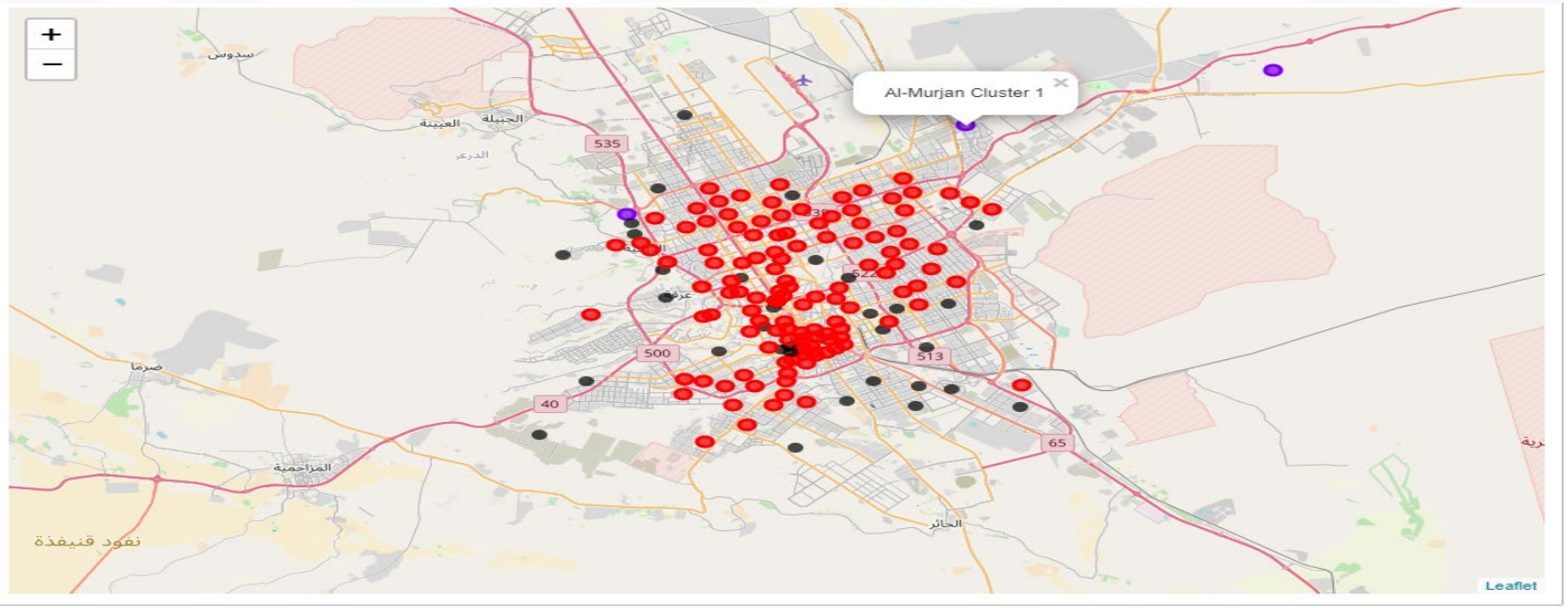
	Municipality_ar	Neighborhood_ar	Postal_code	Municipality_en	Neighborhood_en	Latitude	Longitude	Cluster_labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Cor
0	بلدية العليا	حي العليا	12222	Olaya	Al Alia	24.69502	46.69004	0.0	Coffee Shop	Café	Cupcake Shop	D
1	بلدية العليا	حي الازدهار	12486	Olaya	Al-Izdihar	24.77904	46.72446	0.0	Coffee Shop	Jewelry Store	Falafel Restaurant	Bik
2	بلدية العليا	حي المصيف	12466	Olaya	Al-Masif	24.76386	46.68849	0.0	Coffee Shop	Clothing Store	Hotel	B
3	بلدية العليا	حي التعاون	12475	Olaya	Al-Taawon	24.76724	46.69640	0.0	Coffee Shop	Donut Shop	Gift Shop	
4	بلدية العليا	حي الورود	12252	Olaya	Al-Worood	24.72213	46.68584	0.0	Coffee Shop	Pizza Place	Furniture / Home Store	Pha

