Tourism in Bahrain

By Maryam Almanea

August 2020

Introduction

Find right hotel to stay , place to dine in or top places to visit might be a challenge for tourists .

The main objectives of this project to explore Bahrain top big cities and venues surrounding the cities and recommend best places. Data science methodology and k-means clustering algorithm are used to carry out analysis. The result can help and satisfy visitor's needs, where best hotel to stay? Where best places to visit? And best restaurants to dine in

Data acquisition and cleaning

1. List of largest cities in Bahrain and geographic location , data scraped using BeautifulSoup python library from geonames . The original data were need formatting , extract cities names and split location into geolocation coordinates

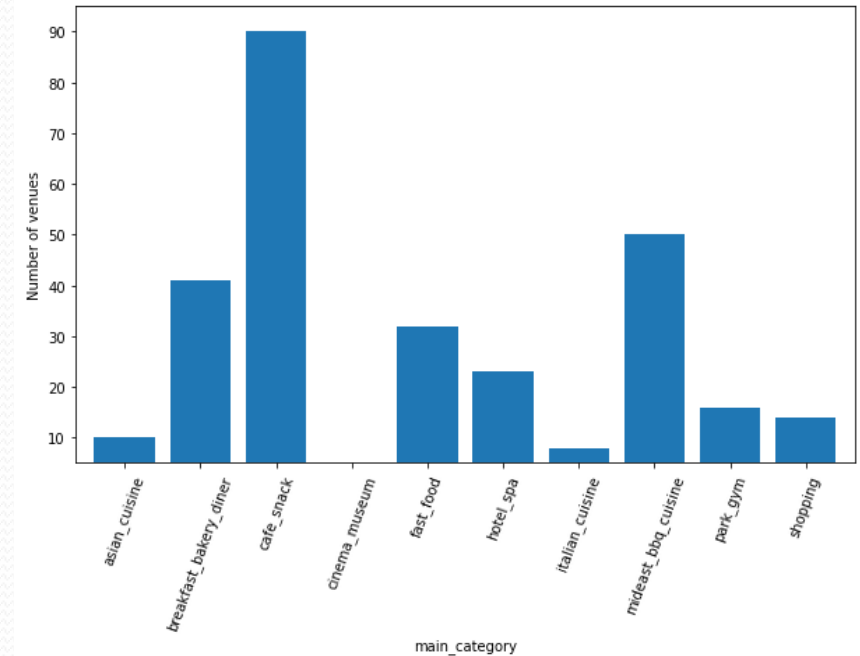|	City	Latitude	Longitude
0	Manama	26.228	50.586
1	Al Muharraq	26.257	50.612
2	Ar Rifā'	26.13	50.555
3	Dār Kulayb	26.069	50.504
4	Madīnat Ḥamad	26.115	50.507
5	Madīnat 'Īsá	26.174	50.548
6	Sitrah	26.155	50.621
7	Jidd Ḥafş	26.219	50.548
8	Al Ḥadd	26.246	50.654

Data acquisition and cleaning

2. Foursquare Venue data : venue name , ID , geolocation coordinates , category ,total likes counts.

- ➢ In total, 392 rows and 115 categories in the raw dataset.
- ➢ Cleaned data contains 46 categories which attaract tourists like : hotels, coffee shops, restaurants, shopping malls.

	index	City	Latitude	Longitude	Venue	id	Venue Latitude	Venue Longitude	Venue Category
0	0	Manama	26.228	50.586	Bahay Kubo Restaurant	4d07a2f18620224b7e34b940	26.228443	50.582647	Asian Restaurant
1	1	Manama	26.228	50.586	Sameeh pastries	4cb0419acbab236acbc29e73	26.226697	50.588958	Turkish Restaurant
2	2	Manama	26.228	50.586	The Lebanese Restaurant	4d288e6ff7a9224b630ef69f	26.228518	50.583612	Middle Eastern Restaurant
3	3	Manama	26.228	50.586	Al Wasmiya Village Restaurant	4e270708ae609b2f94ecc680	26.229518	50.581646	Restaurant
4	4	Manama	26.228	50.586	Jahan Grills	4dd7f705b0fb8af380937ea3	26.224671	50.587759	BBQ Joint
5	5	Manama	26.228	50.586	تكة ابل	4f26d629e4b02a9500c19159	26.229502	50.591782	Restaurant
6	6	Manama	26.228	50.586	تكة أمين	51d09d47498e2a039a702d75	26.228616	50.585451	BBQ Joint

