Exploring Las Vegas Using Foursquare and Zomato API

Gagandeep Kaur

July 27,2020

1. Introduction

1.1. Background

Whenever a person wants to visit any city or place that is new to them they eventually look for the places to eat in addition to the hangout places. For finding hangout places there are various other options. Finding better restaurants and café is problem when there are many options and the city is very dense. Thus to find any eating place we check their ratings and reviews which are give us pretty much fair opinion about the restaurant. Now how about we getting an idea about the price range? This is what we are going to explore and analyze here. Also, we will also be combining the information about the location of the venues in the city .

Las Vegas is is the 28th most populated city in the united states and the most populated city in the state of Naveda. Las vegas is internationally renowned major resort city, known primarily for shopping, fine dine, entertainment and nightlife. Las vegas also serves as the leading financial, commercial and culture center for Naveda. This project explores various venues in Vegas and attributes the data based on user ratings and price range or average price.

To explore the information stated we have made use of two location API that is Foursquare API and the Zomato API. Using Foursquare API we were able to set the boundaries (using radius parameter) and Using Zomato API we were able to fetch the information about various venues like name, address, rating, price range and category. Using the data from the foursquare and Zomato API maps have also been plotted using Folium to make more clear visualization.

1.2. Target Audience

This project target audience that are going to visit the city. This will help people better understand the locality and choose a budgeted a hotel to stay, restaurant or cafe to eat and branded showrooms. In short this will help people to explore area more with help of provided map.

2. Data

2.1. Data Sources

To get location and other information about various venues in Las Vegas, I used two APIs and decided to combine the data from both of them together.

Using the Foursquare's explore API (which gives venues recommendations), I fetched venues up to a range of 3 kilometers from the center of Las Vegas and collected their names, categories and locations (latitude and longitude).

Using the name, latitude and longitude values, I used the Zomato search API to fetch venues from its database. This API allows to find venues based on search criteria (usually the name), latitude and longitude values and more. Given that the data from the two APIs did not align completely, I had to use data cleaning to combine the two datasets properly.

From Foursquare I retrieved the following information
1. Name: Name of the Venue
2. Category : Type of the venue(eg. cafe, restaurant etc.)
3. Latitude : Latitude Value of the venue
4. Longitude : The longitude value of the venue