- *Cluster 0 : 131 neighborhoods*
- *Cluster 1 : 3 neighborhoods*
- *32 neighborhood were identified as noise*

RESULTS

The City's Map

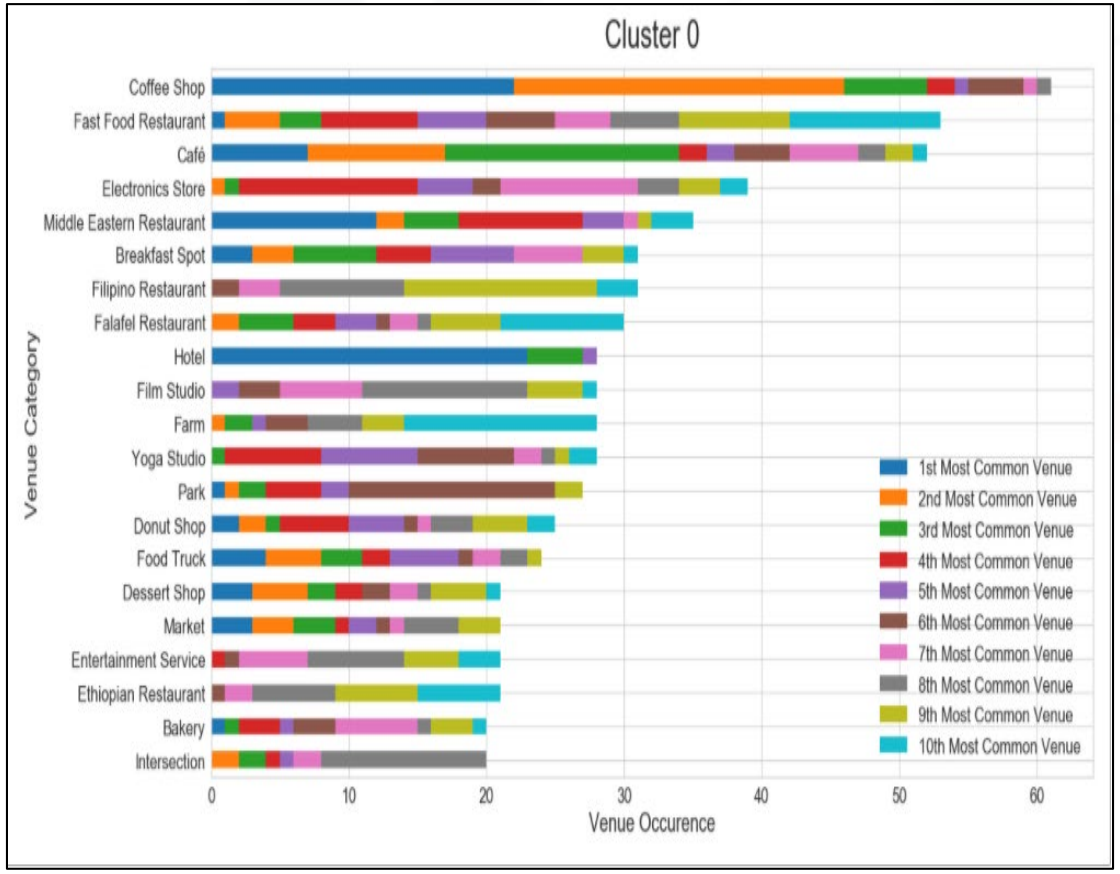
Geographic distribustion of the neighborhoods