Methodology

1. Define 10 new categorical variables to better group venues based on similarities. For example:

   ❖ Hotel_sap (hotel , spa , lounge ,resort)

   ❖ Shopping (Men's Store, Boutique, Shopping Mall, Clothing Store)

   ❖ Café_snack (Café , Coffee Shop , Ice Cream Shop, Juice Bar ,Cafeteria , Sandwich)

   ❖ breakfast_bakery_diner (Breakfast Spot ,Bakery , Department Store ,Grocery Store)

   ❖ italian_cuisine (Italian Restaurant, Pizza Place)

Methodology

2. Based on descriptive statistic of total likes, new categorical variables were created as follows:

- 1 star : total likes 6 or less
- 2 star : total likes more than 6 and less or equal to 11
- 3 star : total likes more than 11 and less or equal to 27
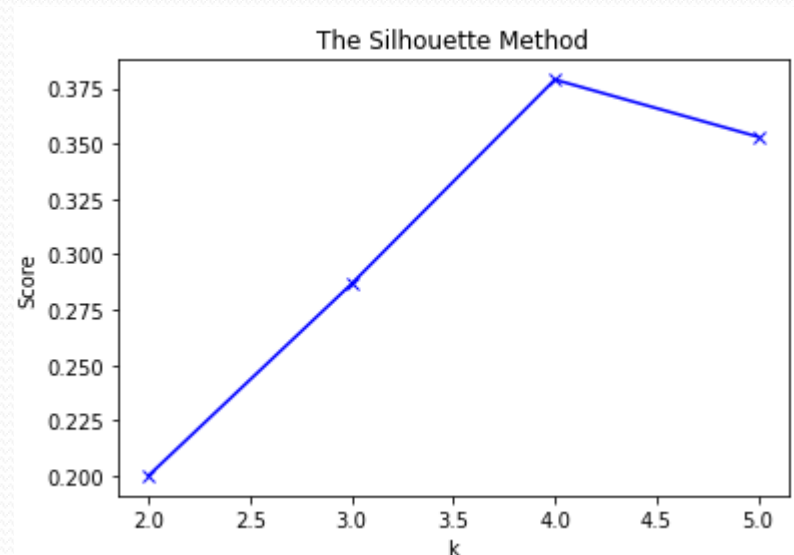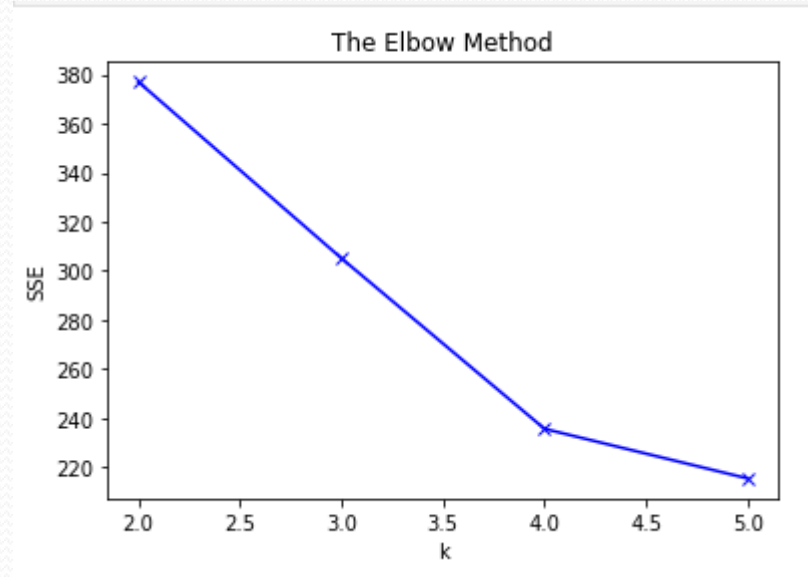- 4 star : total likes more than 27

	total likes
count	290.000000
mean	33.510345
std	75.758048
min	0.000000
25%	6.000000
50%	11.000000
75%	27.000000
max	787.000000

