***BATTLE OF THE NEIGHBORHOODS PROJECT***

**1. Introduction**

1.1 Background:

Building a good chain of outlets for any business is a necessity in the current world to survive as a brand. Many brands are unable to sustain in the market even after good quality of products and services for just one most important reason, which is, bad placement of the outlet/branch or not expanding to the correct location at the correct time.

Correct time of expansion depends upon the brand, which usually depends on the quality or products and services as good quality of products and services gain them the necessary funding to expand. After that, it's upon the brand to invest their resources on expansion or modification of existing outlets.

Though modifications of current outlet/branch is a good step, but in most of the cases, in contrast to expansion, its effects on the profits is very less.

Correct placement of the outlet/branch in a given neighbourhood is a very important step which must be done with all the necessary background studies done as one wrong placement can result into huge loss, and thus we decided to deal with this particular problem. Our area of concern for this project will be the state of New York.

## 1.2 Problem Statement:

Keeping in mind the problem stated in the background study, and for a sample client in our scenario, i.e., a Pizza Place owner. Thus the problem statement can be stated as:  
"T**o find the best locations in New York State for the expansion of a Pizza Place based in Carnegie Hill, Manhattan, NY**."

1.3 Interest:

Our current client is very much interested in the project as he will get a narrowed down list of all the places where he can possibly expand, based on location.

Any other business who wish to expand their business might also be interested in this project based on the success/satisfaction of our current client.

**2. Data Acquisition and Cleaning**

## 2.1 Data Sources:

The main data source for our project is the neighbourhood JSON data found [here](https://geo.nyu.edu/catalog/nyu_2451_34572). This dataset contains all the neighbourhoods in New York State along with their Latitude and Longitude values.

We also used the [foursquare](https://foursquare.com/) API to retrieve all the nearby venues in the form of JSON data, which consisted of all the venues with their Latitude and Longitude values along with their Venue category.

## 2.2 Data Cleaning:

Data downloaded is in below format:

{'type': 'FeatureCollection',

'totalFeatures': 306,

'features': [{'type': 'Feature',

'id': Place ID,

'geometry': {'type': 'Point',

'coordinates': [Latitude, Longitude]},

'geometry\_name': 'geom',

'properties': {'name': 'Place Name',

'stacked': 1,

'annoline1': 'Place name annotation name 1',

'annoline2': ‘Place name annotation name 2’,

'annoline3': ‘Place name annotation name 3’,

'annoangle': 0.0,

'borough': 'Borough Name',

'bbox': [Top left x,

Top left y,

Bottom right x,

Bottom right y]}},…

…],

'crs': {'type': 'name', 'properties': {'name': 'urn:ogc:def:crs:EPSG::4326'}},

'bbox': [Entire State Top Left x,

Entire State Top Left y,

Entire State Bottom Right x,

Entire State Bottom Right y]

}

From the above data, we need the data under “features” tag only. So we access it by calling file[‘features’], and ignore rest of the data. We can use “totalFeatures” tag as well so as to confirm we have read all the features data.

The Foursquare data obtained is in the following format:

{“response”:

{“groups”:{

“items”: {

“venue”:{

“name”: “Venue Name”,

“location”: {“lat”: “latitude”, “lng”: “Longitude”},

“categories”: {

{“name”: “Category Name”

(other tags)

}

{“name”: “Category Name”,

(other tags)

}…

}

}

}

}…

…

}

}

## 2.3 Feature Selection:

After cleaning the data, we select features by accessing each feature using file[“features”][feature number – 1]. We need not perform any specific operations on the neighbourhood JSON file since it is well made.

Of all the features available to us, we will be needing “coordinates”, “name” and “borough” tags for our use and we may discard all the other tags since they serve no purpose to our project as per the requirements.

From the Foursquare data retrieved, the tags of our concern are as follows:

1. Name

2. Location

3. Categories, which are under the “venue” tag. We may discard all other tags as they serve no purpose to our project.

**3. Exploratory Data Analysis**

3.1 Initial JSON Data Analysis:

Our initial JSON data consisting of the details of all the neighbourhoods in each Borough of the state of New York. We mine that data and obtain the results that the state of New York consists of 5 Boroughs, namely, “Bronx”, “Manhattan”, “Brooklyn”, “Queens” and “Staten Island”. These boroughs consist of a number of neighbourhoods whose numbers can be identified from the following graph:

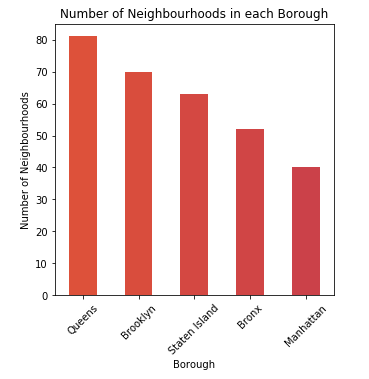


Fig. 1: Number of neighbourhoods in each Borough

From the graph, we can clearly identify that the number of neighbourhoods in each borough are as follows in descending order:

1. Queens: 81
2. Brooklyn: 70
3. Staten Island: 63
4. Bronx: 52
5. Manhattan:40

Below map picturizes all the neighbourhoods in each separate borough:

P.T.O.

