Choosing Restaurant Location

Saad Elshazly

July 12, 2020

1. Introduction

1.1 Background

Riyadh city has many neighborhoods, each neighborhood has multiple venues, the venues are categorized where multiple venues fall into the same category. Venue Category may be a Coffee shop, Gym or Restaurant and so on.

1.2 Problem

If you need to start your business and you don`t know the type of business that is favorite to people and what is the location that will make it successful, so it is not easy mission to choose the previous criteria, in this case you have to choose the best business type as well as the best location.

1.3 Interest

Our analysis focuses in restaurant Business, this helps people who are interested in restaurant business and they want to know the most attractive restaurant categories based on neighborhoods. We will focus on about 19 neighborhoods of Riyadh city as below:

1. Zahrat al Badiah District ظهرة البديعة
2. Al Aziziyah District حي العزيزية
3. Al Nahdhah District حي النهضة
4. Al Malaz District حي الملز
5. Al Aqeeq District حي العقيق
6. Al Muhammadiyah District حي المحمدية
7. Al Rabiea District حي الربيع
8. Al-Ma'athar Al-Shimali District حي المعذر
9. As Sulaymaniyah District حي السليمانية
10. Almuruj District حي المروج
11. Ghirnatah District حي غرناطة
12. King Fahd District حي الملك فهد
13. Al Yasmin District
14. Al Malga District حي الملقا
15. King Faisal District حي الملك فيصل
16. Al Olaya District حي العليا
17. Al Ghadir District حي الغدير
18. Hittin District حي حطين
19. Al Murabba District حي المربع

2. Data Gathering and Cleaning

2.1 Data Source

We collected location data from foursquare API here . First we got all categories here, and then we filtered those categories into just food categories. In The next step we collected all venues for each category here . To assign each venue with its neighborhood we used Nominatim. Finally we collected more data about each venue here.

2.2 Data Cleaning

Data cleaning started with venue data collected for each category where JSON Data converted into dataset that was having many duplicates, those duplicates has been removed and columns has been reordered where venue id, name, category and has park features come first then location data such as latitude, longitude, etc.

Features names were not clear, so they renamed. Because we interested in just restaurants we have removed venues data that is not restaurants.

The second stage on cleaning data was data about neighborhoods where we collected them to assign each venue with its neighborhood, so in this dataset we assigned each venues id with a neighborhood name; we removed neighborhoods with less than 13 restaurants, then venue dataset and address or neighborhood dataset have been merged to produce one dataset having venues data plus neighborhoods.

The third step in cleaning our data was venues details, many features removed and just 4 features has been excepted (Id, Verified, Likes, Rating) , null values in rating columns filled with the mean of all ratings, the final details dataset merged with venue dataset by venue Id.

The final step was cleaning the merged dataset that contains all data we need for analysis, firstly duplicated data removed, Verified column not assigned values filled with mode, rating data rounded up to one decimal value, likes column data type converted from float into integer.

3. Explanatory Data Analysis

3.1 analysis overview

The target from the analysis, as said before, is to help people who intend to open their business as restaurant and after cleaning data we have about 19 Neighborhoods. To ease analyzing data and get our target we will cluster or group neighborhood into 3 groups. Each group will be analyzed to find the most common ten restaurant categories, the best rated and the most liked ones. We have grouped the neighborhoods into 3 groups by using K-MEAN Clustering Algorithm,(Figure 1); each group has set of neighborhoods where groups are colored in the map as green, blue and red for groups 1, 2 and 3 consecutively.

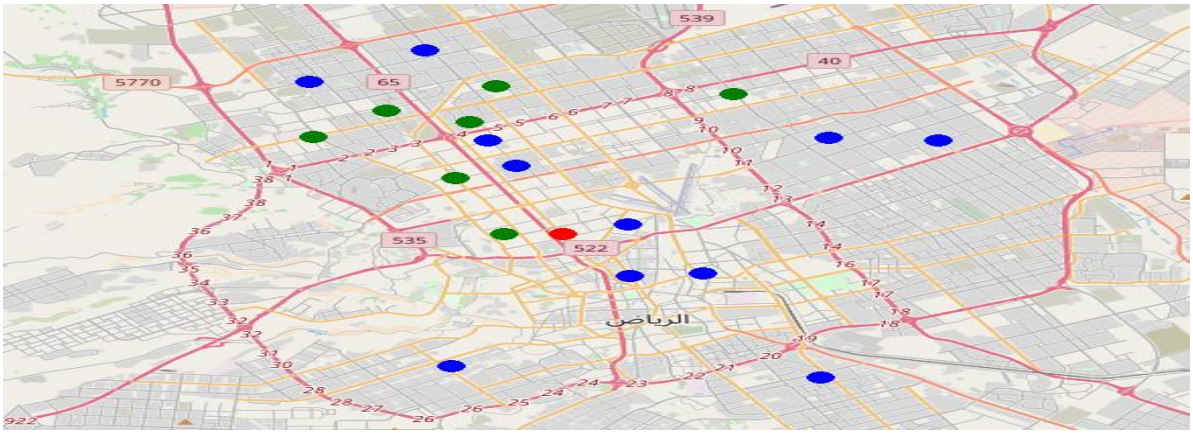

Figure 1. Map to show grouped neighborhoods in Riyadh City

3.2 First Group Analysis

3.2.1 Overview and Statistics

We have about 39 restaurant Categories with total 167 restaurants in 7 neighborhoods, the most common restaurant category is the French, and also Ghirnatah is the most frequent Neighborhood, most of restaurants in the first group are not verified.