Cluster 0 Cluster 1 Noise Points

RESULTS

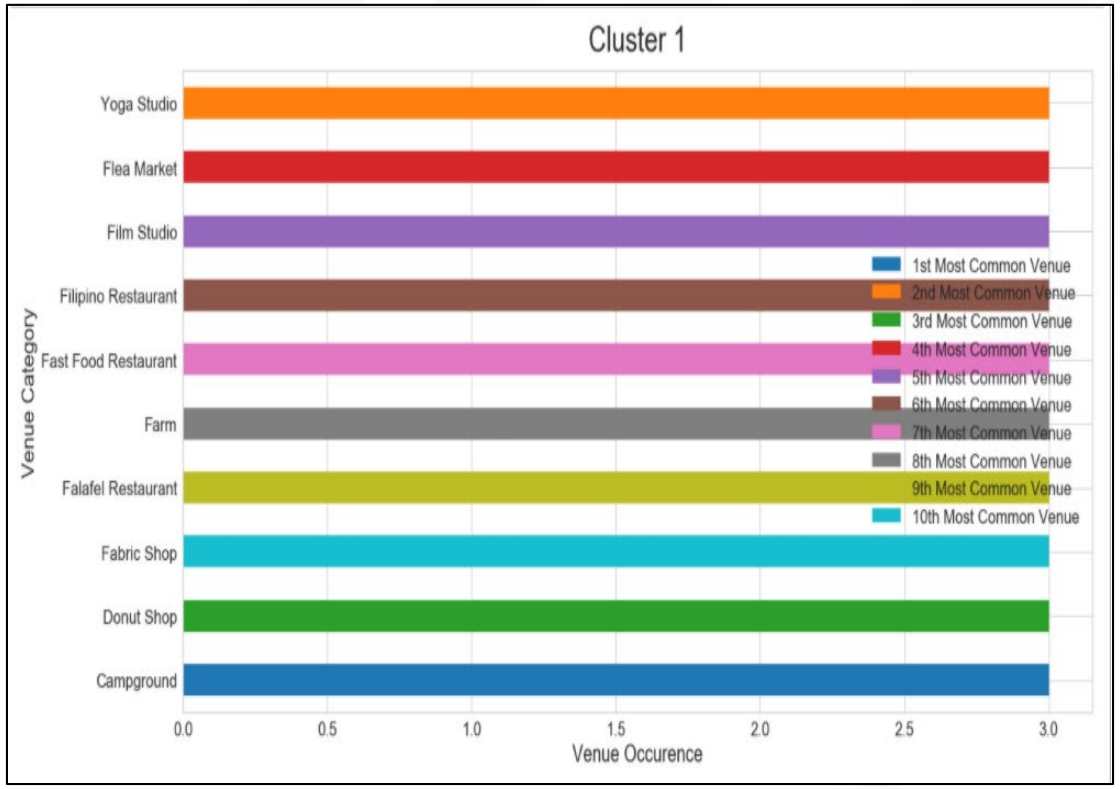
Clusters Common Venues



- Coffee shops appeared more than 60 times in the most 10 common venues, and more than 20 times as the first most common one.
- A lot of restaurants exist with different specialties (fast food, middle eastern, Filipino etc.)
- Hotels (22 time as first common venue) and café places (more than 50 appearances) are also so common in these neighborhoods.

