

Capstone Project - The Battle of the Neighbourhoods

Applied Data Science Capstone by IBM/Coursera

Background and History

New York City is the largest city in the United States of America and the city's demographics show us that it has a wide and very diverse range of ethnic groups in its metropolis. It is home to approximately 8.3 million people in a recent count (year 2019) which accounts for over 42.67% of New York's total population, which approximates to 19.45 million in total. Over the past years, the city's population growth rate has been growing faster than the New York region itself. Part of the reason for this has been because New York has had a long history of immigration and continues to be by far the leading metropolitan gateway for legal immigrants admitted into the United States of America.

With this major point of entry for immigrants, a lot of diverse immigrant neighbourhoods were formed and with this formation it brought new different cultures to New York. Many new languages were introduced to the shores of New York and some diverse languages were eventually spoken more frequently than the most common spoken language in the world, English. The English language is the least spoken language in neighbourhoods such as Corona, Flushing and Sunset Park.

Various unique languages were not the only customs introduced to New York. A variety of world cuisines were introduced as well. Diverse cultures brought diverse food items that many populations, not just in the United States of America but in the rest of the world, cook and enjoy with each other. There are many restaurants in New York that offer delicious food and they range from Chinese, Portuguese, Indian, etc.

Introduction: Business Problem

New York is filled with many large and very diverse ethnic groups like Portuguese, Mexican, Vietnamese, etc. These ethnic groups offer their very own unique cuisines for locals and foreigners to enjoy. The purpose of this project is to find the best restaurants that offer these cuisines. We shall narrow our search to find the best Thai cuisine in New York.

We will search for the best Thai restaurants in all the major parts of New York. This report will be targeted to stakeholders, who are predominantly tourists, who are interested in finding highly rated Thai restaurants in New York. Other stakeholders can be entrepreneurs who want to open a Thai restaurant in areas that lack such cuisines.

We will use our data science tools and skills to produce and visualize the boroughs and neighbourhoods that have the highest rated Thai restaurants in New York. From this analysis, the stakeholders will be able to make a confident decision on which restaurant to visit.

Data Acquisition and Cleaning

Data Acquisition

Factors that will influence our decision will be based on the definition of the business problem. The factors are as follows:

- The number of existing restaurants in a neighbourhood
- The number of and distance to Thai restaurants in a neighbourhood from the city centre.

Regularly spaced grid of locations will be used, centred around city centre, to define our neighbourhoods.

Following data sources will be needed to extract/generate the required information:

- New York data that consists of lists of boroughs and neighbourhoods with their respective details.
 - **Data Source:** https://cocl.us/new_york_dataset
 - **Data Description:** This dataset consists of details such as the boroughs, neighbourhoods and their respective latitude and longitude coordinates. This data will then be used to explore the various neighbourhoods of New York.
- The number of Thai restaurants and their venue details in each neighbourhood in New York will be obtained.
 - **Data Source:** [Foursquare API](#)
 - **Data Description:** With the use of the API, a list of all the restaurant venues in each neighbourhood will be presented. Although all the venues will be displayed, we shall filter the venues that only have Thai restaurants. Venue details such as ID, Name, Likes, Rating and Tips will also be collected as well.
- GeoSpace (Geospatial) data
 - **Data Source:** <https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmi-j8zm>
 - **Data Description:** By using this data, we will obtain the New York Borough boundaries and be able to visualize a choropleth map of New York. This dataset will be download into a text file which will be called "NY_Borough_Boundaries.txt"

Data Cleaning

The data preparation for each of the three data sources of data will be done separately.

We will begin by extracting the New York data and from there, the boroughs with the associated neighbourhoods will be collected. The coordinates of each neighbourhood will also be displayed (see **Fig 2.1**).

	Borough	Neighbourhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Fig 2.1 New York borough and neighbourhood data

Next, we will use the FourSquare API to obtain the venues of each Thai restaurant in New York (see **Fig 2.2**). Venue details such as ID, Name, Likes, Rating and Tips will also be collected as well (see **Fig 2.3**).

	Borough	Neighbourhood	ID	Name
0	Bronx	Riverdale	4b94a4a2f964a520787f34e3	Nam Thai
1	Bronx	Kingsbridge	4b94a4a2f964a520787f34e3	Nam Thai
2	Manhattan	Marble Hill	4a739e29f964a520f5dc1fe3	Siam Square
3	Bronx	Country Club	4b4c90d3f964a52046b626e3	Honey's Thai Pavilion
4	Bronx	Parkchester	5585c948498e0bd8a8a706f1	Thai Cuisine