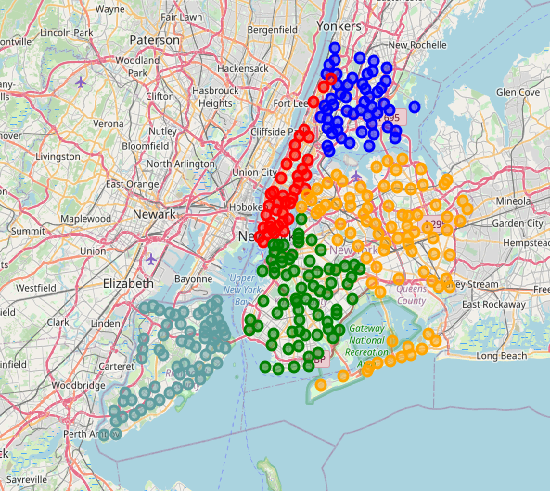


Fig. 2: Map of all Neighbourhoods in each of the 5 Boroughs

## 3.2 Venues in each neighbourhood:

Since we are using the foursquare API to retrieve details of all the venues in 500m range of the neighbourhood, we cover almost all major venues and in the neighbourhood. Since there are a lot of neighbourhoods in the, it is impossible for us to plot them on a single map. So, we classify them on the basis of “Venue Category” and find the top 10 most common venues. Below tables suggest top 10 most common venues in each neighbourhood (we display only 5 of the neighbourhoods in each borough so as to not cover a lot of space):



Table 1: Top 10 most common venues in neighbourhoods of Bronx



Table 2: Top 10 most common venues in neighbourhoods of Manhattan



Table 3: Top 10 most common venues in neighbourhoods of Brooklyn



Table 4: Top 10 most common venues in neighbourhoods of Queens



Table 5: Top 10 most common venues in neighbourhoods of Staten Island

**4. Predictive Procedure**

4.1 Narrowing down based on top common venues in a Neighbourhood:

First we collect all the neighbourhoods where our venue of concern, i.e., “Pizza Place” doesn’t come under top 10 most common place so as to make sure our client doesn’t place their new outlet in a region where their product is already very common. New table of suggested places is as follows (only 5 neighbourhoods are displayed so as to save space):

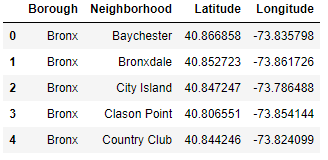


Table 6: Neighbourhoods where Pizza Place is not in top 10 most common venues

Following graph visualizes all the narrowed down areas:

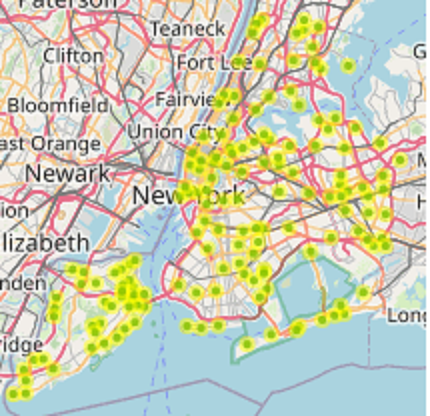


Fig. 3: Map of all suggested Neighbourhoods

In order to find where all the pizza places are present with respect to our identified neighbourhoods, we plot a map with all the Pizza Places, along with all our identified neighbourhoods as given below:

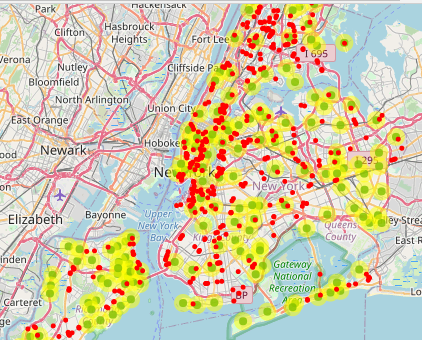


Fig. 4: Map of all suggested Neighbourhoods with all Pizza Places

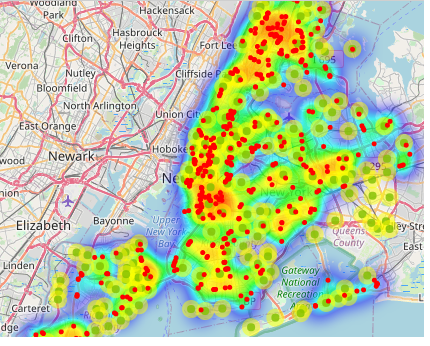


Fig. 4: HeatMap of all suggested Neighbourhoods with all Pizza Places

4.2 Narrowing down based number of competitors in 1Km radius:

Now we will narrow down the list of suggested places further by finding all the competitors of our client within 1Km radius using Foursqaure API again. Based on the number of competitors, we divide the neighbourhoods into 3 categories, namely: “Best”, “Moderate” and “Worst”. This is done based on the number of competitors in 1Km radius.

