The Battle of the Eateries – Exploring Food Venues in Dubai

Introduction

Dubai is a booming cosmopolitan city with a very diverse population. Dubai has also become an international travel hub, with many flights that transit through its airport. Given its increasing popularity, the number of hotels and restaurants has significantly increased over the last decade with many new outlets being opened every year. Additionally, Dubai's growth has meant that construction projects have increasingly sprawled across previously undeveloped areas. Given the hot summers, it is also not surprising that another increase has been experienced in the number of indoor shopping malls, where in addition to stores, food courts, cafes and restaurants abound.

As the most populous city in the UAE and being a central business hub in the middle east, several interesting questions could be explored. Firstly, are there certain areas of Dubai where more growth in the number of food-related establishments has increased more than others? And if so, what type of establishments are the most popular? Other interesting questions of interest include the cost of visiting such establishments? Does Dubai's often-cited population wealth mean that more establishments are on the high-end range of the pricing spectrum? If so, are there any establishments that cater to those with lower incomes and if so, are those establishments located in different areas of the city?

Audience

The target audience of this Capstone project and analysis varies greatly. Firstly, given that it is a hub for commerce and travel, knowing where certain food establishments are clustered/located might be of interest to any visitors to the city. Other variables of interest that the analysis will explore are the ratings of various establishments (provided via other patrons) as well as the average cost of visiting different establishments in different areas of the city. Having this information would allow visitors to make informed decisions that cater to their personal preference. Another target audience of interest are individuals seeking to open new restaurants, cafes etc. The information that the analysis will provide can shed some interesting insights into what locations are most popular, what type of restaurants are most prevalent and by extension are there certain categories of restaurants that are not available and may present an interesting business opportunity? The analysis that will be performed will include informative visualizations of the data that could reveal interesting trends.

Data Sources

In order to perform this analysis, two main sources of data will be used. First, in order to get location information about different venues that currently exist in Dubai, Foursquare API data

will be used to gather complete information about different types of venues (location, name, category, etc.)

In order to explore the ratings of each venue, a second set of data will also be used. Urbanspoon rating data will also be collected (Zomato API) and the ratings from this API will be combined with the Foursquare data. I chose to utilize this second dataset because of the popularity of the Urbanspoon/Zomato API, and as such, combining the information from both sources could lead to increased confidence in the findings of the analysis and allows me to answer the questions that are of interest. Below are the types of information that can be retrieved from each source.

Foursquare API: (https://developer.foursquare.com/places)

- Name of venues
- Category of venue
- Location (longitude, latitude)

Urbanspoon/Zomato API: (https://developers.zomato.com/api)

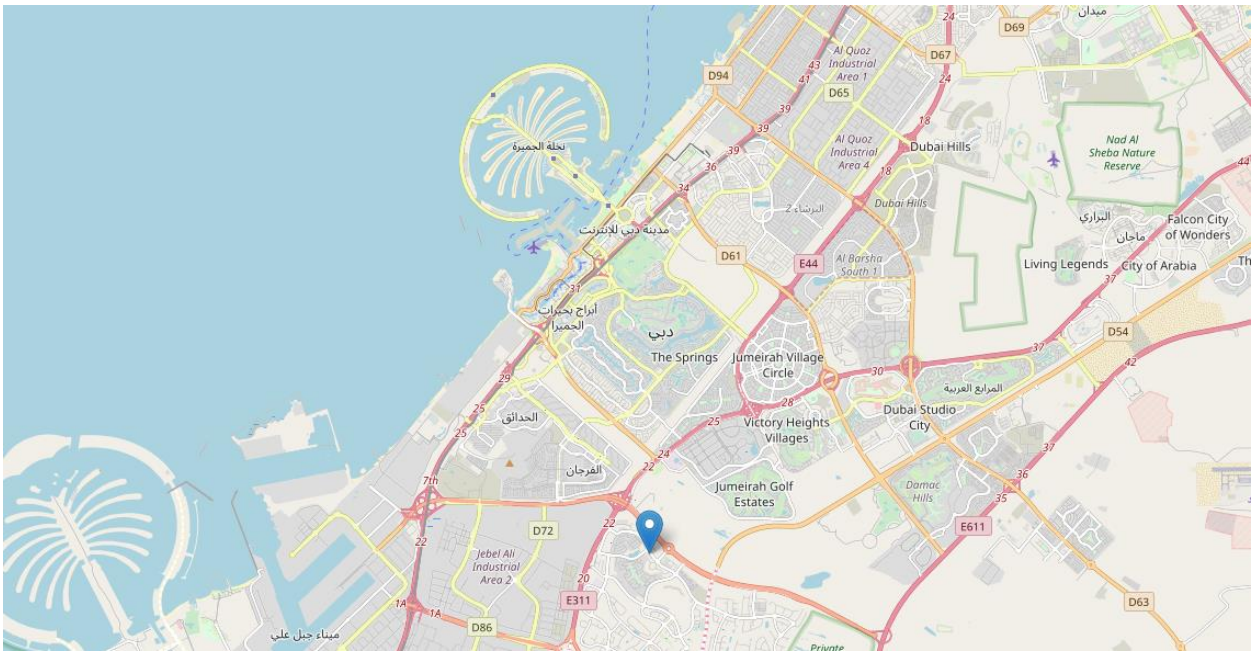
- Name of venue
- Ratings
- Price range
- Location (longitude and latitude)

Once the two datasets are combined, I will perform exploratory analysis as well as provide different visualizations in order to explore the questions listed in the introduction. Some techniques that will be used include production of descriptive statistics, maps and other visualizations and clustering.

Gathering the Data

Dubai is composed of a number of many different areas, especially with the recent expansion and building boom. Dubai's metropolis covers an area of 35 sq km, however recent development has meant that the city has sprawled over a considerably larger area.