Fig 2.2 Thai restaurant venues

	Borough	Neighbourhood	ID	Name	Likes	Rating	Tips
0	Bronx	Riverdale	4b94a4a2f964a520787f34e3	Nam Thai	3	7.2	5
1	Bronx	Kingsbridge	4b94a4a2f964a520787f34e3	Nam Thai	3	7.2	5
2	Manhattan	Marble Hill	4a739e29f964a520f5dc1fe3	Siam Square	15	7.0	13
3	Bronx	Country Club	4b4c90d3f964a52046b626e3	Honey's Thai Pavilion	19	7.2	28
4	Bronx	Parkchester	5585c948498e0bd8a8a706f1	Thai Cuisine	17	7.5	1

Fig 2.3 Thai restaurant venue details

Finally the geospatial data that will be obtained online and saved as a text file will be used to draw up the boundaries of each borough which will later be used to find the average ratings of Thai restaurants for each borough. Those results will be displayed in the Results and Discussion section.

Methodology

Overview

1. Find the New York data from this link: https://cocl.us/new_york_dataset
2. Find all the venues for every neighbourhood in New York using the FourSquare API
3. Filter out all the Thai Restaurant venues
4. Find all the venue details such as the Likes, Ratings and Tips for all the Thai Restaurants using the FourSquare API
5. The data will be sorted in a way that will keep the Ratings as the constraint
6. The ranking of the neighbourhoods and restaurants will be visualized using Python's Folium Library

Exploratory Data Analysis

A statistical summary will be conducted at this stage of the analysis.

The total count of Thai restaurants equates to 136 restaurants throughout New York (see **Fig 2.4**).

```
# Displaying the shape of the above dataset  
Thai_Cuisine_df.shape  
(136, 4)
```

Fig 2.4 Total number of Thai Restaurants

Queens is the borough that contains the most Thai restaurants compared to the other boroughs with 67 restaurants in total, while Staten Island has the least (see **Fig 2.5**).

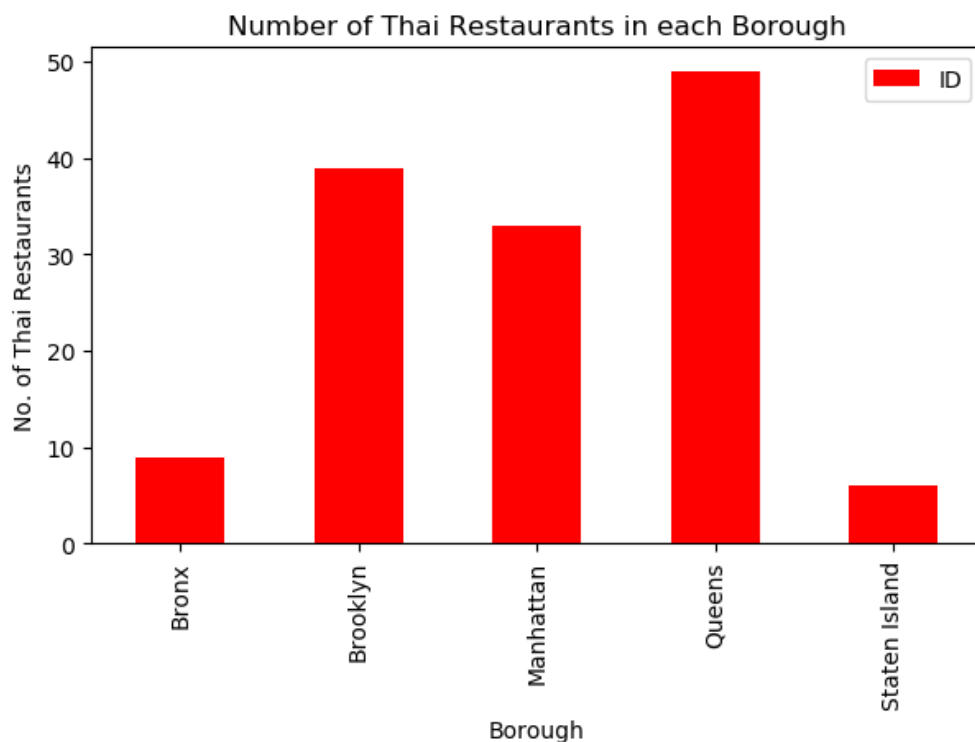


Fig 2.5 Number of Thai Restaurants in each Borough

Elmhurst is the neighbourhood that contains the most Thai restaurants compared to the other neighbourhoods in New York with 9 restaurants in total (see **Fig 2.6**).