3.2.1 What are the Neighborhoods in the first group?

In figure 2, we see that there are 7 Neighborhoods and Ghirnatah and Al Aqeeq are the most Crowded with restaurants, Al Ma`athar Alshimali and Al Ghadir are less crowded with restaurants.

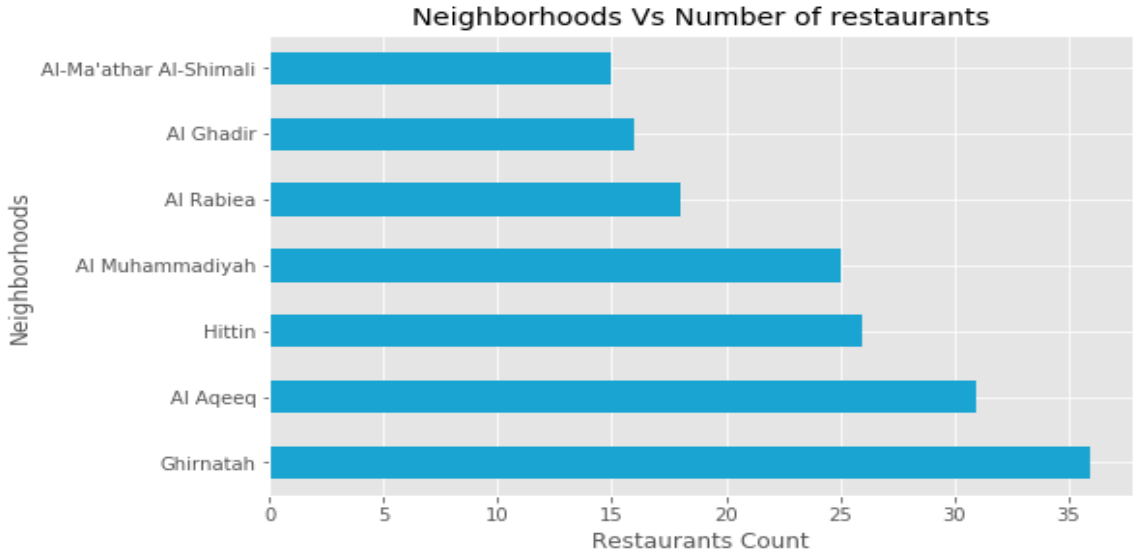

Figure 2. Showing number of restaurants for each Neighborhood in the first group.

3.2.2 What are the most common restaurant Categories in the first group?

In the word cloud image we see that the most common restaurant type has bigger font size, so we can observe that French, Mexican, American and Indian are the most common categories in the first group.

Figure 3. Show the most common Categories in the first group of neighborhoods

3.2.3 What is the most rated restaurant Categories in the first group?

Restaurant rating is a very significant feature in our analysis and it determines how people like and what is the degree of likeness of a specific category, so people in the first group of neighborhoods prefer Lebanese and Japanese restaurants and less preference for American food.