I used the geopy library to extract the latitude and longitude values of Dubai, and then plotted them on a map using folium.



Foursquare API

First, I fetched all the venues in Dubai within a range of 10 kms from the coordinates above using the Foursquare API. The Foursquare API has the explore API function which allows us to find venue recommendations within a given radius from the given coordinates. I will use this API to find all the venues needed for my analysis. This resulted in 238 venues in a 10km radius around my coordinates.

Urbanspoon API

The Urbanspoon/Zomato API allows using its 'search' API to search for any given venue based on certain search filters such as query, latitude, longitude and more. In order to use Zomato, I had to generate a user key which can done by registering for a developer account (similar to Foursquare).

I will use the name, lat, and lng values of various venues fetched from Foursquare API to use the search API in zomato in order get more information regarding each venue.

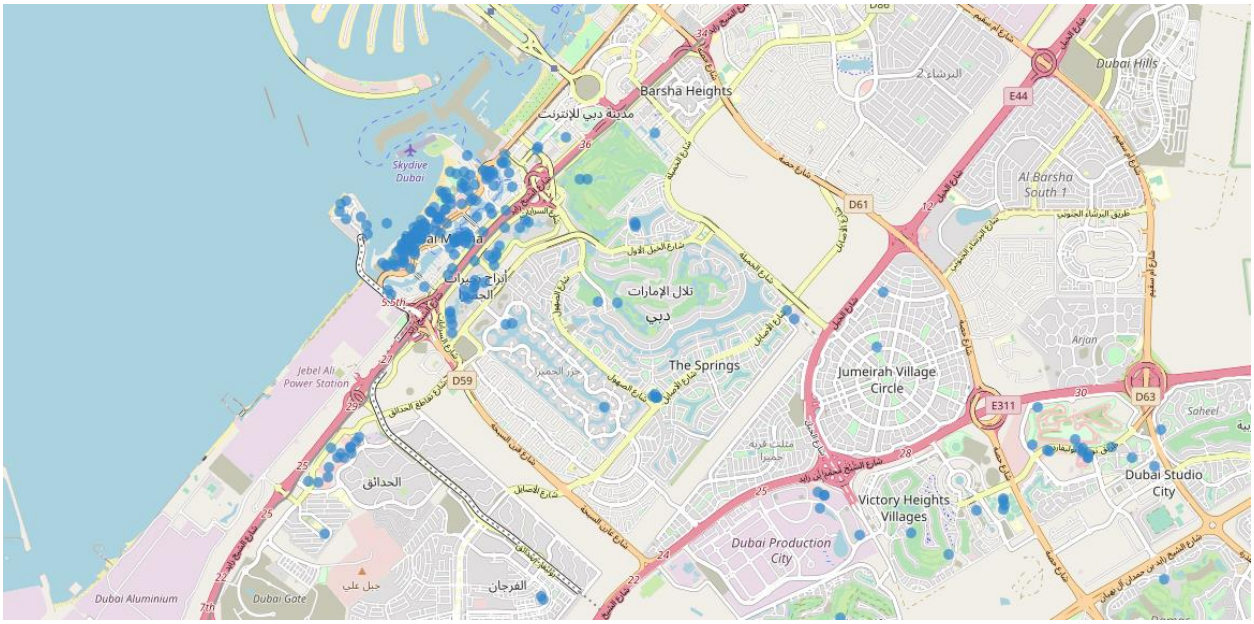
The query will be the name of the venue.
The start defines from what offset we want to start, so it remains at 0.
The count defines the number of venues I want to fetch. As I have the exact location coordinates (from Foursquare), I will fetch only one.
I will supply the latitude and longitude values.

and finally, I will set the sorting criteria as real_distance so each time we get the venue information we would be searching based on location coordinates. This API fetched the required information for all 238 venues.

Cleaning the Data

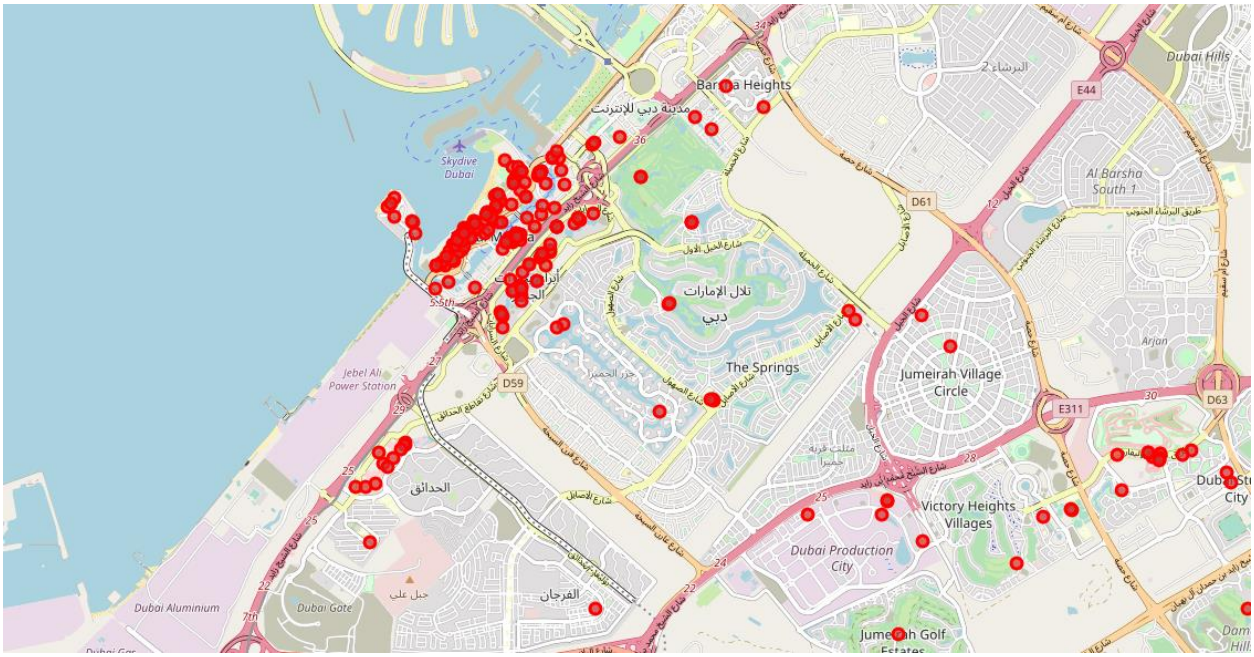
Given that I have data from 2 different sources, it is necessary to check whether the data points align. One quick way to do so is to plot each dataset on a separate map and compare them. The figures below are the result of plotting the Foursquare and Urbanspoon data on separate maps using Folium. I did this to see whether the locations overlapped as I had expected them to.