From Zomato API I restrieved the following Information:
1. Name : The name of the venue
2. Address: The complete address of the venue.
3. Price Range : The price range of the venue as defined be the Zomato.
4. Rating : The ratings as provided by many Users.
5. Latitude : The Latitude value of the venue.
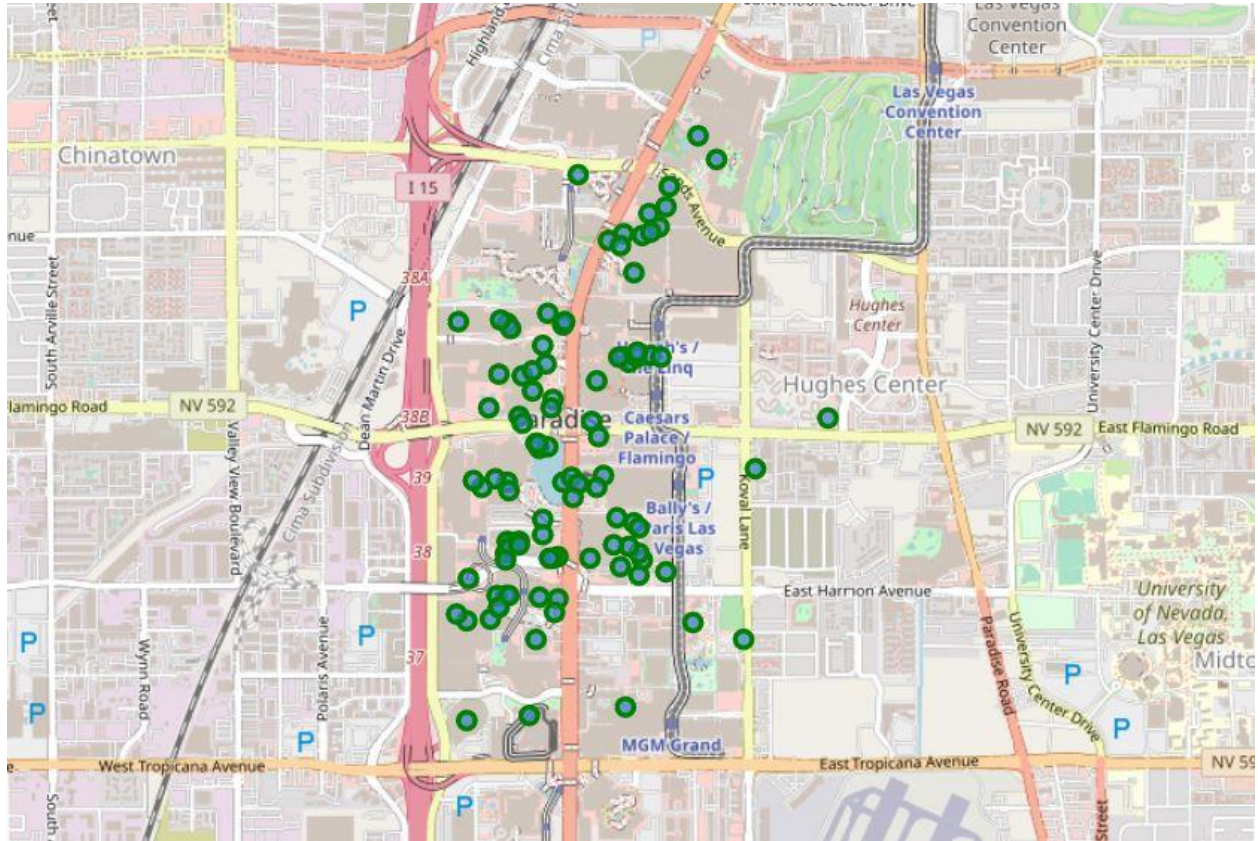6. Longitude : Longitude value of the venue.