RESULTS

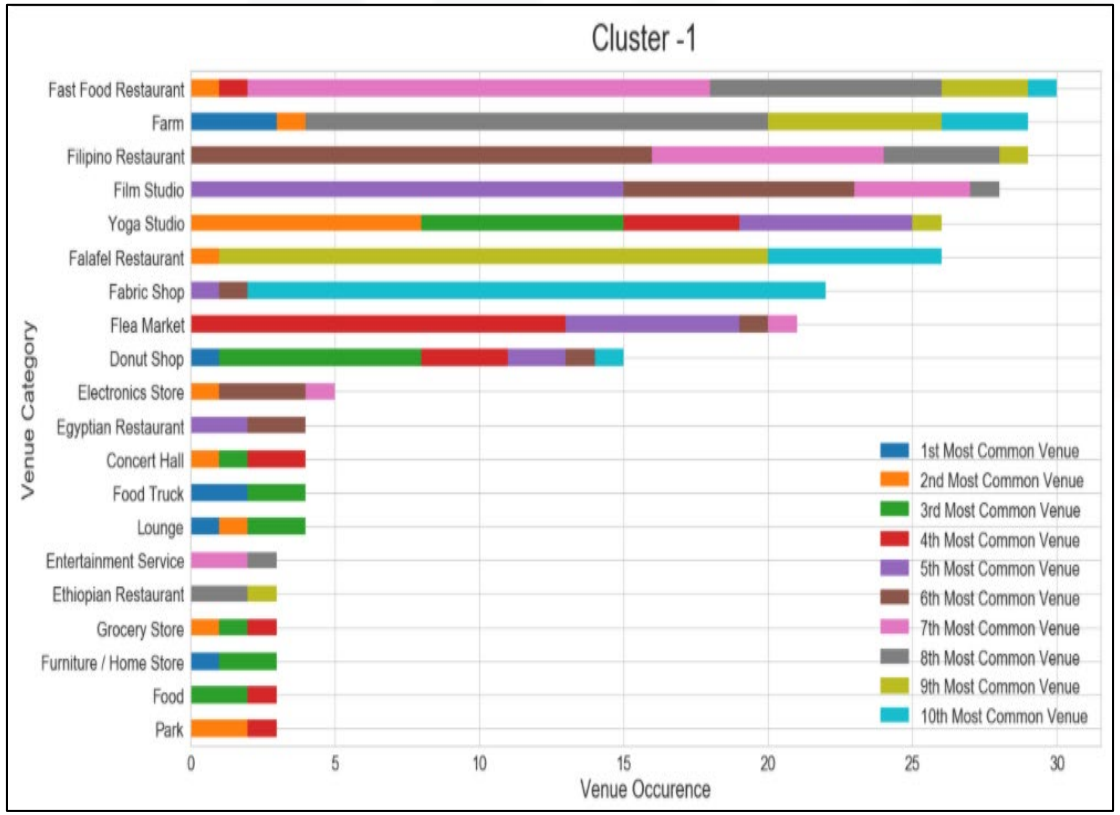
Clusters Common Venues



- Campgrounds is the most common venue.
- Yoga studios is the second common venue and film studios is the fifth in all these neighborhoods.
- There exist some farms, some restaurants (with no fast food restaurants) and some shops in these neighborhoods.

RESULTS

Clusters Common Venues



- neighborhoods clustered as noise have a lot of restaurants as common places but also studios, farms shops, parks, lounges, concert halls etc.

RESULTS DISCUSSION

- The city has the character of the modern cities
- Over than 70% of the city neighborhoods shares the same characteristics
- In cluster 0 neighborhoods we found a lot of coffee shops, café places, hotels, restaurants etc.
- The city is so populated, receives a lot of tourists each year and it contains a lot of companies, that explains the popularity of this kind of places.

RESULTS DISCUSSION

- It’s clear why cluster 1 have only three neighborhoods : Venues share the same ranking, that’s the power of clustering algorithms.
- Neighborhoods of cluster 1 are quitter place to live they reside outside the city (points in purple).
- Rather than classifying the rest of the neighborhoods incorrectly, DBSCAN identify them as noises.
- These neighborhoods share some venues with both previous clusters but with a different ranking and a less number, some restaurants, shops, studios.

RESULTS DISCUSSION

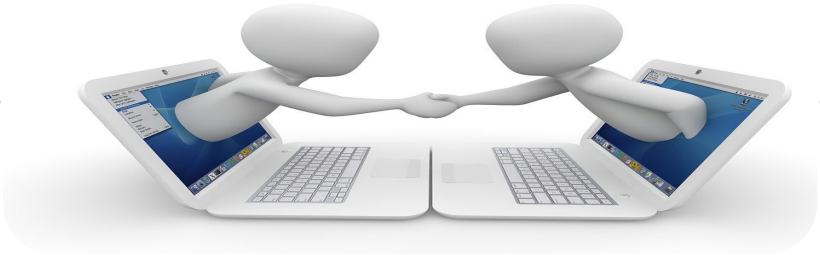
Recommandations

- Run further analysis on unclustered neighborhoods and identify similarities between them.
- Build a little application that facilitate to a user to find results about a specified neighborhood.