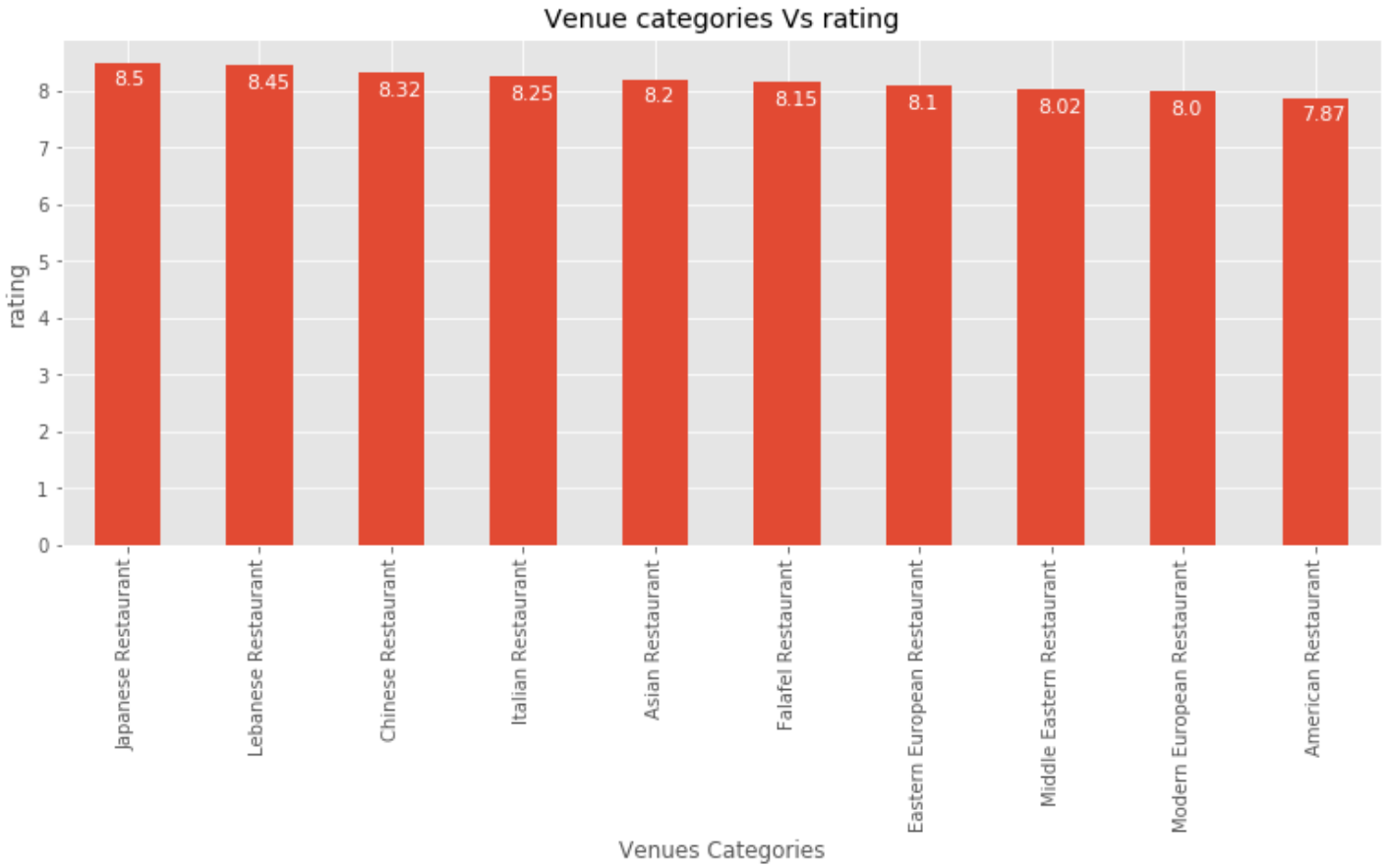

Figure 4. Show the most ten rated categories in the first group of neighborhoods

3.2.4 What about rating distribution?

The difference in rating is not big between restaurants in all neighborhoods but there are some outliers that didn`t stick to normal rating distribution such as there is a restaurant with about 5.5 rating and in the other side there is 8.5 rating.

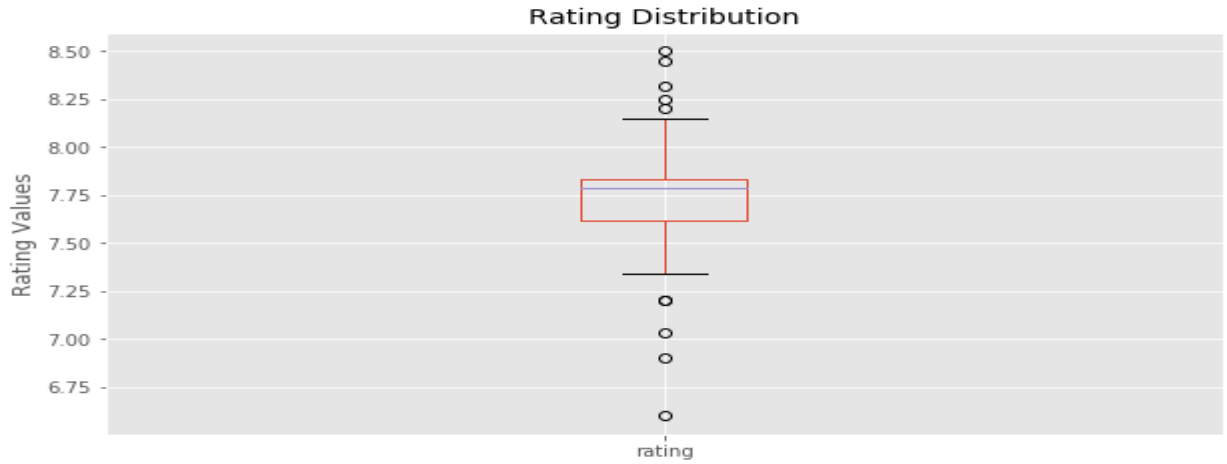

Figure 5. Show rating distribution between all restaurants in the first group of neighborhoods

3.2.5 What is the most liked categories in the first group?

Like is a good feature in the analysis but it may be affected by the count of the restaurants in as specific category, So French and American are the most liked categories, Japanese and Turkish have less likes.