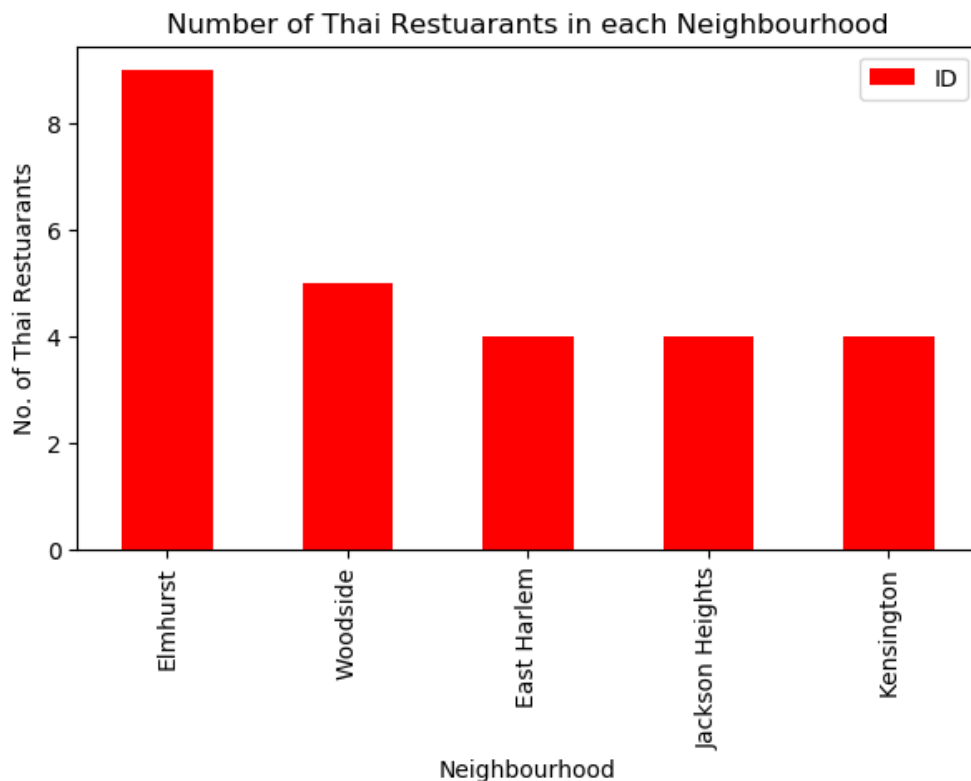


Fig 2.6 Number of Thai Restaurants in each Neighbourhood

Next, we will look at the analysis pertaining to the details of each Thai restaurant. Details such as the restaurant with the most likes, the restaurant with the highest rating and the restaurant with the most tips.

The restaurant with the most likes is Uncle Boons with 1015 likes in total. The restaurant is situated in a neighbourhood called Little Italy in Manhattan (see **Fig 2.7**).

```

Borough      Manhattan
Neighbourhood Little Italy
ID           516863b1011ca6684e37a88f
Name         Uncle Boons
Likes        1015
Rating       9.1
Tips         328
Name: 62, dtype: object

```

Fig 2.7 Thai Restaurant with the most Likes – Uncle Boons

The restaurant with the highest rating of 9.2 is Wayla and the restaurant is situated in a neighbourhood called Chinatown in Manhattan (see **Fig 2.8**).

```
Borough           Manhattan
Neighbourhood      Chinatown
ID                 5cc4e9d0c876c8002c3010cb
Name               Wayla
Likes              126
Rating             9.2
Tips               26
Name: 46, dtype: object
```

Fig 2.8 Thai Restaurant with the highest Rating – Wayla

The restaurant with the most tips is SriPraPhai with a tip of 357 tips in total. The restaurant is situated in a neighbourhood called Woodside in Queens (see **Fig 2.9**).

```
Borough           Queens
Neighbourhood      Woodside
ID                 3fd66200f964a5204ef11ee3
Name               SriPraPhai
Likes              766
Rating             8.9
Tips               357
Name: 71, dtype: object
```

Fig 2.9 Thai Restaurant with the most Tips – SriPraPhai

Next, we will further our analysis by calculating and extracting the average ratings of Thai restaurants for each borough and the highly rated neighbourhoods only.

Manhattan has most of the highly rated Thai restaurants than the other boroughs with an average rating of 8.44 followed by Brooklyn with an average rating of 8.07 (see **Fig 2.10** and **Fig 2.11**)