CONCLUSION

Foursquare API is a powerful tool that helps to explore neighborhoods and cities

DBSCAN is a versatile algorithm for clustering

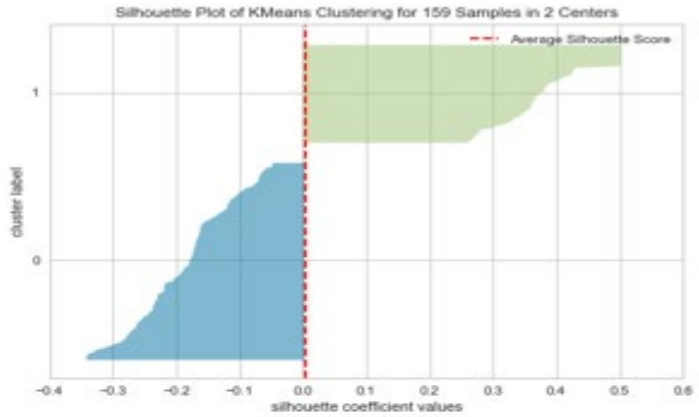


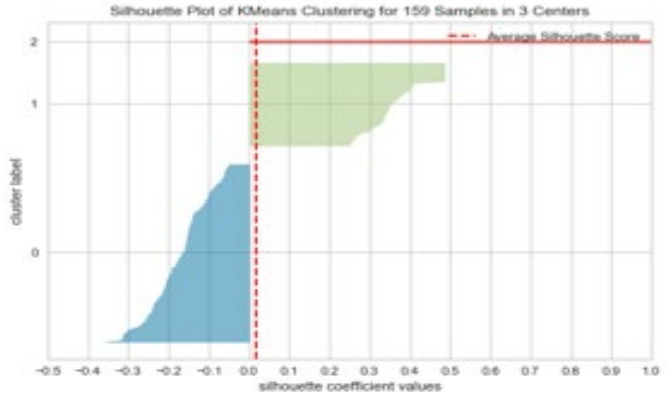
The number and the type of venues in over than 70% of the neighborhoods explains why *Riyadh* is the most populated and the main financial hub of the country

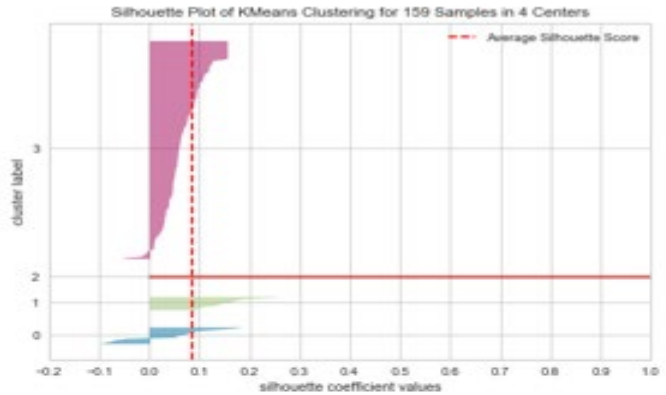
Three of the neighborhoods are so similar and represents the calmest places, the rest of the neighborhoods are unique