2.2. Data Cleaning
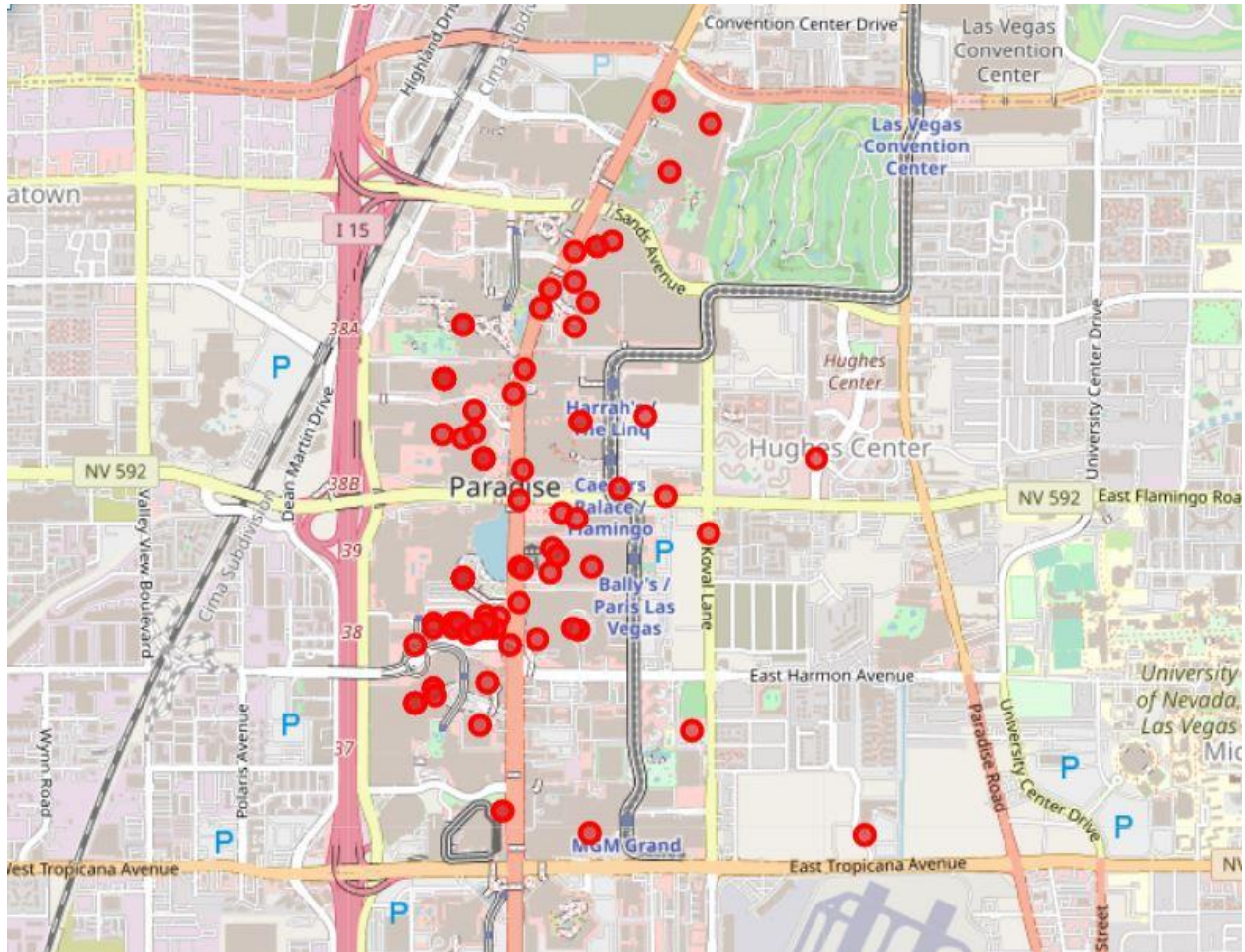

Figure 1-Venues by Foursquare API

Figure 2 Zomato Venues

From above figures, we can clearly see that some venues from the two APIs do not align with each other. Thus, I decided to combine them using their latitude and longitude values.

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 all corresponding venues from the two datasets that had their latitude and longitude values different by more than 0.0004 from one another. Thus, I rounded both the latitude and longitude values up to 4 decimal places. Then, I calculated the difference between the corresponding latitude and longitude values and saw if the difference was less than 0.0004 which should ideally mean that the two locations are the same. This removed many outliers from the two datasets. Once this was done, I observed that there were still some venues which were not correctly aligned.

As a final Dataset we are only left with 29 venues as described in figure below.

	categories	venue	latitude	longitude	price_range	rating	address
0	Fountain	Cafè Bellagio - Bellagio	36.1117	-115.1753	3.0	3.6	Bellagio, 3600 South Las Vegas Boulevard, The ...
1	Casino	Bacchanal Buffet - Caesars Palace	36.1161	-115.1744	4.0	4.5	Caesars Palace, 3570 Las Vegas Boulevard South...
2	French Restaurant	Mon Ami Gabi	36.1121	-115.1694	4.0	4.6	3655 Las Vegas Boulevard South 89109
3	Hotel	Bacchanal Buffet - Caesars Palace	36.1161	-115.1744	4.0	4.5	Caesars Palace, 3570 Las Vegas Boulevard South...
4	Hotel	Cafè Bellagio - Bellagio	36.1117	-115.1753	3.0	3.6	Bellagio, 3600 South Las Vegas Boulevard, The ...

3. Methodology and Exploratory Data Analysis

As a first step, I retrieve the venues in Las Vegas from Foursquare and Zomato APIs. I extract the location data from the Foursquare API for all venues up to a distance of 3 kilometers from the center of Las Vegas. Using this, I fetch the venue information including price and rating data from Zomato API.