Foursquare Map



We can already see some interesting things! There seems to be a cluster of venues right at the waterfront. A quick search revealed that the Dubai Marina region is highly popular, with many hotels, restaurants and other venues. The location at the waterfront also makes this an unsurprising (yet interesting) finding.

Urbanspoon Map



The maps appear to show a similar pattern, but a closer inspection I can see that some venues from the two APIs do not align with each other. To explore further, I combined the datasets.

To combine the two datasets, I had to check that the latitude and longitude values of each corresponding venue match. After careful analysis, I decided to drop venues from the two datasets where the latitude and longitude values different by more than 0.0004 from one another. To do so, I rounded both the latitude and longitude values up to 4 decimal places. Then, I calculated the difference between the corresponding values and to determine whether the delta between them was less than 0.0004 which should ideally mean that the two locations are the same. This removed several outliers from the two datasets. This left me with 130 venues to work with.

	name	categories	lat	lng	venue	latitude	longitude	price_for_two	price_range	rating	address	lat_diff	lng_diff
0	Al Furjan Pavilion	Shopping Mall	25.0246	55.1529	Al Arrab	25.0246	55.1528	205.0	3.0	2.7	Ground Floor, Al Furjan Pavilion, Al Furjan, D...	0.0000	-0.0001
1	Me'aisem City Center (مقيصم سيتي سنتر)	Shopping Mall	25.0398	55.1979	Starbucks	25.0397	55.1980	70.0	2.0	3.7	Ground Level, City Centre Me'aisem , Al Fay Ro...	-0.0001	0.0001
2	Jumeirah Islands Club	Beach Bar	25.0525	55.1627	The Glasshouse	25.0523	55.1627	320.0	4.0	3.4	Jumeirah Islands Clubhouse, Emirates Hills, Dubai	-0.0002	0.0000
3	Aloft Me'aisam, Dubai	Hotel	25.0379	55.1972	City Limits	25.0379	55.1972	235.0	3.0	3.6	Ground Level, Aloft Me'aisam, Dubai Production...	0.0000	0.0000
4	Pinkberry	Frozen Yogurt Shop	25.0539	55.1710	Pinkberry	25.0540	55.1709	90.0	2.0	3.9	Ground Floor, Town Center, Emirates Hills, Dubai	0.0001	-0.0001
5	Baskin Robbins	Ice Cream Shop	25.0423	55.1185	Baskin Robbins - Ibn Battuta Mall	25.0421	55.1186	45.0	1.0	3.7	Tunisia Food Court, Ibn Battuta Mall, Garden C...	-0.0002	0.0001
6	Ghaya Grand Hotel I غايا جراند	Hotel	25.0340	55.2036	Lobby Lounge - Ghaya Grand Hotel	25.0341	55.2035	150.0	3.0	3.1	Lobby Level, Ghaya Grand Hotel, IMPZ, Dubai	0.0001	-0.0001
7	Paul Café	French Restaurant	25.0398	55.1980	PAUL Bakery & Restaurant	25.0398	55.1981	210.0	3.0	3.9	Ground Level, City Centre Me'aisem, Al Fay Roa...	0.0000	0.0001
8	The Hamptons Cafe	Restaurant	25.0646	55.1479	The Hamptons Cafe	25.0645	55.1478	240.0	3.0	4.3	Street 1, Jumeirah Islands, Emirates Hills, Dubai	-0.0001	-0.0001
9	Starbucks	Coffee Shop	25.0539	55.1711	Starbucks	25.0539	55.1711	70.0	2.0	3.6	Ground Floor, Town Center, Emirates Hills, Dubai	0.0000	0.0000
10	Premier Inn	Hotel	25.0465	55.1189	Costa Coffee	25.0465	55.1191	100.0	3.0	3.2	Premier Inn, Ibn Batuta Mall, Jebel Ali Villag...	0.0000	0.0002

Additional data manipulation

I decided to use the venue name from Urbanspoon as it appeared to be more descriptive and accurate (using Arabic script). I also wanted to get the average price per person by dividing the column price_for_two by 2 and removing this column from the dataset along with some additional columns that I am not interested in. My data set was now ready for additional data analysis.