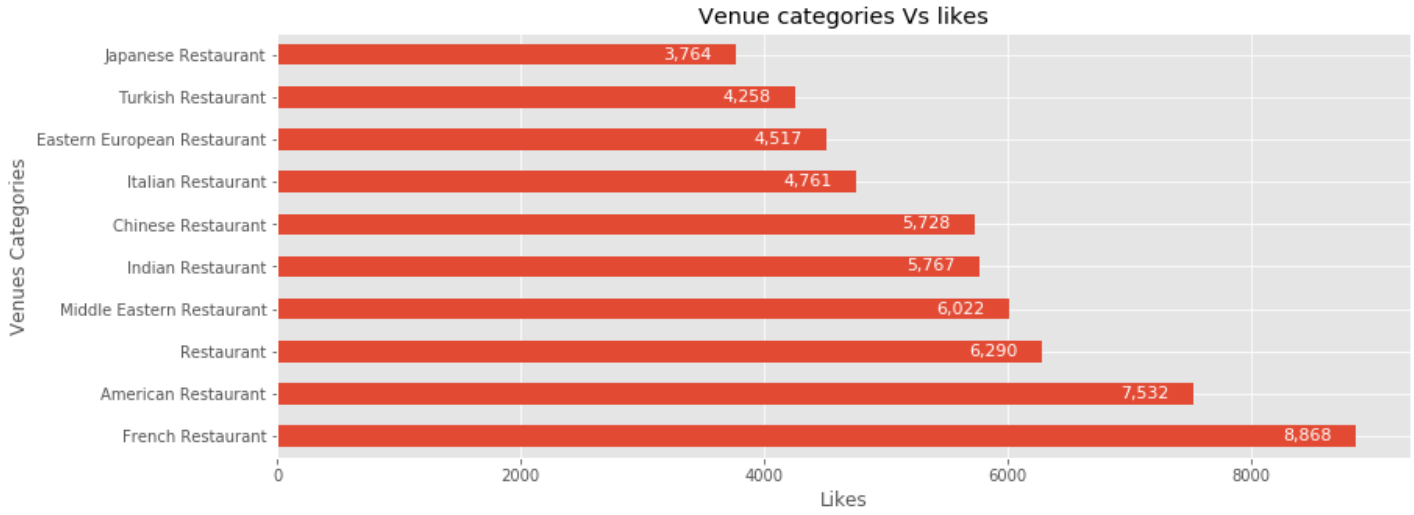

Figure 6. Show the best liked restaurant categories in the first group

3.3 Second Group

3.2.1 Overview and Statistics

We have about 47 restaurant Categories with total 215 restaurants in 11 neighborhoods, the most common restaurant category is the Kebab Restaurant, and also As Sulaymaniyah is the most frequent Neighborhood, most of restaurants in the first group are not verified.

3.3.1 What are the Neighborhoods in the Second group?

In figure 7, we see that there are 11 Neighborhoods where Yasmeen and Al Sulaymaniyah are the most Crowded with restaurants, Zahrat Al-Badiah and King Faisal are less crowded with restaurants.

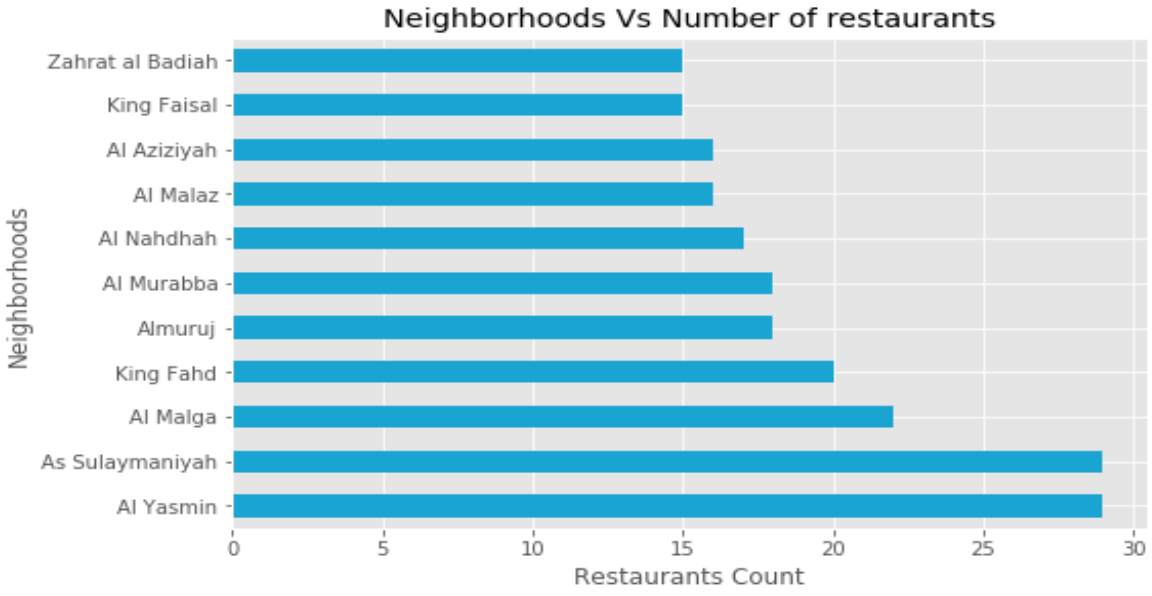

Figure 7. Number of restaurants for each Neighborhood in the second group

3.3.2 What are the common restaurant types in the Second group of neighborhoods?

In the word cloud image we see that the most common restaurant type has bigger font size, so we can observe that Kebab, Seafood, Falafel and Italian are the most common restaurant categories in the second group.

Figure 8. Show the most common Categories in the second group of neighborhoods

3.3.2 What are the best rated restaurants in the second group?

Restaurant rating is a very significant feature in our analysis and it determines the degree of likeness of a specific category, so people in the second group of neighborhoods prefer South Indian and English restaurants and less preference for Falafel restaurants.