1. Best: 0 or 1 competitors

2. Moderate: 2-4 competitors

3. Worst: 5 or more competitors

Following table suggests all the neighbourhoods categorized into the stated categories:

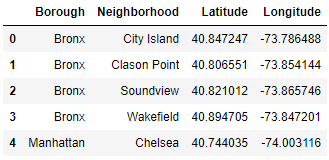


Table 7: Best suggested neighbourhoods

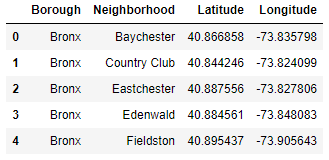


Table 8: Moderate suggested neighbourhoods

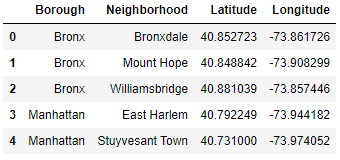


Table 9: Worst suggested neighbourhoods

Then we plot all the regions on the map as given below:

P.T.O.

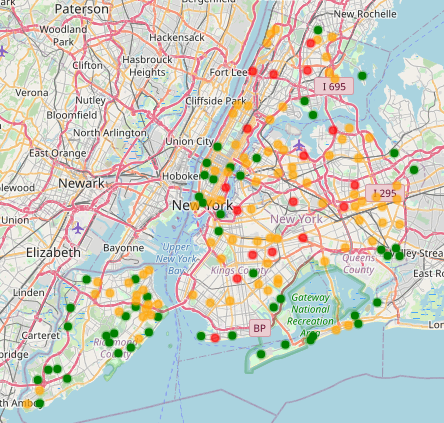


Fig. 5: Best (in green), Moderate (in orange) and Worst (in red) suggested neighbourhoods

4.3 Narrowing down based on clusters of Best Suggested Neighbourhoods:

Now we use DBSCAN clustering algorithm to cluster the Best Suggested Neighbourhoods with at least 3 neighbourhoods per cluster with epsilon value of 0.005. Upon using the algorithm, we find the centres and radii of each cluster and plot them on the map as given below:

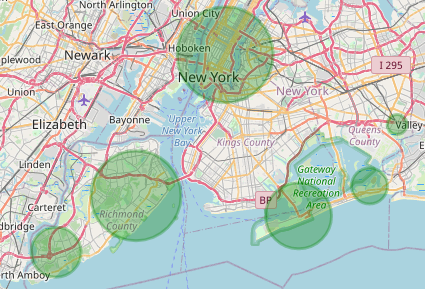


Fig. 6: Clusters of best suggested neighbourhoods

We also plot all the neighbourhoods with qualify as “Best” neighbourhoods for our client to open a new branch/outlet as given below:

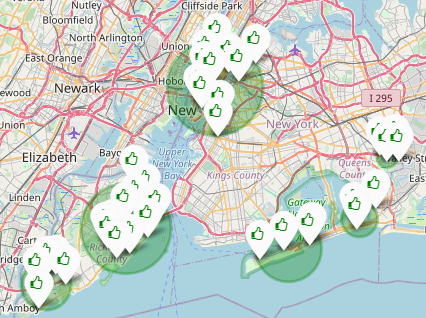


Fig. 7: Clusters of best suggested neighbourhoods along with their locations

The biggest cluster among all these clusters suggests that there are a large number of neighbourhoods in close proximity where there are very few “Pizza Places” compared to other neighbourhoods and thus will be best suitable for our client to open a new branch/outlet, thus increasing their chances of success by a great factor.

Below table includes a list of all the neighbourhoods which come under the biggest cluster of “Best” suggested places:

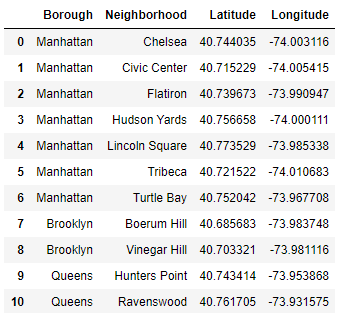


Table 10: Neighbourhoods which come under the biggest cluster of “Best” suggested Neighbourhoods

**5. Conclusions**

From all the above data processing and analysis, we can come to a final conclusion that following neighbourhoods are the best for our client to carry out further research on the likings and average money spent on eating out, by the residents, to further narrow down their new outlet location:

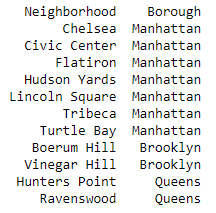


Table 11: Best suggested Neighbourhoods

Client may also want to look into following neighbourhoods, although they do not make a huge cluster and thus may not be as beneficial for our client to invest their resources on:



Table 12: Other suggested Neighbourhoods

**6. Future Directions:**

Based on our current research and methodology, we were only able to suggest our client a possible neighbourhood based on location. Later on for further improvements, we may also want to look into the average salaries and average spending of people living in our target neighbourhoods so as to better analyse and further narrow down the list of suggested neighbourhoods. We may also consider the general likes and dislikes of people as they also play an important factor in deciding whether any business will be successful in the region or not.