Final Report

.

Introduction

A new restaurant owner would have a lot of decisions to make when opening a new restaurant , each of these decisions is of course of high importance and relevance to the success of this restaurant. Factors like food, price, employees ,etc. A very important factor to consider is the location of the restaurant to show why this is of high importance let us consider the following scenario if a quiet and fancy type of restaurant open in a location that is full of teenagers and kids odds are the kind of customer that this restaurant would target is not going to casually be in the area. On the other hand if that same restaurant would open in let us say in an upper class neighborhood It would be a huge success. This is why locations are vital for restaurant. What I will try to do is to analyze where should a restaurant open in New York city.

Data

We downloaded the data about New York city from this link <https://cocl.us/new_york_dataset> . This data has information about places in New York City . We extracted from it Neighborhood , Borough ,latitude and longitude and read that data into a dataframe of 306 samples and 4 features to further process in python. After that we used the Foursquare API to explore these neighborhood and get in return the recommended places in that particular longitude and latitude. By sending an API request to foursquare and getting in return a json object from which we extracted a venue’s name ,longitude ,latitude, and category (Restaurant ,gym, etc ).We limited the API results to 20 result due to the call limit. For some neighborhoods the API requests returned an empty json object which we disregarded and dropped from our analysis. The final analyzed vector had 103 neighborhoods and 10 Columns.

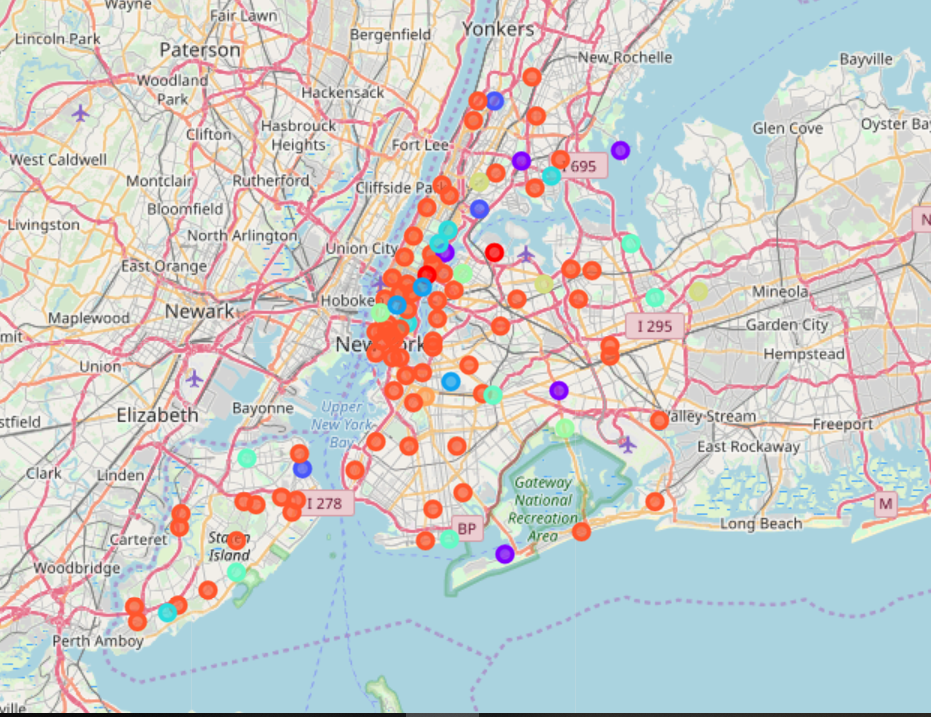
Methodology

Having already saw the dataset in a previous assignment before we knew what we were dealing with.

We explored the json objects returned from the API calls and the json object downloaded from this URL <https://cocl.us/new_york_dataset>. We found for the new York dataset that most properties were in the key features which was basically a list of neighborhoods. The API calls were done using the explore endpoint which returned recommended venues to go it given a certain longitude, latitude, radius , limit and API version. We used a limit of 20 places and a radius of 50 Meters from the exact longitude and latitude. We extracted from the API response the recommended venues which were in the first item of the groups key. The response contained a venue category feature which was of great interest to us to analyze which venues were popular in which neighborhood we created a one hot encoding of this venue category feature. We calculated how many each category is repeated per neighborhood and then we created a new data frame with the top 5 common venue category per neighborhood. From here we used K means clustering with K =10 to partition our neighborhoods into 10 clusters each with a similar taste in venues. Meaning that for example cluster 1 would be interested in gyms and sports center and cluster 2 would be more interested into home appliances stores and so on. We used K means to find the clusters where neighborhood are more interested and keen on restaurants and coffee shops for a restaurant opening there would mean attracting the right kind of people.

Results

We visualized our clustering results as you can see here



What we can see from here is that in most of these locations the red cluster seems to have the upper hand as it is very dominant. Below we defined the prefences of every cluster

|  |  |
| --- | --- |
| Cluster One | Fancy neighborhoods preference for expensive food, and activities like sushi restaurants and yoga studios |
| Cluster Two | Neighborhoods that prefer diners and bars |
| Cluster three | Neighborhoods that prefer discount and department stores |
| Cluster Four | Neighborhoods that prefers gyms and shopping stores |
| Cluster Five | Neighborhoods that prefers pizza places |
| Cluster Six | Neighborhoods with many bus stops and convenience stores |
| Cluster Seven | Neighborhoods that prefer Chinese restaurants |
| Cluster Eight | Neighborhoods that have a lot of parks |
| Cluster Nine | With one output is an outlier and more like belongs to Cluster eight |
| Cluster ten | The dominant cluster that shows neighborhoods with 72 outputs in almost all of them there is at least one restaurant, food court or food truck and it seems like the common trend of the city to have this variability. |

Discussion

So our question which we posed to the very beginning which was were should a restaurant owner open still depends on the genre and type of restaurant. If it is a fancy one I would recommend opening in neighborhoods in cluster 1 if it is a Chinese one it should open in neighborhoods in cluster 7 and so on.

So there is no perfect one answer for all restaurant owners but for every type of restaurant there is an answer for them. If it is general restaurant I would recommend neighborhoods in Cluster 10.

Conclusion

In this assignment We have clustered neighborhoods in New York City according to their most featured venues in order to answer our question of where should a restaurant open in New York City. This assignment can be handy to restaurant owners still researching for the best venue for their restaurant.