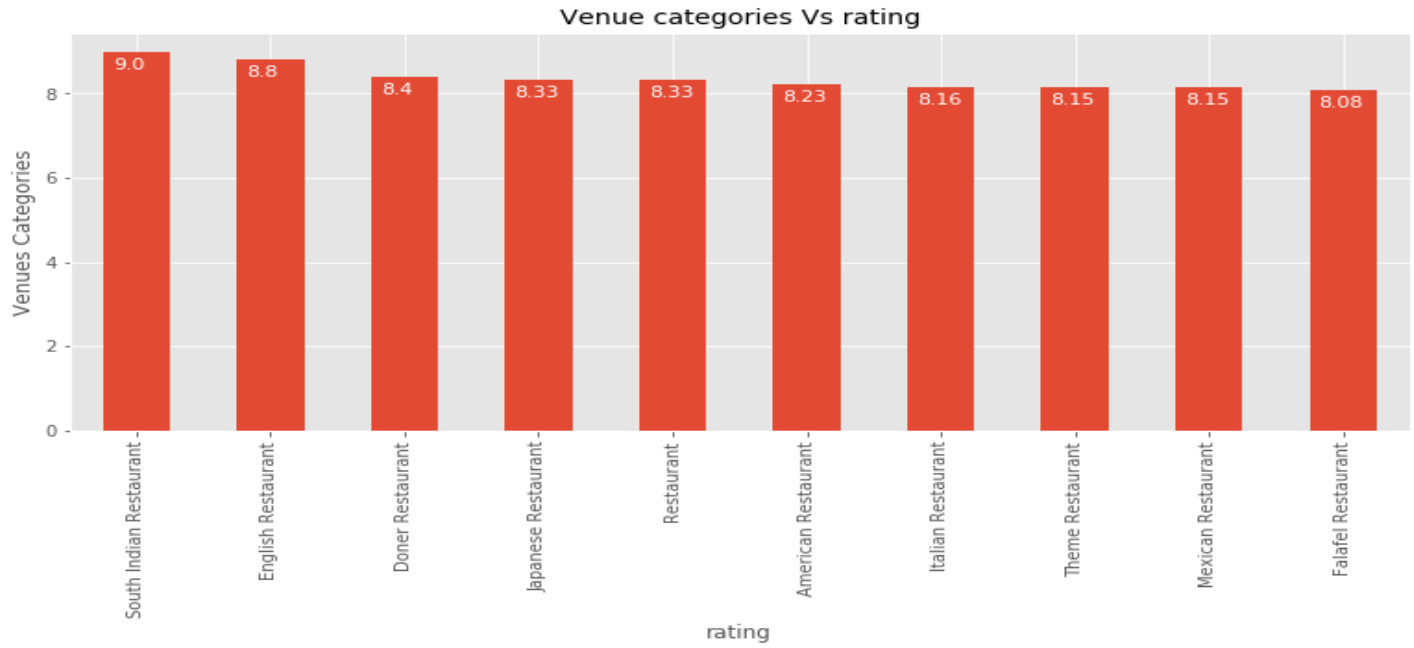

Figure 9. Show the most ten rated categories in the second group of neighborhoods

3.3.3 What are the most liked restaurants?

Like is a good feature in the analysis but it may be affected by the count of the restaurants in a specific category, So Japanese and Italian are the most liked categories and French and fast-food have less likes.