	categories	venue	latitude	longitude	price_range	rating	address	average_price
0	Shopping Mall	Al Arrab	25.0246	55.1528	3.0	2.7	Ground Floor, Al Furjan Pavilion, Al Furjan, D...	102.5
1	Shopping Mall	Starbucks	25.0397	55.1980	2.0	3.7	Ground Level, City Centre Me'aisem , Al Fay Ro...	35.0
2	Beach Bar	The Glasshouse	25.0523	55.1627	4.0	3.4	Jumeirah Islands Clubhouse, Emirates Hills, Dubai	160.0
3	Hotel	City Limits	25.0379	55.1972	3.0	3.6	Ground Level, Aloft Me'aisam, Dubai Production...	117.5
4	Frozen Yogurt Shop	Pinkberry	25.0540	55.1709	2.0	3.9	Ground Floor, Town Center, Emirates Hills, Dubai	45.0
5	Ice Cream Shop	Baskin Robbins - Ibn Battuta Mall	25.0421	55.1186	1.0	3.7	Tunisia Food Court, Ibn Battuta Mall, Garden C...	22.5
6	Hotel	Lobby Lounge - Ghaya Grand Hotel	25.0341	55.2035	3.0	3.1	Lobby Level, Ghaya Grand Hotel, IMPZ, Dubai	75.0
7	French Restaurant	PAUL Bakery & Restaurant	25.0398	55.1981	3.0	3.9	Ground Level, City Centre Me'aisem, Al Fay Roa...	105.0
8	Restaurant	The Hamptons Cafe	25.0645	55.1478	3.0	4.3	Street 1, Jumeirah Islands, Emirates Hills, Dubai	120.0

Methodology

I want to identify venues in Dubai based on their rating and average price. This would enable visitors to Dubai to select venues they would be interested in visiting based on their preferences.

As a first step, I retrieved the data from two APIs (Foursquare and Urbanspoon). I extracted the venue information from the center of Dubai, up to a radius of 10 km. The latitude and longitude values were then used to fetch venue rating and price from Urbanspoon.

Next, I then explored the data retrieved from the two APIs on two separate maps. The data from the two sources was then merged based on the name, latitude and longitude values from the two datasets. The final dataset now only included the rating and average price per person for each venue.

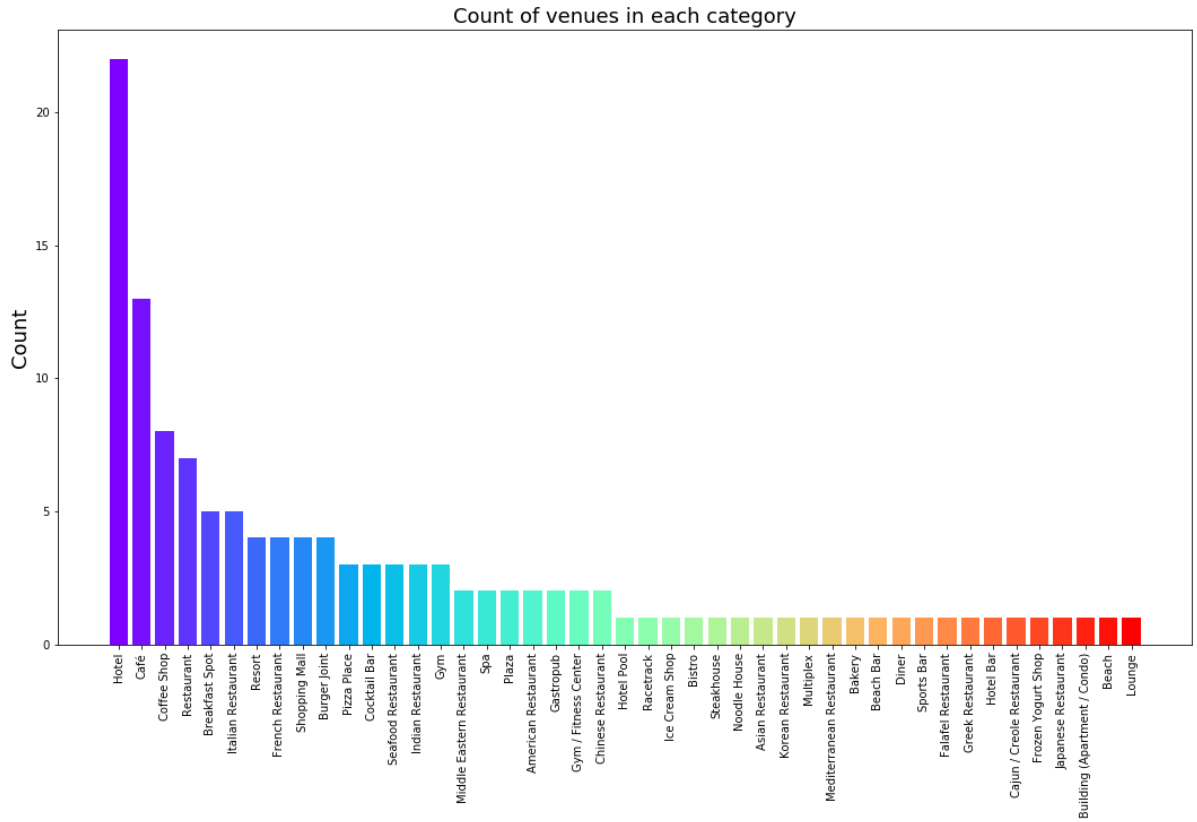
To continue the analysis, I will look at the dataset I created and produce some visualizations based on the ratings and price of each venue. This will help identify places where there may be clusters of venues that may be of interest. I will also explore venues with high ratings and those with low ratings (as well as prices) to determine whether any observations or inferences can be made. Finally, I will use clustering to identify whether venues with certain characteristics could be identified, and what those characteristics may be.

Analysis

The merged dataset is now ready. I will inspect venues based on their rating. The rating of a venue is based on user reviews and ranges from 1 to 5. I will also analyze the venues based on their price per person as well as the price range.

Categories

I have various types of venues in the final dataset. I will first take a look at the venues and explore the most popular venue categories. This can show two different things, popularity as well as whether there are any growth opportunities (for those venue categories with low counts).