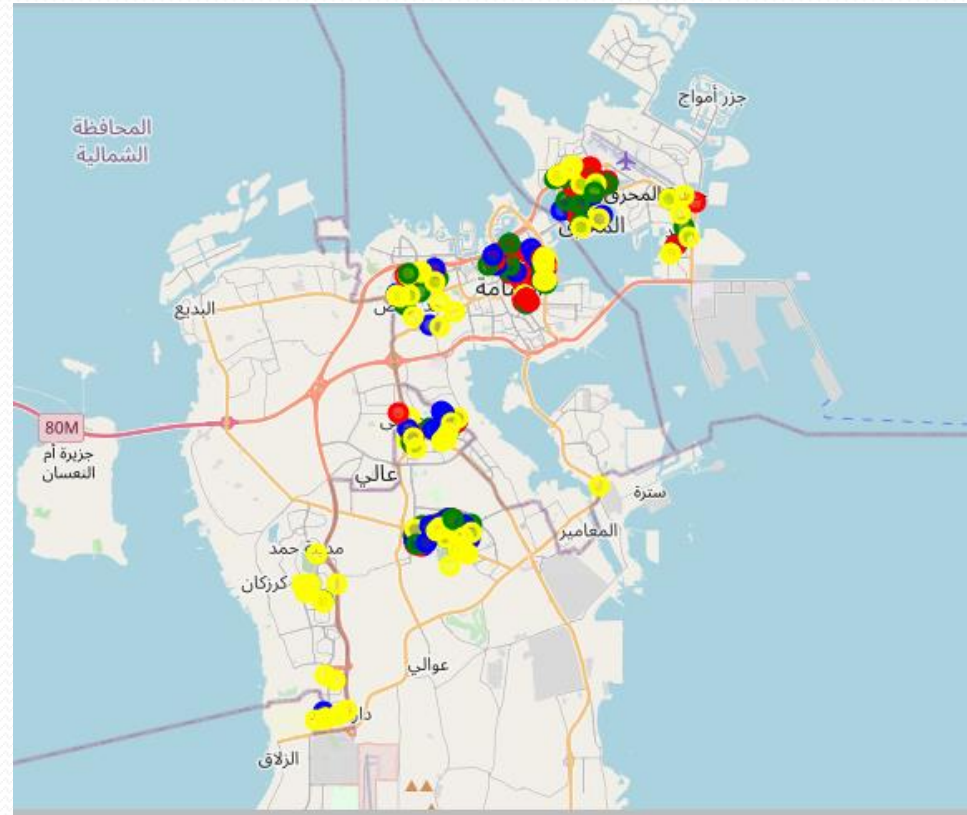| index	City	Latitude	Longitude	Venue	id	Venue Latitude	Venue Longitude	Venue Category	main_category	total likes	venue rate
0	Manama	26.228	50.586	Bahay Kubo Restaurant	4d07a2f18620224b7e34b940	26.228443	50.582647	Asian Restaurant	asian_cuisine	48	star4
1	Manama	26.228	50.586	Sameeh pastries	4cb0419acbab236acbc29e73	26.226697	50.588958	Turkish Restaurant	mideast_bbq_cuisine	38	star4
2	Manama	26.228	50.586	The Lebanese Restaurant	4d288e6ff7a9224b630ef69f	26.228518	50.583612	Middle Eastern Restaurant	mideast_bbq_cuisine	45	star4
3	Manama	26.228	50.586	Al Wasmiya Village Restaurant	4e270708ae609b2f94ecc680	26.229518	50.581646	Restaurant	mideast_bbq_cuisine	11	start2
4	Manama	26.228	50.586	Jahan Grills	4dd7f705b0fb8af380937ea3	26.224671	50.587759	BBQ Joint	mideast_bbq_cuisine	55	star4