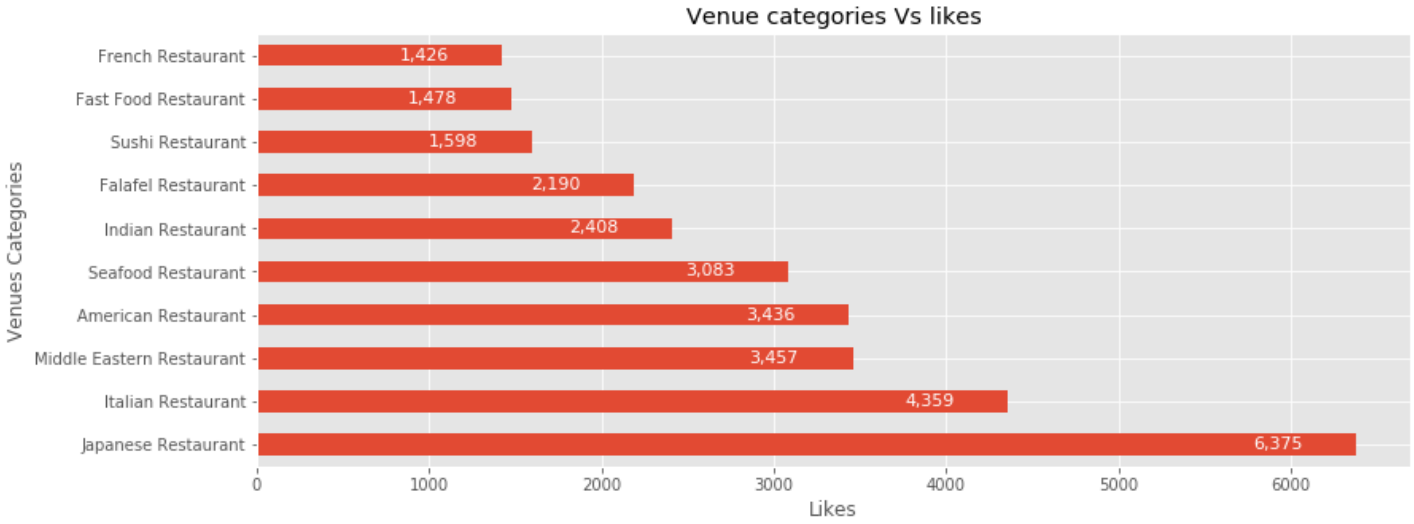

Figure 10. Show the best liked ten restaurant categories in the second group

3.4 Third Group

3.4.1 Overview and Statistics

We have about 53 restaurant Categories with total 99 restaurants in 1 neighborhood, Al Olaya Neighborhood, the most common restaurant category is the Italian Restaurant, and most of restaurants in the first group are not verified.

3.4.2 What are the Neighborhoods in the Third group?

There is only one Neighborhood in third group, **Al Olaya**

3.4.3 What is the common restaurant type in the third group?

In the word cloud image we see that the most common restaurant type has bigger font size, so we can observe that French, Japanese, Middle-Eastern, Italian and Eastern-European are the most common restaurant categories in the second group.

Figure 11. Show the best liked ten restaurant categories in the third group

3.4.4 What are the best rated restaurants in the third group?

Restaurant rating is a very significant feature in our analysis and it determines the degree of likeness of a specific category, so people in the second group of neighborhoods prefer Chinese and Fast-food restaurants and less preference for Indian and Eastern-European restaurants.

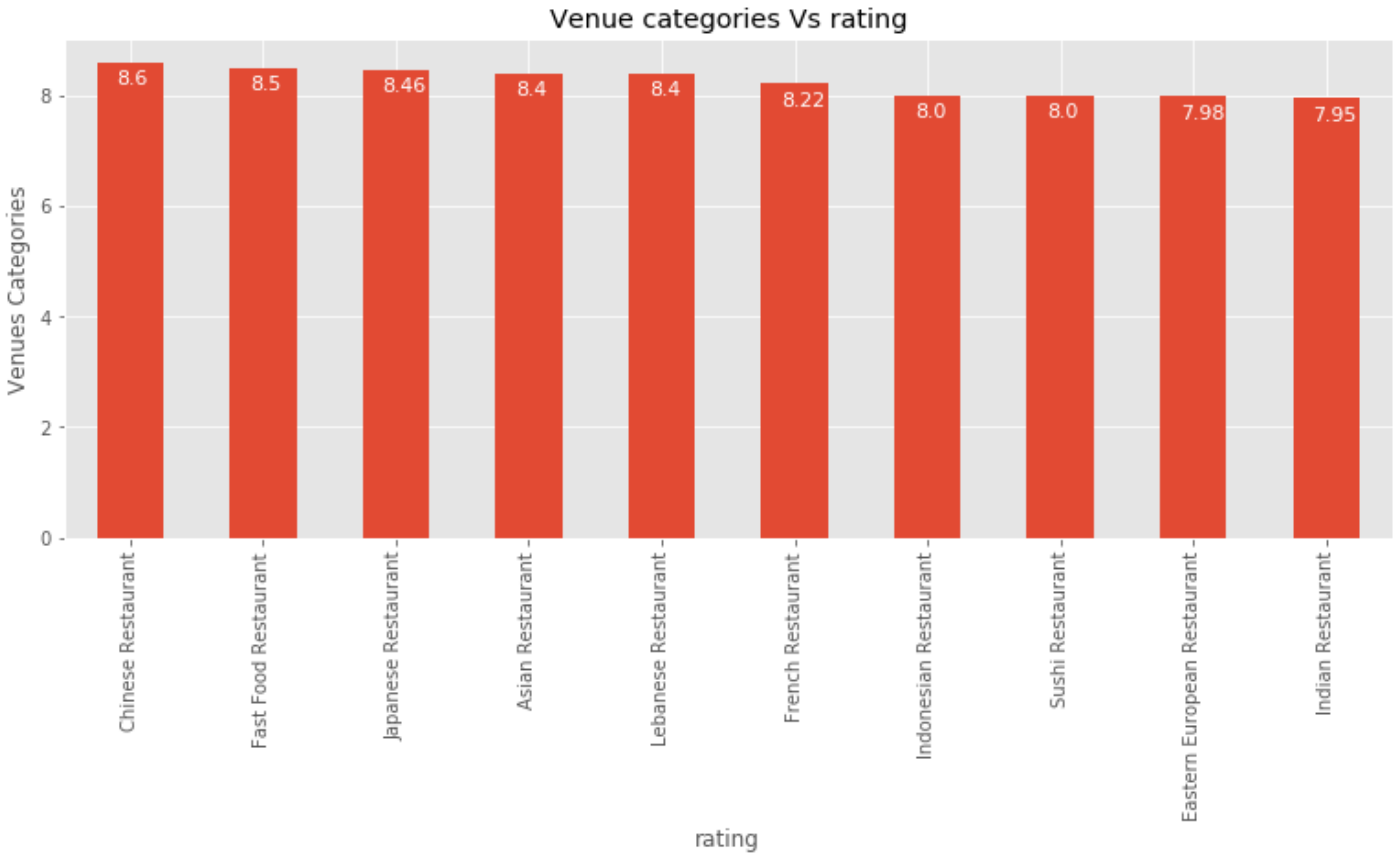

Figure 12. Show the most ten rated categories in the third group of neighborhoods

3.4.5 What are the most liked restaurants?

Like is a good feature in the analysis but it may be affected by the count of the restaurants of a specific category, So Japanese and Italian are the most liked categories and Asian and Chinese have less likes, and there is a big diverse in likes between categories in the third group.