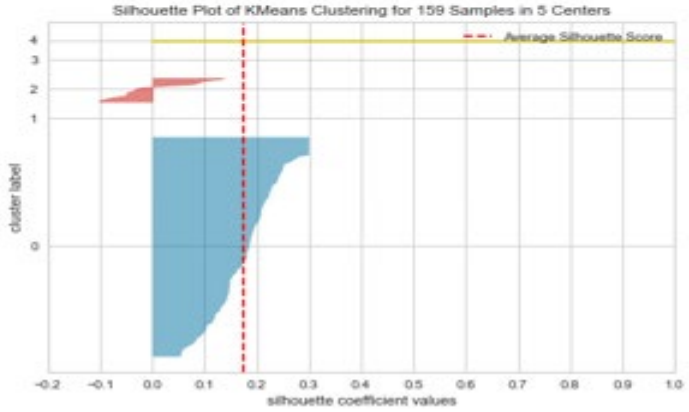
APPENDIX

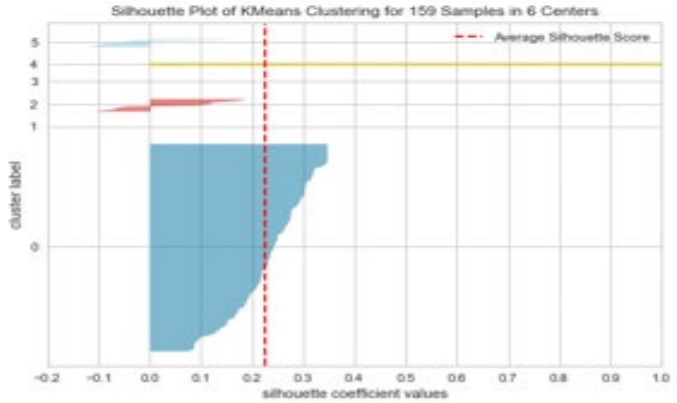
Silhouette Score For Different Values Of K With KMeans Algorithm