Methodology

3. Clustering

    1. Perform one hot encoding to convert categorical data to integer data .

    2. Find optimal number of clusters using elbow and silhouette score methods in range between 2 to 5 . Result shows that k=4 is optimal number of clusters.

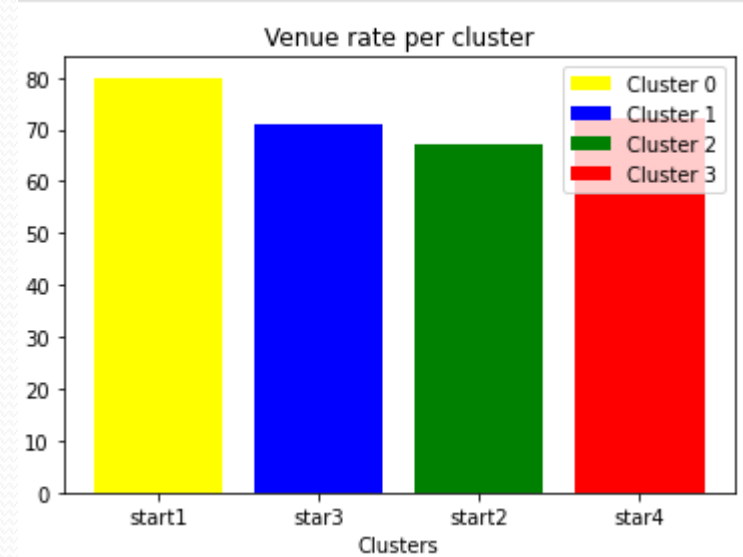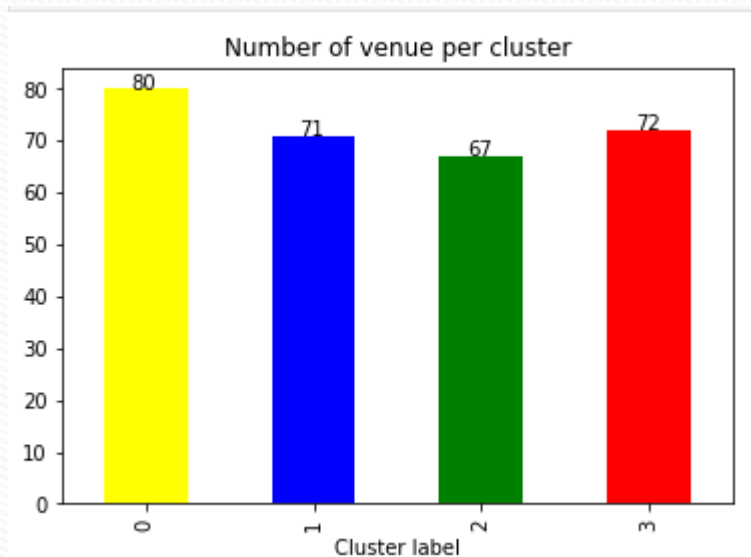   3. K-means algorithm is used to perform data clustering with 4 clusters

The Elbow Method

The Silhouette Method

Results

The results of K-means clustering for 289 venues are visualized on map .