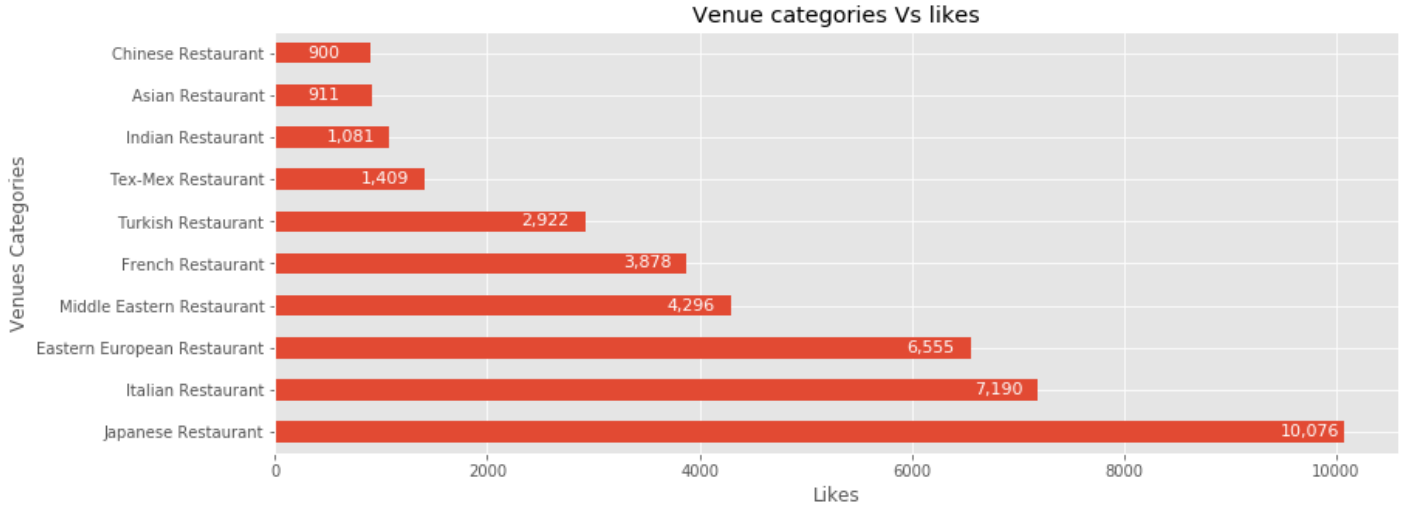

Figure 13. Show the best liked ten restaurant categories in the third group

4. General analysis

4.1 Is the rating affected by verification?

Verification feature indicating whether the owner of this business has claimed it and verified the information and as we see that the rating affected by verification.

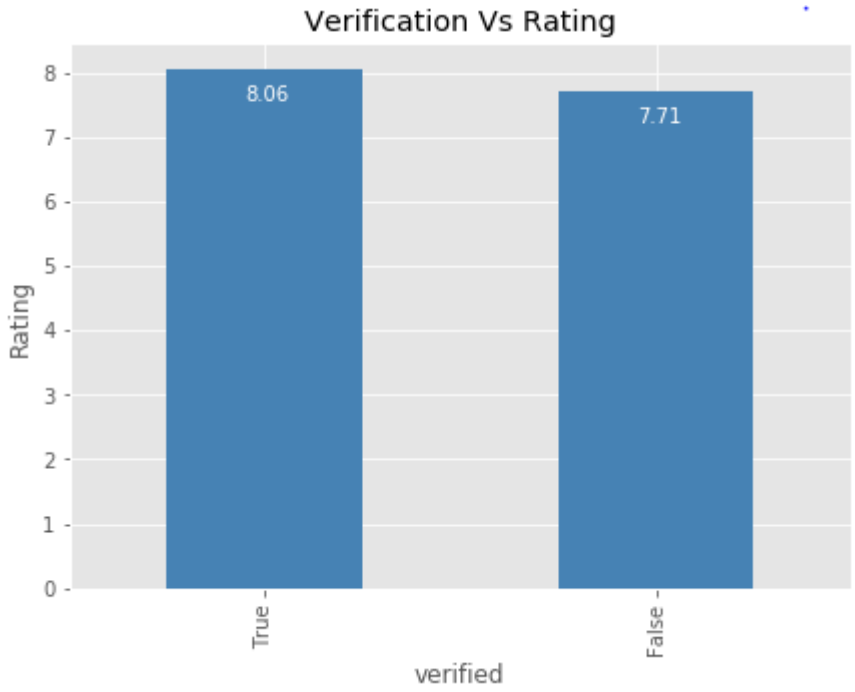

Figure 14. Showing verification effect on restaurants rating.