Using data cleaning, the dataset from the two APIs will be combined based on the venue names, latitude, and longitude values. One to one matching and careful data inspection would be used to remove any remaining outliers such as multiple venues at the same location from the two datasets. The final data will include the venue name, category, address, latitude, longitude, rating and price range.

Using this dataset, I begin by analyzing the top venue types that exist in Las Vegas. I will then explore the venues on maps. This will allow us to better understand the location of various venues and the places where many venues co-exist and create place worth visiting. I'll also explore the venues based on the ratings and price range of various venues. The venues will be plot using proper color coding such that a simple glance at the map would reveal the location of the venues as well as give information about them. I aim to identify places which can be recommended to visitors based on their rating preferences. I'll also cluster the venues and see if we can draw meaningful information out of what kind of venues exist in Las Vegas.

As a final step, I will analyse these plots and try to draw conclusions on what places can be recommended to visitors. I'll discuss my findings and any inferences I can draw.

3.1. Categories

I begin my analysis by taking a look at the various categories of venues that exist in Las Vegas.
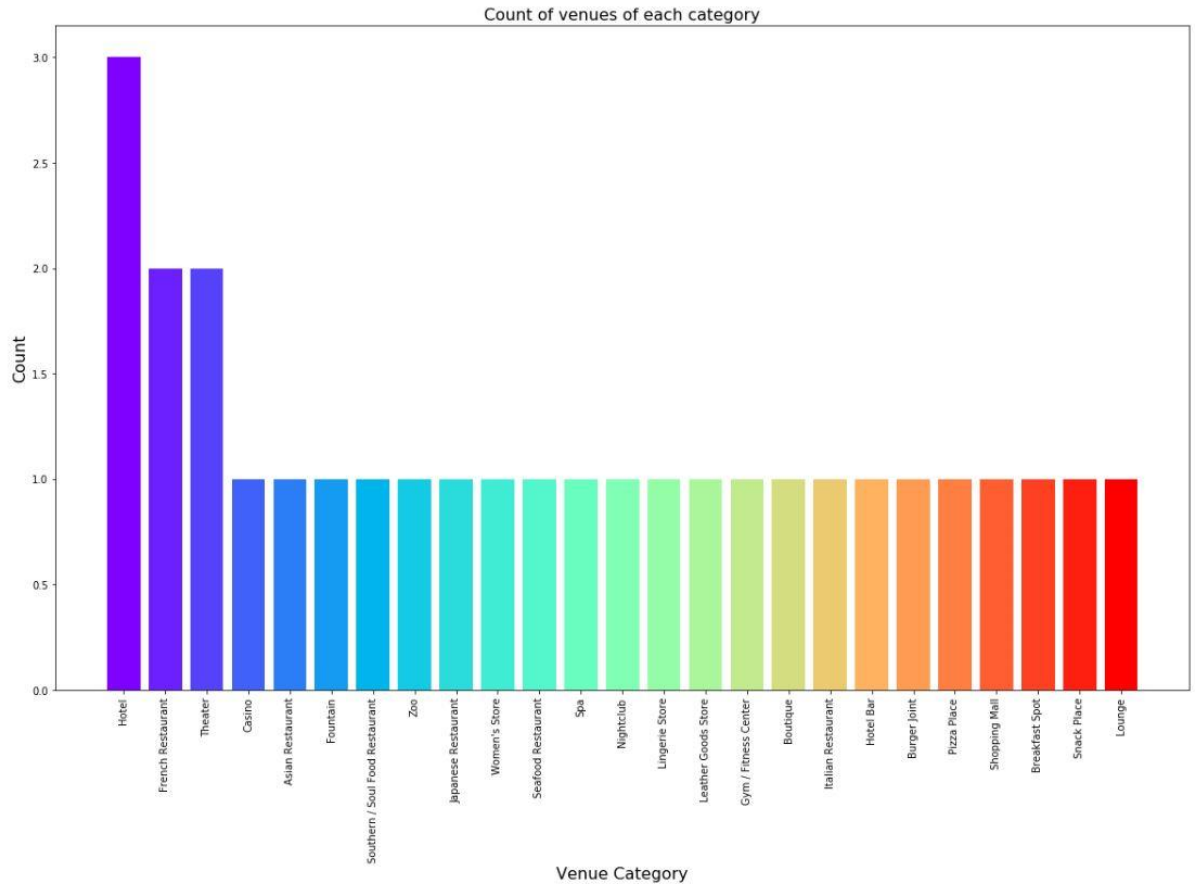
Figure 3 Count of venues of each category