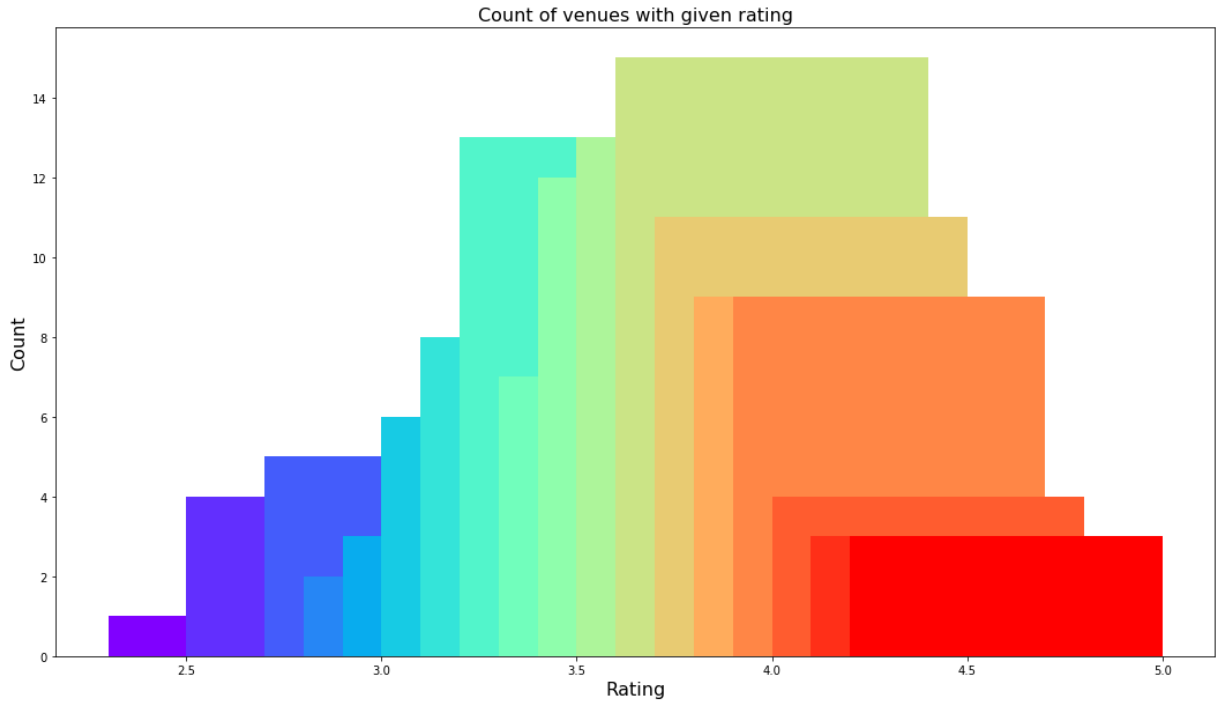
The figure reveals that hotels, cafes/coffee shops and restaurants are the most prevalent venues in the area I have explored. On the far end of the spectrum, it appears that lounges, frozen yogurt shops and some types of restaurants aren't as prevalent. This may be a function of several things, there are categories (subtypes of restaurants) that are listed outside the overall "restaurant" category. This means that a more nuanced analysis of the dataset may be necessary. Given the aims of this project and its scope, I will not be merging additional categories, however, it is a limitation that needs to be noted. The fact that hotels are the most popular venue is not surprising given Dubai's growth and popularity as a travel destination, as well as the waterfront cluster we observed on the map.

Ratings

The rating of a venue is an important factor that can influence a visitor's decision as to whether they should visit a certain venue or not. It also gives an idea for investors as to what the competition looks like in certain categories of venues, which could impact whether they want to enter the market or not.

To further explore ratings, I will first examine the average rating for all the venues in the area of the city I am exploring. Next, I will plot the venues on the map and color code them to represent their rating.

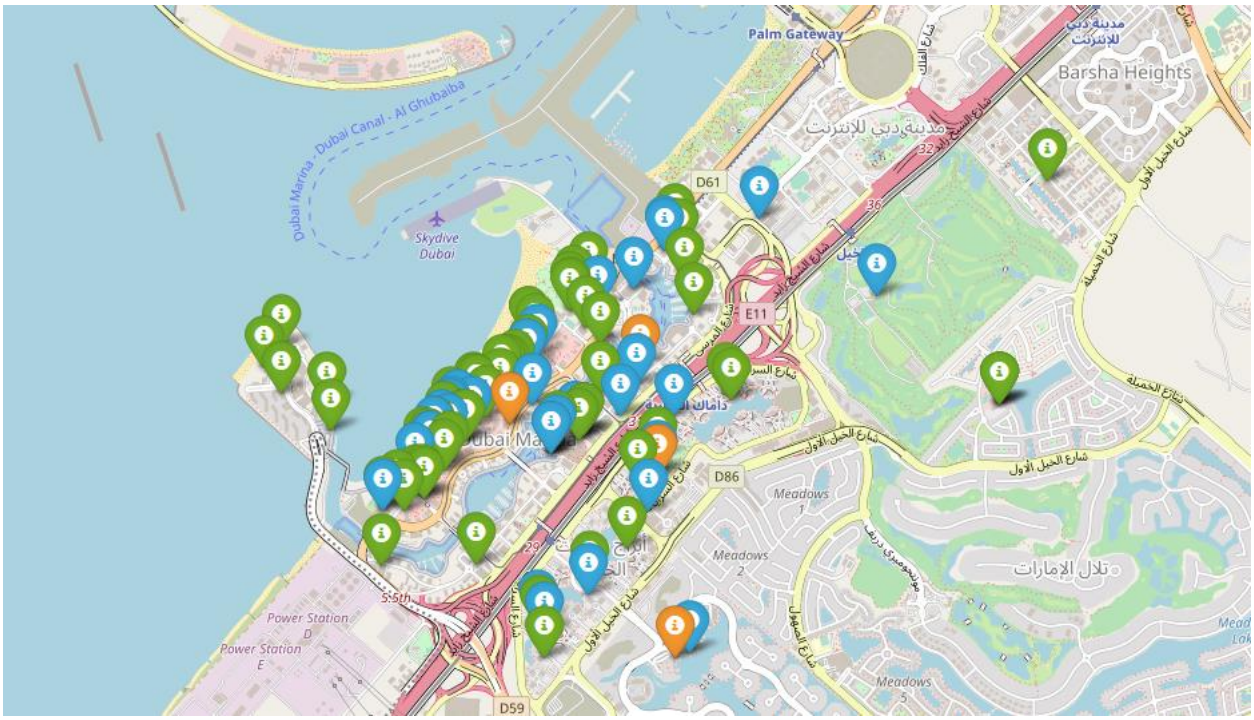
The first step is to identify rating values and plot them as a bar plot with their counts to see the most common rating.



