Analysing restaurants in Brooklyn, New York

Aditya Deepak Nayar

Introduction:

Background

New York in one of the most diverse cities in the world with people from several countries residing there .One of the most visible effects of this can be seen in the variety of eateries and restaurants that are found there with restaurants serving almost all types of cuisines. But all of them are not evenly distributed across the city and usually restaurants serving a particular cuisine might be more popular in one neighbourhood while other cuisines may be popular in some other neighbourhoods. These are affected by a lot of factors and some of which are evident whereas others are not. So a chef or a restaurant chain looking to open a restaurant in a particular neighbourhood will have to consider the popularity of the dishes or cuisine it is planning to serve in that given neighbourhood and it would also be beneficial to know the areas where they are popular.

Problem

Grouping different neighbourhoods in Brooklyn having similar food patterns of eateries so as to make decisions on which neighbourhood to open the restaurant in so as to get maximum customers.

Interest

This problem will be considered by several large fast food chains, burger joints and so on to decide the area in which they should start their branch to get maximum customers. They can use the data to decide on the kind of dishes they should focus more on in different outlets and also the kind of competition they will have in different areas and the offers that should be given and so on. This data can also be used by chefs who are specialised in a particular cuisine to decide the area where they would like to open their restaurant and the taste of people in different neighbourhoods.

Data

Data Sources

The details about the latitude and longitude of the neighbourhoods in New York were available in <https://geo.nyu.edu/catalog/nyu_2451_34572> which I had downloaded from an IBM sever where it was present in json format which was cleaned and stored into a data frame. The details about the restaurants were obtained by using the Foursquare API.

Data Cleaning

Initially the json file containing the latitude and longitude details were downloaded and then they were converted into a data frame with neighbourhood, borough, latitude and longitude as the features. As this data contained details of all boroughs in New York I filtered it so as to contain neighbourhoods in Brooklyn alone.

Then requests were made using the Foursquare API which returned the details of all the venues in a radius of 500 meters around the neighbourhood’s latitude and longitude. The request also returned the category of the given venue. These details were then put into a data frame.

To filter out the other venues that were not restaurants or eateries, by checking the venue category one can find out that all eateries had either a restaurant, joint, bistro or place in their venue type. Shops like ice cream shops and candy shops were not included. Then these new columns having one of the above four words were put into another list and then their values were appended into the data frame through which all the eateries were stored in a single data frame.

Feature Selection

For the analysis all the venues that were eateries were chosen. Along with this the neighbourhood name was also chosen to enable clustering later on was also selected. The neighbourhood coordinates are also stored to map the clusters at a later stage. There were no redundancies in the results of the Foursquare API.