➢ Yellow markers - cluster 0
➢ blue markers – cluster 1
➢ green markers- cluster 2
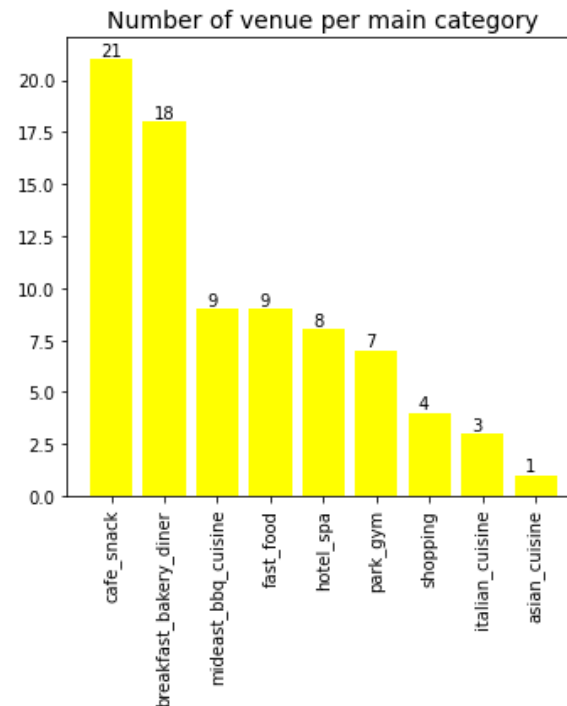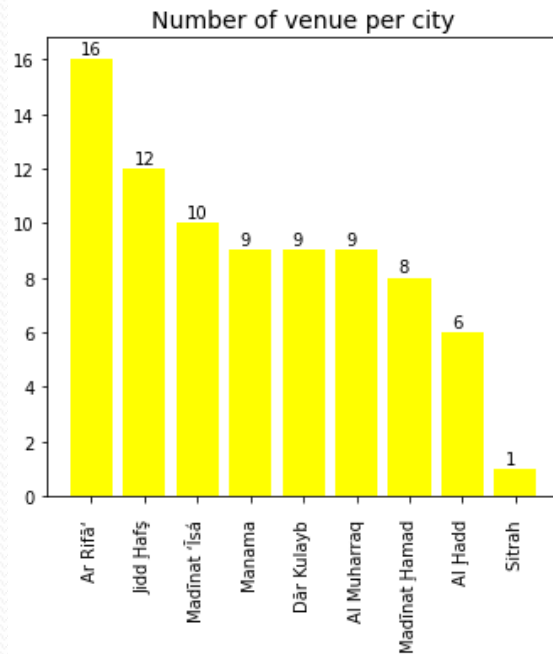➢ red markers – cluster 3

Results

That total number of venues is largest in cluster 0 followed by cluster 3 and 1 with 72,71 venues respectively and lowest in cluster 2. Despite convergence in number of venues, the highest user recommended venues are in cluster 3 then cluster 1,2 while least in cluster 0 based on user total likes
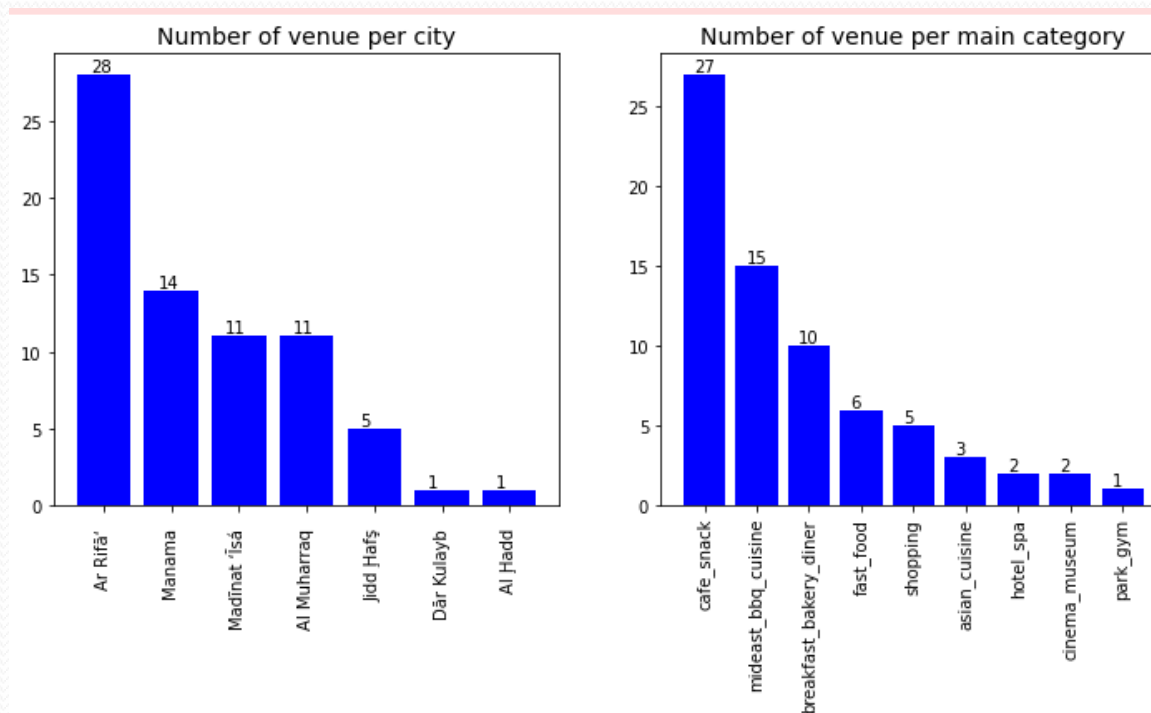
Results