While the above figure is pretty, it is difficult to examine the questions of interest from it. One thing that can be noted is that many venues have ratings hovering around 4.

A better way to visualize ratings is by creating "bins" for ranges of ratings and then plot them. The ratings will be divided between 4 bins:

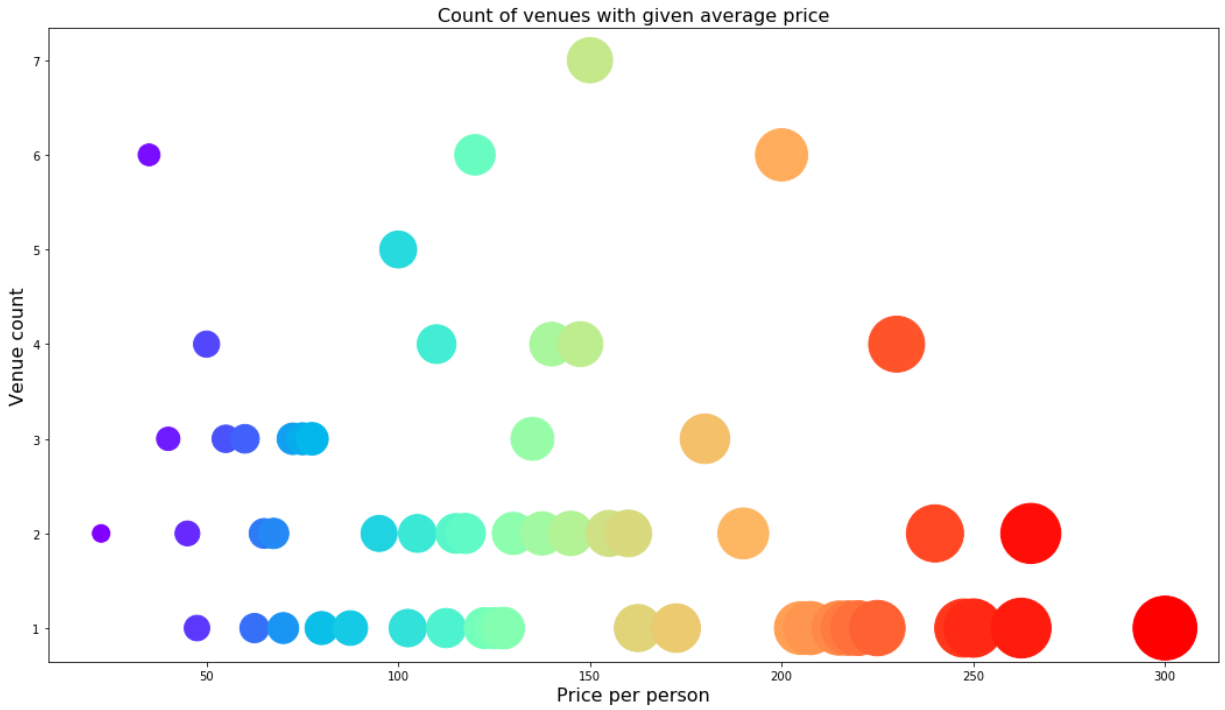
- 1 to 2
- 2 to 3
- 3 to 4
- 4 to 5



Blue (from the notebook) corresponds to a “very good” rating, while green corresponds to “good”. It appears that most venues have a rating of either “good” or “very good”.

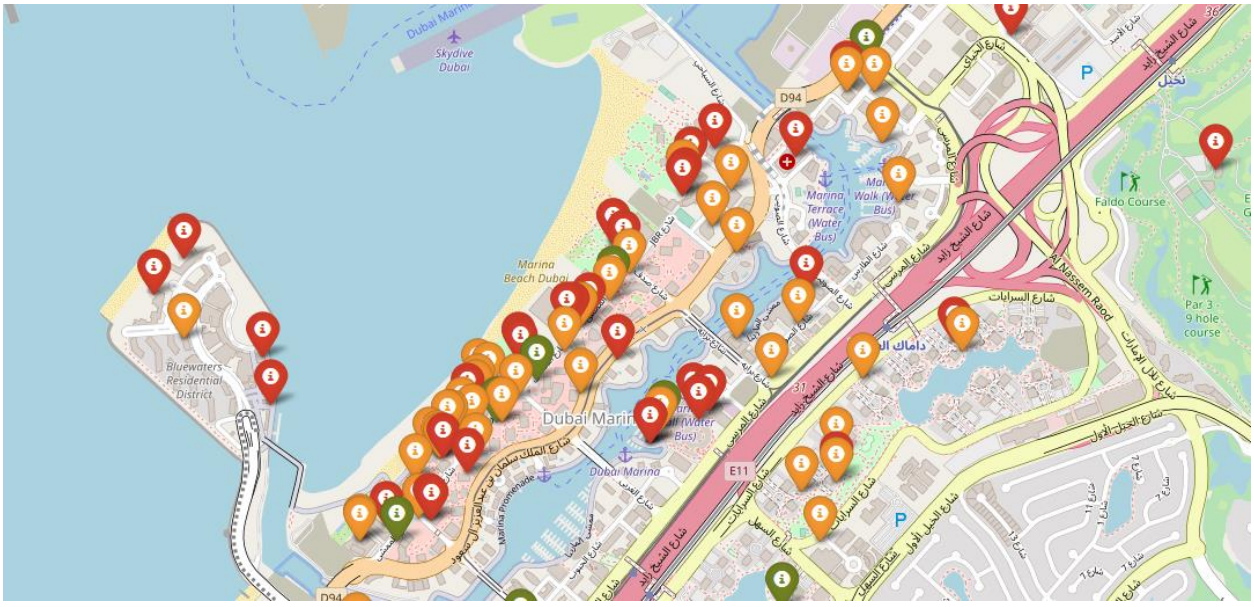
I then took a look the venues based on their prices. In the dataset, I have two price features for our venues, one is average_price which defines the average cost for one person and the other is price_range which determines the price range as defined by Urbanspoon.