4.2 Is there a relation between likes and number of restaurants?

Figure 15. Showing relation between likes and number of restaurants for each category.

4.3 What about relation between rating and number of restaurants?

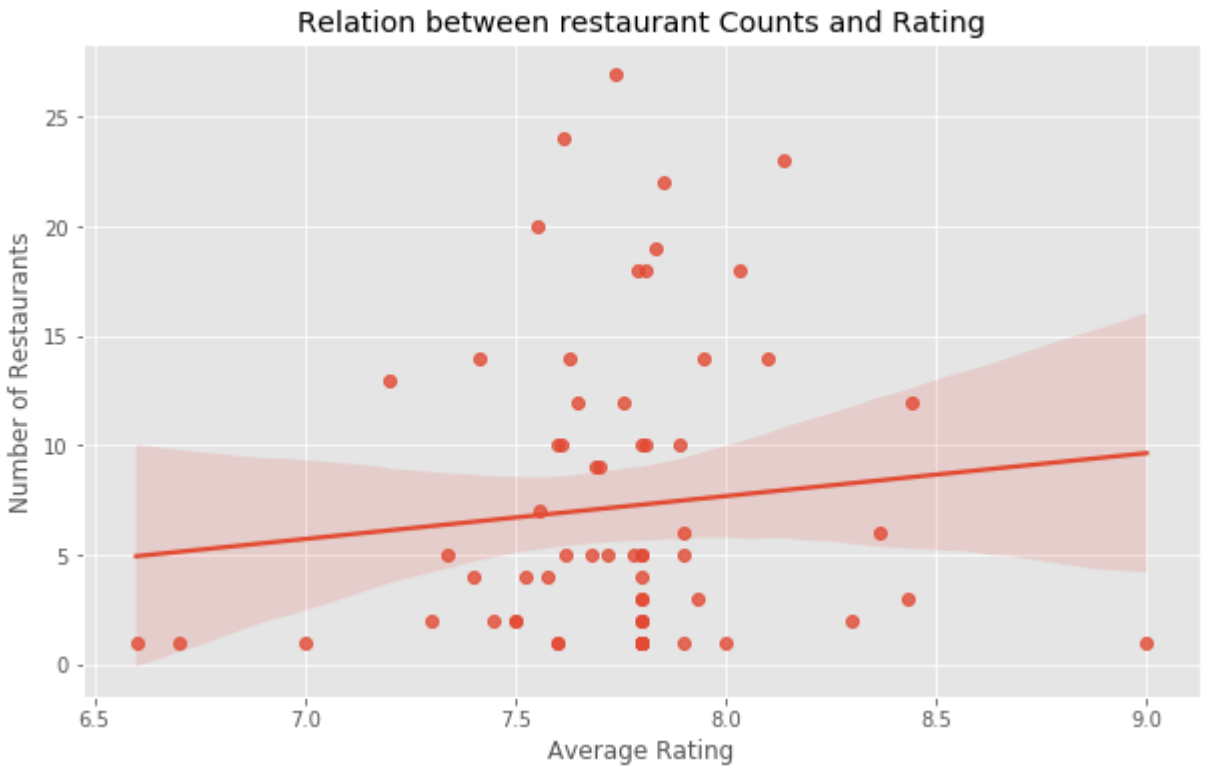

Figure 16. Showing relation between rating and number of restaurants for each category.

5. Result

As the above analysis for each group of neighborhoods we can recommend some restaurant categories for each group.

For the first group, there are some restaurants are good rated and liked despite they are not common such as Japanese, Lebanese, Chinese, Asian, Falafel, Eastern European and Modern European Restaurants. So we recommend any type of restaurants for the previous categories in any neighborhood of the first group.

For the second group, the best rated and liked and not common restaurants are South Indian, English, Doner, Japanese, American, Theme and Mexican Restaurants. So they are the best choice restaurants the second group of neighborhoods.

For the third group, the best rated and liked and not common restaurants are Chinese, Fast Food, Asian, Lebanese, Indonesian, Sushi and Indian restaurants. So they are the best choice restaurants to open in Al-Olaya neighborhood.

In general, you have to take in consideration that your rating will be affected by you restaurant verification where verified restaurant will get better rating than not verified one (figure 14).

5. Conclusion

From our analysis we observed that there is a good positive relationship between the number of restaurants and number of likes. Which is mean that the result will be biased to the restaurant

that has more branches than others, so we didn`t depend totally on likes in selecting best restaurant categories (figure 15).

There is a very week relationship between number of restaurants for specific category and rating which makes us confident about or result where we depended on rating to select the best restaurant which it will not be biased to specific restaurant category because specific restaurant have more branches than another one(Figure 16).

6. Future Direction

Our analysis was focuses on 19 neighborhoods in just one city (Riyadh) in Saudi Arabia, so the analysis will be more effective if it generalized on the entire cities of the country.

This will help people all over the country who are interested in restaurant business to choose the best restaurants type in the best location.

The flow of analysis can be applied on any type of business not just restaurants; it can be coffee shop, Gym, etc.