From the above graph we can infer that majorly there are **Hostels** next we have **Theatre** and **French Restaurants**.

3.2. Rating

Next, I'll explore the ratings of various venues in Las Vegas. I decided to plot a bar chart with x-axis as the rating from 1 to 5 and the y-axis as the count of venues with that rating. I decided to plot the bar chart to see what average rating venues get in Chandigarh. This can be seen in following.

We can in the picture that majority of places that can be visited are **Hotels** followed by **Theaters.**

I followed this information by plotting the venues on the map of Las Vegas. The venues that were rated below 3 were marked by red and orange while the venues that were rated more than or equal to 3 were plot as green and dark green. Taking a look at figure 6 reveals the same results as the bar plot.
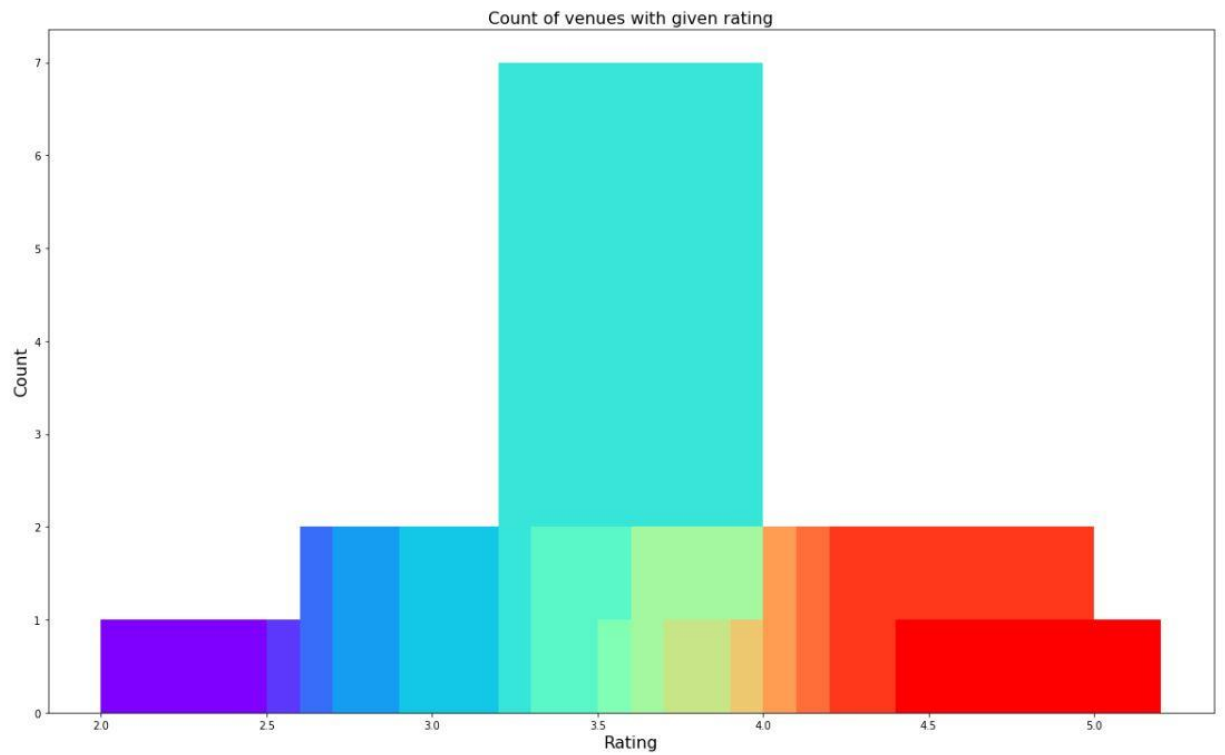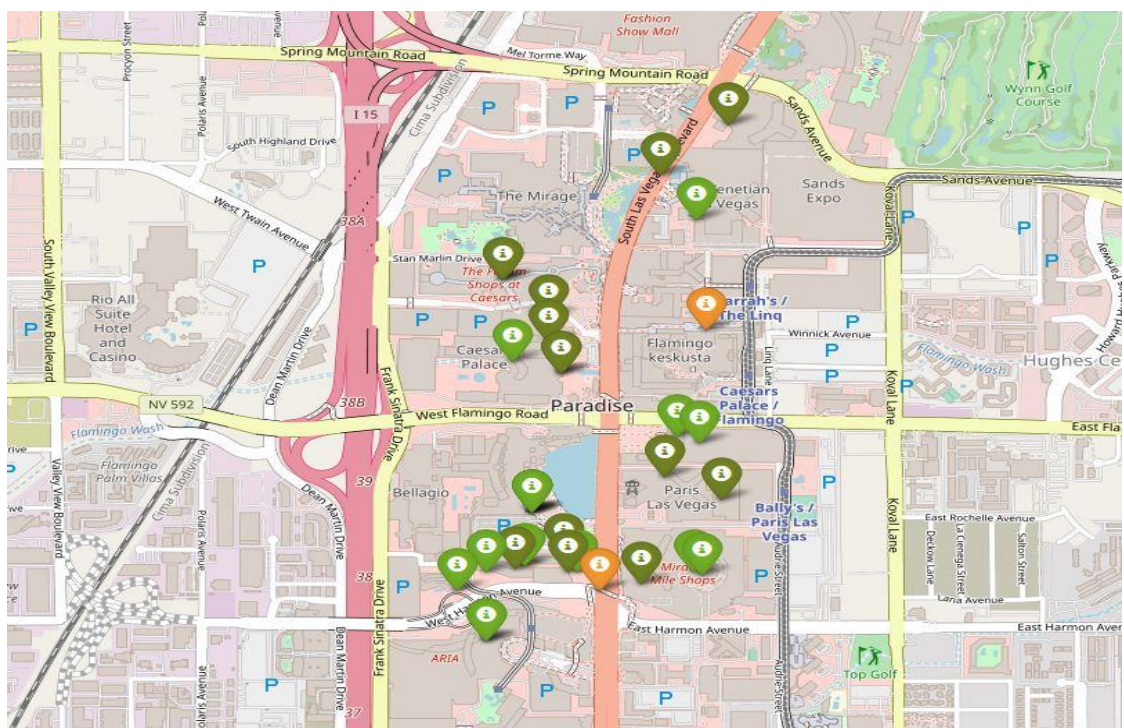
Figure 4 Count of venues with given Rating

Figure 5 Plot of venues with different ratings.