I first explored the average_price using a scatter plot between of price and the count of venues with that average price. The scatterplot points are sized based on the price for additional clarity.



The analysis reveals that a large number of venues have an average price between 100 to 250 AEDs.

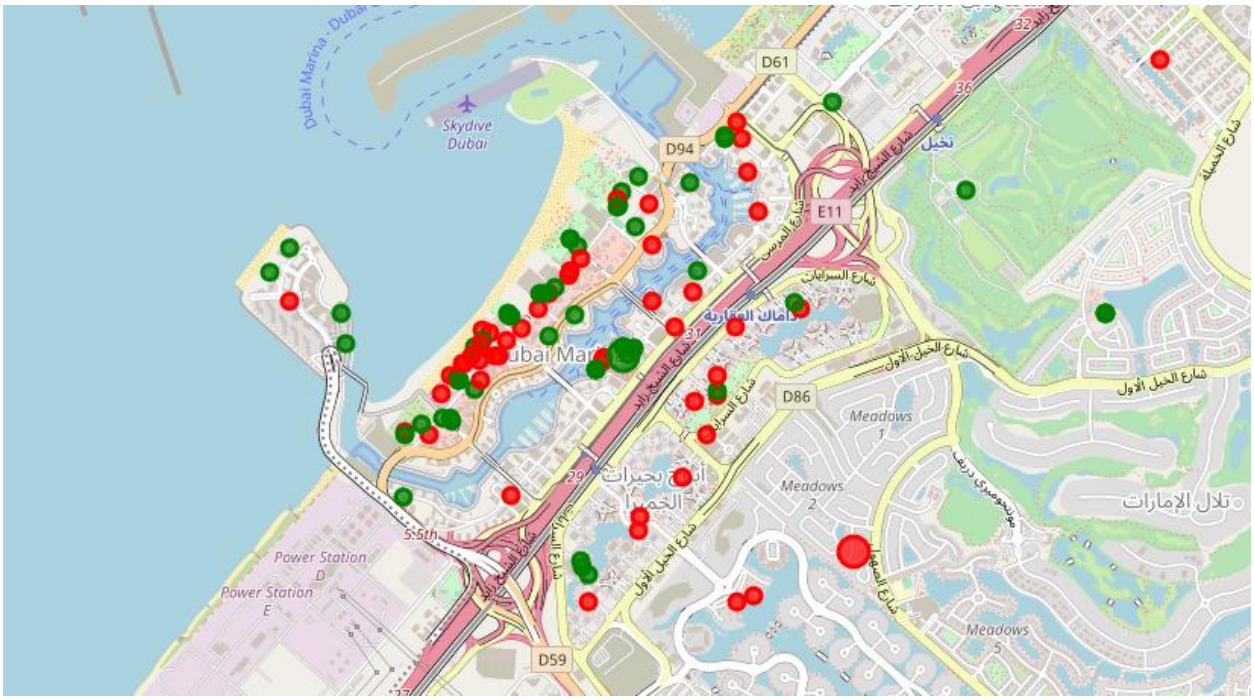
Since visitors and travelers are usually looking to go to venues that fit within their budgets. I explored the price_range column in my dataset and used it to plot the venues on a map. Venues with lower prices are shown in green and we move towards red as the price increases.



From this zoomed in perspective of the map, it appears that there are few lower priced venues; which given Dubai's emphasis on luxury as well as the location I have explored is somewhat expected.

Clustering

For the final step of my analysis, I clustered all datapoints for venues based on their price range, location. This was done so I could identify similar venues and any relationships between them. For simplicity/clarity, I will cluster the venues into two separate groups.



From this zoomed in map, it is difficult to tell what the features are for each cluster. But it appears as if there is an even geographical spread. To explore this some more, I examined each cluster separately.

Cluster 1

	cluster_labels	categories	venue	latitude	longitude	price_range	rating	address	average_price	rating_bin
0	1	Shopping Mall	Al Arrab	25.0246	55.1528	3.0	2.7	Ground Floor, Al Furjan Pavilion, Al Furjan, D...	102.5	Okay
1	1	Shopping Mall	Starbucks	25.0397	55.1980	2.0	3.7	Ground Level, City Centre Me'aisem , Al Fay Ro...	35.0	Good
2	1	Hotel	City Limits	25.0379	55.1972	3.0	3.6	Ground Level, Aloft Me'aisam, Dubai Production...	117.5	Good
3	1	Frozen Yogurt Shop	Pinkberry	25.0540	55.1709	2.0	3.9	Ground Floor, Town Center, Emirates Hills, Dubai	45.0	Good
4	1	Ice Cream Shop	Baskin Robbins - Ibn Battuta Mall	25.0421	55.1186	1.0	3.7	Tunisia Food Court, Ibn Battuta Mall, Garden C...	22.5	Good
5	1	Hotel	Lobby Lounge - Ghaya Grand Hotel	25.0341	55.2035	3.0	3.1	Lobby Level, Ghaya Grand Hotel, IMPZ, Dubai	75.0	Good
6	1	French Restaurant	PAUL Bakery & Restaurant	25.0398	55.1981	3.0	3.9	Ground Level, City Centre Me'aisem, Al Fay Roa...	105.0	Good
7	1	Restaurant	The Hamptons Cafe	25.0645	55.1478	3.0	4.3	Street 1, Jumeirah Islands, Emirates Hills, Dubai	120.0	Very good
8	1	Coffee Shop	Starbucks	25.0539	55.1711	2.0	3.6	Ground Floor, Town Center, Emirates Hills, Dubai	35.0	Good
9	1	Hotel	Costa Coffee	25.0465	55.1191	3.0	3.2	Premier Inn, Ibn Batuta Mall, Jebel Ali Villag...	50.0	Good

