Predicting the Popular Areas in the city

Shilpi Sirohi

02/01/2020

1. Introduction

1.1 Background

Tourism is the most promising industry which has huge potential to grow. The new generation wants to travel, explore the world. Also, they are tech savvy. They don't want to just travel, they want to experience the culture, explore the local cuisines, visit the local markets. Adding the predictive model will certainly help the visitors to plan the trip more efficiently. Adding the new model will help the website or the mobile application to get more traffic. So, this model will be advantageous for the business as well as for the visitors.

1.2 Problem

It is easier to find the popular restaurants, tourist spots online. But to plan the trip, one has to calculate the travelling time, cost, etc. If one plans to visit two places which are on opposite corner of the city, then so much of the time will go in travelling. This won't leave the right impression of the city on the visitor. Also, visitor is unable to enjoy the stay thoroughly.

2. Data acquisition and cleaning

2.1 Data Sources

The data about the venue like venue name, category, location can be retrieved using the Foursquare API. As this study is only for Orlando, Florida. So, the coordinates of the city are retrieved from google Maps API.

2.2 Data Cleaning and Feature Selection

The venue data is retrieved based on the popularity using Foursquare API. So, instead of retrieving the data of all the venues listed in Orlando, only the popular venues are retrieved. This makes the dataset concise and reduced the number of records. There are more than 60 unique categories in the dataset. After analyzing it closely, it has been observed that there are so many categories related to Restaurants like Ramen Restaurant, Tapas Restaurant., etc. All the basically restaurants. So, a new blanket

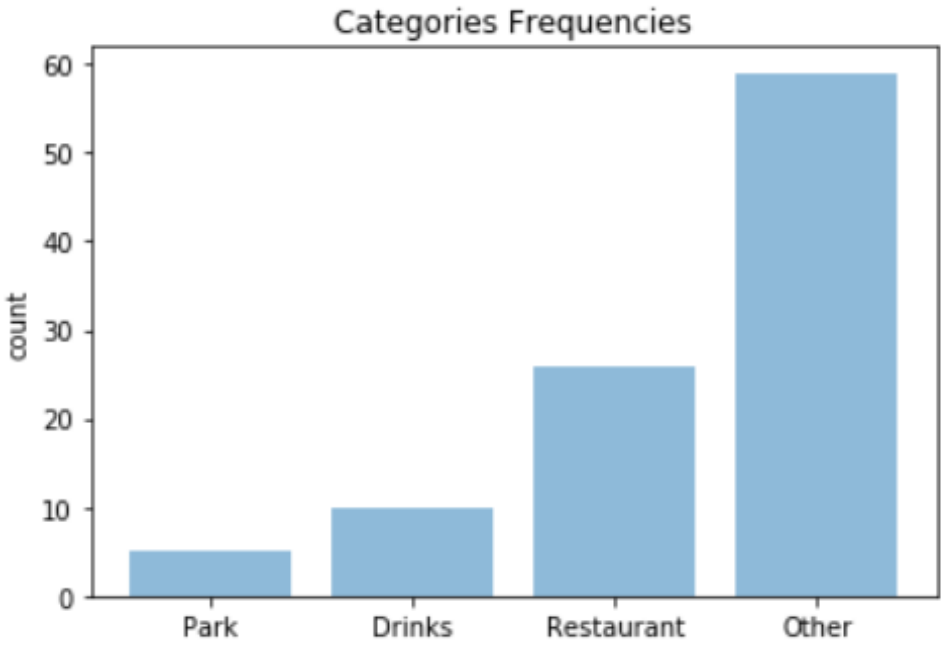
category is created called Restaurants where all the different kind of restaurants are under it. Same goes for the “Drinks”. This helps in focusing on the 3 major categories – Restaurants, Drinks and Parks. Everything else goes in the blanket category “Others”.

There are so many details attached to the venue in the dataset. Only the required fields have been extracted from the complete dataset; which are Venue Name, Categories, Latitude and Longitude.

3. Data Analysis

3.1 Calculation of the frequency

This gives the concise picture of the popular venues in different categories. This will help the expected visitors to get some idea about what to expect.



3.2 Top 5 venues in different categories

Usually tourists are interested in the options available. So, giving them the name of the venues will be a good add-on features in the application.

Parks

Lake Underhill Park
Lake Eola Park
Lake Davis Park
Delaney Park
Orlando Loch Haven Park

Restaurants
Tijuana Flats
Chipotle Mexican Grill
Chick-fil-A
Gringos Locos
Firebirds Wood Fired Grill

Drinks
Lucky Lure
The Hammered Lamb
Southern Nights
Roque Pub
Stonewall

Others
Publix
Wawa
ABC Fine Wine & Spirits
Starbucks
The Country Club Of Orlando

4. Conclusion

This will help visitors to narrow down the area and places based on their interest. Adding this functionality to the application will be useful for the application users. This will help users to make the informed decision. In return application/website will be get good traffic which will help in revenue generation.

5. Future Work

Currently, this is only for one city. This can be done for most popular cities around the world. Monthly email letter based on the predictive analysis will be another add-on functionality for the users of the application.