3.3. Clustering

Finally, I cluster all the venues based on their price range, location and more to identify similar venues and the relationship amongst them. I used KMeans clustering and decided to cluster the venues into two separate groups.
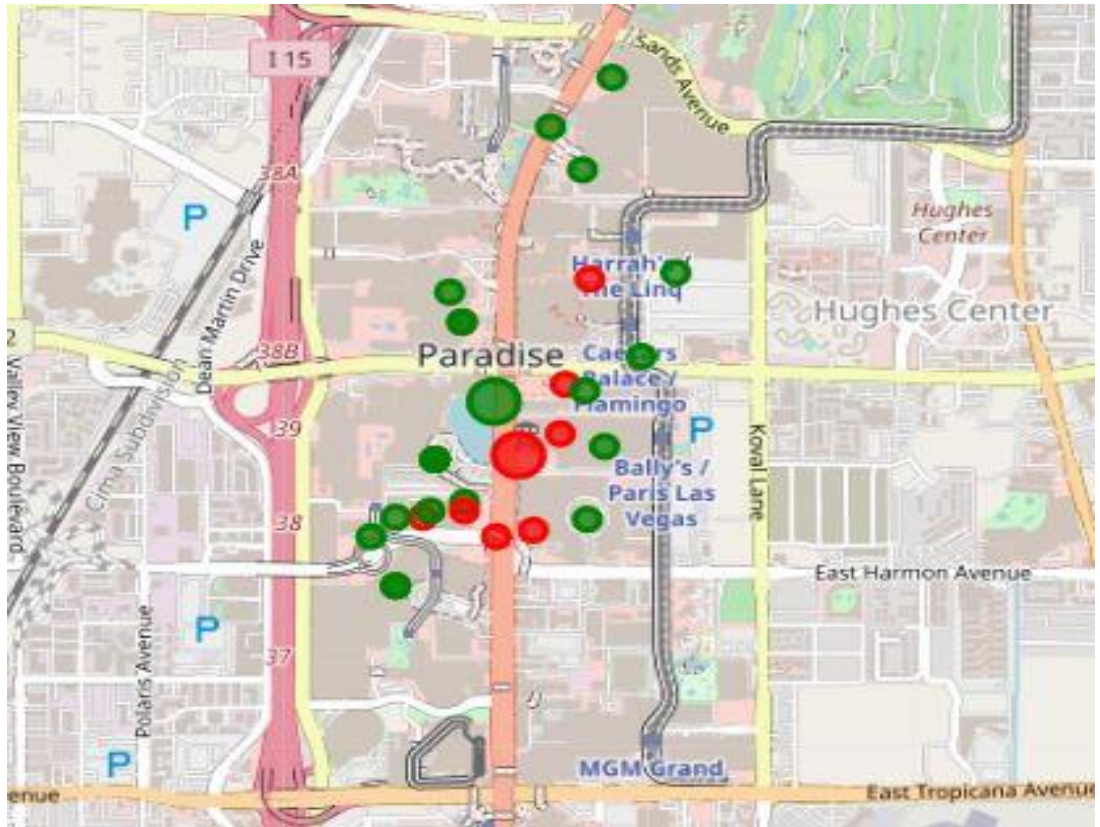

Figure 6 Venue Clusters

In figure 6, we see the two clusters:

- The first cluster (green) is spread across the whole city and includes the majority venues. These venues have mean price range of 3.64 and rating spread around 3.91.
- The second cluster (red) is very sparsely spread and has very limited venues. These venues have mean price range of 1.14 and rating spread around 3.26.

4. Results and Discussions

Based on the analysis done above we came to the fllowing results:

- On retrieving information from Foursquare and Zomato API we get the 100 venues in about radius of 3km.After cleaning the data we get only 32 venues
- We infer that from all the venues majority of them were **Hotels** followed by **Theaters and French Restaurants**.
- While the complete range of ratings range from 1 to 5, the majority venues have rating between 3.0 and 4.0, which means they provide good quality of service.
- Finally Through Clusters we identified that there are many venues which were relatively low priced but have an average rating of 3.91. On the other hand, there are few venues which are high priced and have average rating of 3.26.

5. Conclusion

The very purpose of this project was to explose the visiting places in the city of Las Vegas or simply Vegas. The information about the venues have been fetched using two APIs namely Foursquare and ZOmato. Using that information we were able to plot the map and get more clear view of the places that where they are located.The maps reveals that **Jockey Club, Ceaser Palace and Miracle Mile Shops** are the major areas where people visit.