Cluster 0

	cluster_labels	categories	venue	latitude	longitude	price_range	rating	address	average_price	rating_bin
0	0	Beach Bar	The Glasshouse	25.0523	55.1627	4.0	3.4	Jumeirah Islands Clubhouse, Emirates Hills, Dubai	160.0	Good
1	0	Hotel	Revo Cafe - Mövenpick Hotel Ibn Battuta Gate D...	25.0418	55.1156	3.0	3.4	Mövenpick Hotel Ibn Battuta Gate Dubai, Next t...	140.0	Good
2	0	French Restaurant	Couqley French Bistro & Bar	25.0656	55.1384	4.0	4.5	Laguna Tower, Mövenpick Hotel, Cluster A, Jume...	207.5	Very good
3	0	Pizza Place	Jazz@PizzaExpress	25.0663	55.1380	3.0	4.6	Basement Level, Mövenpick Hotel, Cluster A, Ju...	137.5	Very good
4	0	Hotel	Urban Bar & Kitchen - Mövenpick Hotel Jumeirah...	25.0660	55.1381	4.0	3.9	Level B1, Mövenpick Hotel Jumeirah Lakes Tower...	150.0	Good
5	0	Greek Restaurant	Mythos	25.0750	55.1457	4.0	4.4	Basement Level, Armada BlueBay Hotel, Cluster ...	160.0	Very good
6	0	Hotel	Bistro Des Arts	25.0769	55.1406	4.0	3.9	Supermarket Level, Dubai Marina Mall, Marina P...	230.0	Good
7	0	Restaurant	Cargo	25.0761	55.1388	4.0	3.9	3rd Floor, Pier 7, Behind Dubai Marina Mall, A...	172.5	Good
8	0	Hotel	Quench - Sofitel Dubai Jumeirah Beach	25.0751	55.1320	4.0	3.1	Sofitel Dubai Jumeirah Beach, The Walk, Jumeir...	150.0	Good
9	0	Hotel	The Daily - Rove Dubai Marina	25.0696	55.1279	3.0	3.8	Al Seba Street, Dubai Marina, Dubai	145.0	Good

Inspection shows that Cluster 0 (price range 3.74) includes more higher end restaurants than Cluster 1 did (price range 2.77). Both clusters appear to be close in terms of ratings. A closer inspection of the types of venues shows that cluster 0 includes venues located in hotels while cluster 1 has more venues located in shopping malls, which could also justify the price difference.

Results and Discussion

Based on our analysis above, there are some conclusions that can be made that could be be useful to visitors to Dubai as well as to someone exploring the idea of opening up a venue.

After collecting data from the Foursquare and Urbanspoon APIs (for a radius of 10 km around the coordinates I was exploring), I got a list of 238 different venues. However, not all venues from the two APIs were identical. By inspecting their coordinates more closely I was able to combine the two datasets and remove the outliers. This resulted in a total venue count of 128 venues which is not bad for an area of 10 sq kms.

I was able to identify that the majority of venues were hotels, coffee shops and restaurants; whereas lounges were not as common in that area. This type of information could assist visitors who wish to discover Dubai in figuring out where they would like to stay (hotel) or eat based on different variables which I explored next.

While the complete range of ratings range from 1 to 5, the majority venues have ratings close to 4. This means that on average, this area of Dubai has venues that are considered to be of high quality or viewed favourably. By plotting the ratings on Dubai's map, I also noted that the highly rated venues were closest to the waterfront (Dubai Marina) which is a very popular area.

By looking at prices for each venue, we found a wide range of prices, suggesting that the particular area I was exploring in Dubai could cater to visitors with different budgets. By plotting the prices on the map, it also appeared that the most expensive venues were near the waterfront (Dubai Marina).

Finally, through clusters I was able to cluster the data into two groups, with price and type of venue being the dominant features. The two clusters had similar ratings but diverged in price range. It also appeared that the lower priced venues were located in shopping malls, whereas the cluster of higher priced venues were located mostly in hotels.

A business-oriented individual who is seeking to explore opportunities in Dubai could use the insights derived from this analysis to learn about possible opportunities (not many lounges appeared as venues) and can also understand that location will highly impact how popular a venue is. Dubai Marina is a very popular area, but it also includes venues that appeared to be mostly in hotels and that were highly priced. Setting up a business in a shopping mall does not mean that the venue will be less popular (since ratings appeared to be similar in both clusters), but the average price appeared to be more reasonable.

Conclusion

The purpose of this project was to explore an area of Dubai that visitors to the city could possibly want to visit. The venues were identified using two different data sources, Foursquare and Urbanspoon. Each dataset provided important information, and by combining the datasets, I was able to examine these venues based on their location and draw some insights regarding their ratings, price ranges among others. It appears that Dubai Marina is a highly popular area in Dubai that visitors will have a wide range of options in terms of venues to visit. Staying in a hotel in Dubai Marina would give visitors ample options in terms of dining and recreation.