Cluster 0

➢ 80 venues with lowest rating (1 star) where mainly located at Ar Rifa city. The most common categories are food, hotels, park and shopping.
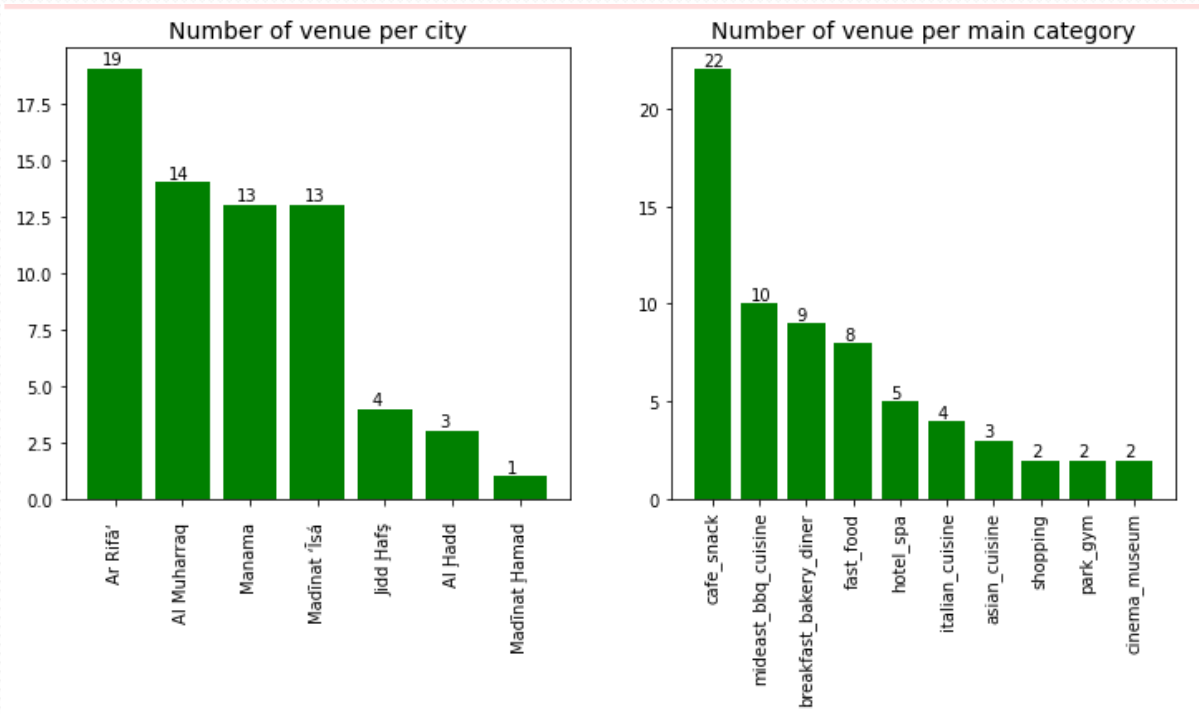
Results

