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 falls 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 start it.

1.3 Interest

Our analysis focuses in restaurant Business, this services 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 follow:

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 the

all venues for each category here . To specify 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 comes 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 clean the merged dataset that contain all data we need for analysis, firstly duplicated removed, Verified columns null 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, so 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 frequent ten venue categories, the best rated and the most liked ones.

3.2 Group Analysis

We have grouped the neighborhoods into 3 groups, 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.