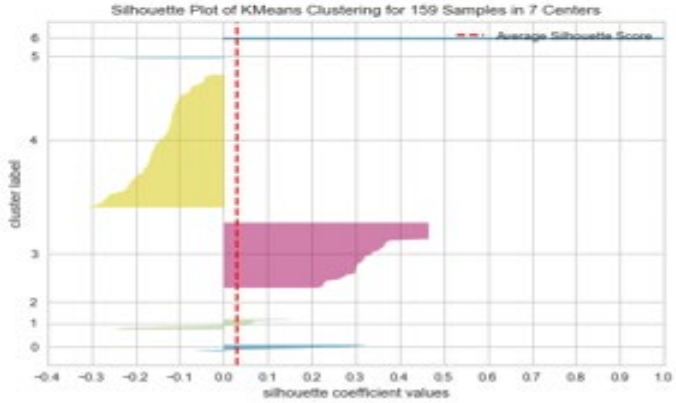






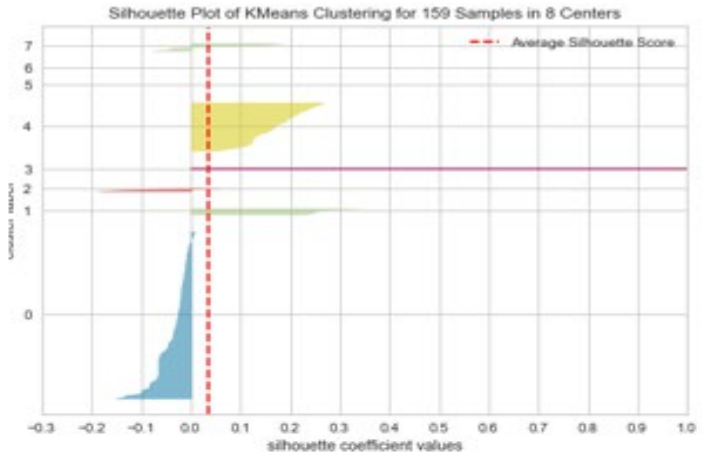


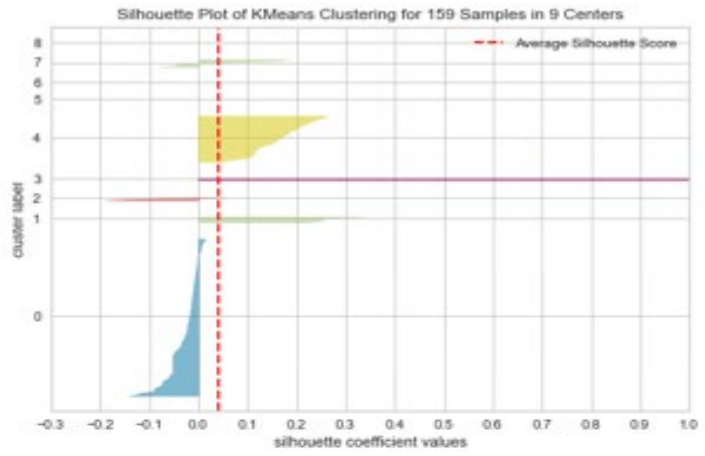


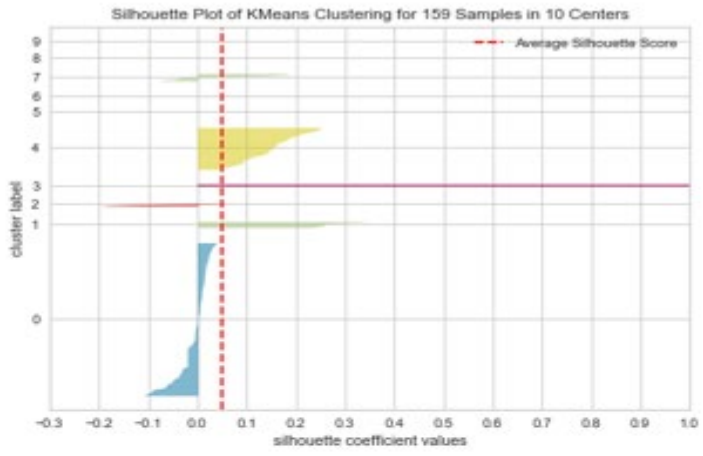


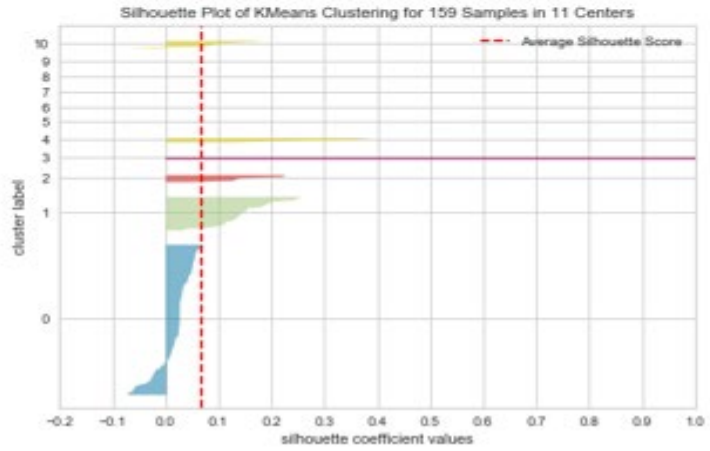
APPENDIX

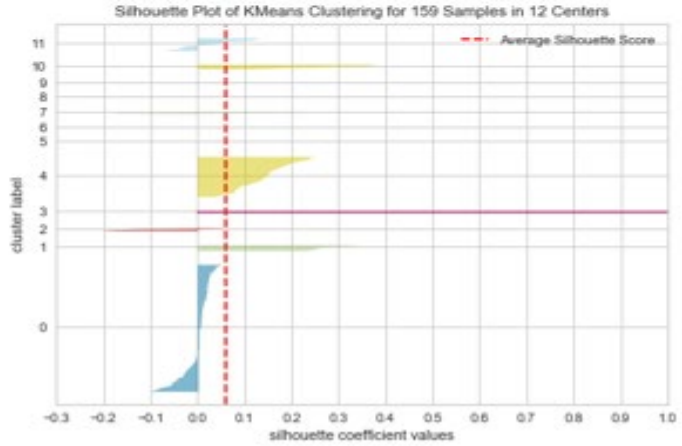
Silhouette Score For Different Values Of K With KMeans Algorithm

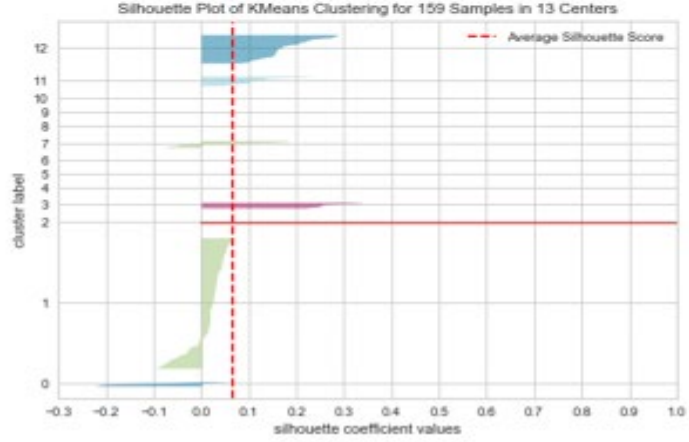












IBM Developer

SKILLS NETWORK