Cluster 1

➢ 71 venues of 3 star rating where mainly located at Ar Rifa city. The most common categories are food, shopping, hotel .
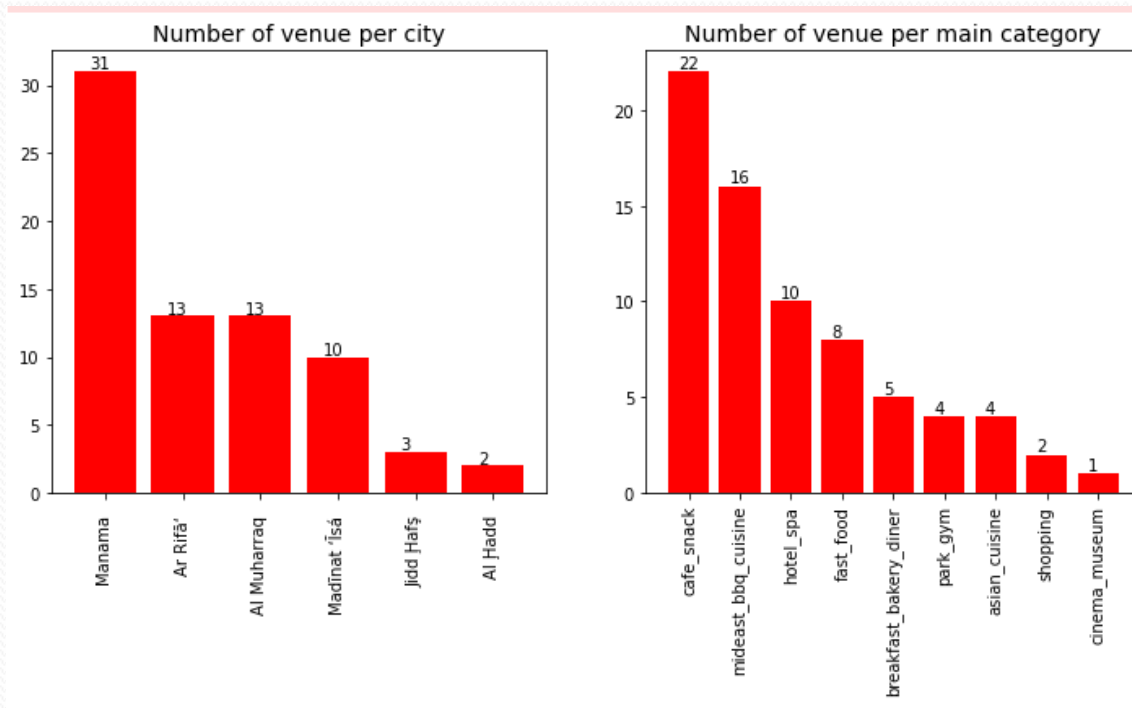
Results

Cluster 2

➢ 67 venues of 2 star rating where mainly located at Riffa , Muharraq , Manama and Madinat Isa. The most common categories are food, ,hotel, shopping ,park and cinema.

Results

Cluster 3

➤ 80 venues with lowest rating (1 star) where mainly located at Ar Rifa city. The most common categories are food, hotels, park and shopping.

Limitation and Future Research

- We focused only on top largest cities in Bahrain , it would be valuable if could add more cities to analysis .
- Paid Foursqaure account to overcome limitation of free account .

Conclusion

This project started with identifying business problem, extracting and cleaning data then data clustering using k-means into 4 clusters. The findings show that venues of cluster 3 are highly recommended where highly rated places, starting from hotels , park ,gym ,historical site, shopping malls , coffee shops and cuisine from different cultures