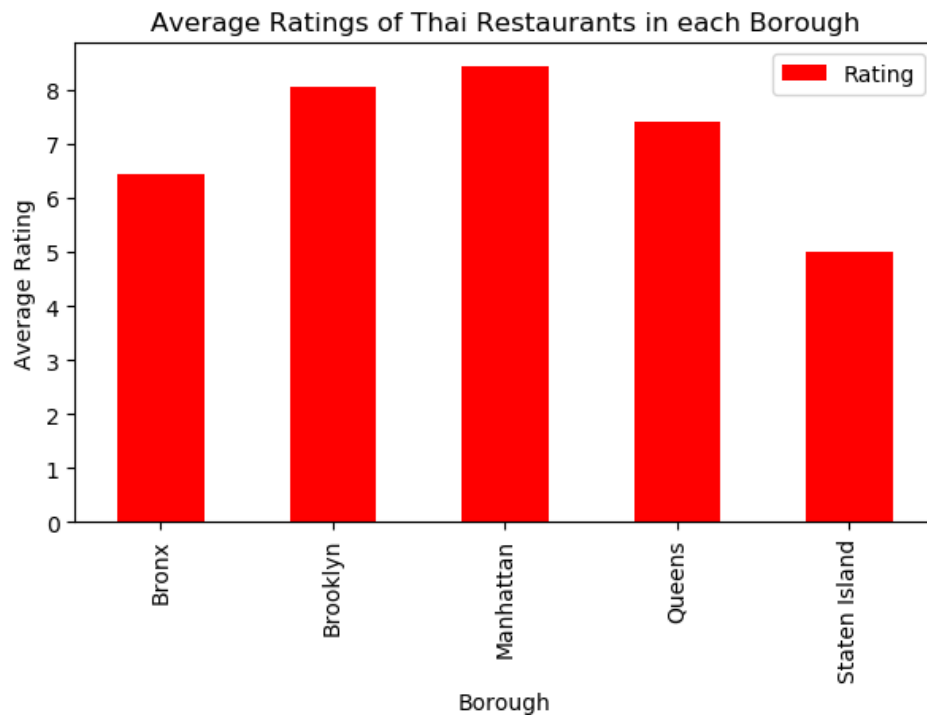


Fig 2.10 Bar Plot of Average Ratings of Thai Restaurants in each Borough

	Borough	Average Rating
2	Manhattan	8.442424
1	Brooklyn	8.066667
3	Queens	7.410204
0	Bronx	6.444444
4	Staten Island	5.000000

Fig 2.11 Average Ratings of Thai Restaurants in each Borough

Chinatown has most of the highly rated Thai Restaurants than the other neighbourhoods in New York with an average rating of 9.20 followed by Noho, Greenwich Village and Clinton which all have the same average rating of 9.10 (see **Fig 2.12** and **Fig 2.13**)

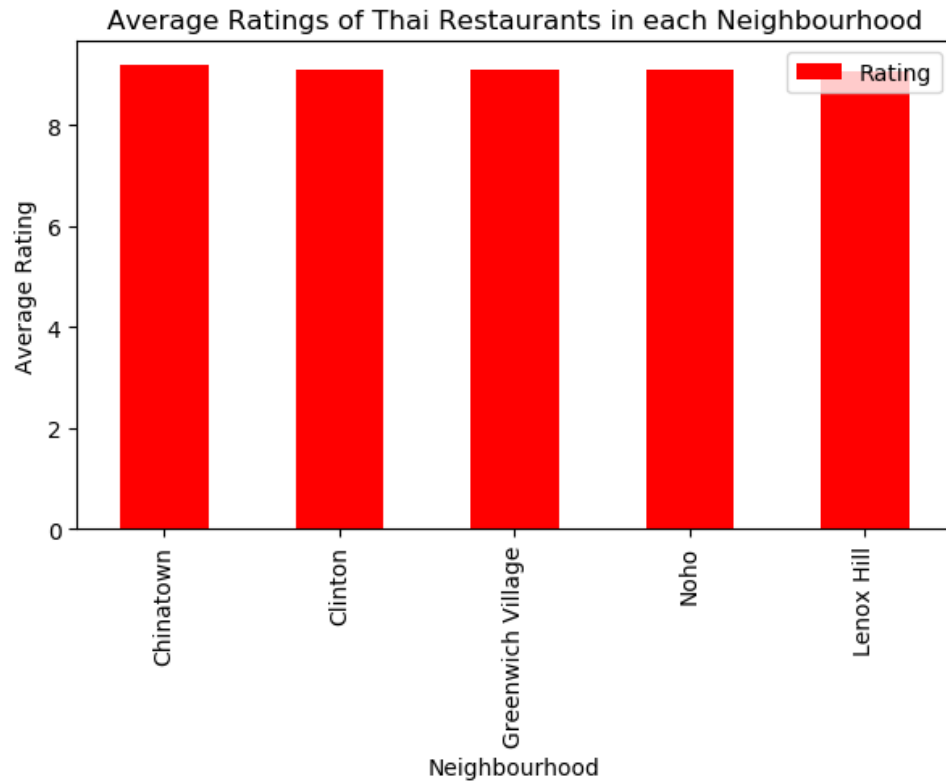


Fig 2.12 Bar Plot of Average Ratings of Thai Restaurants in each Neighbourhood

	Neighbourhood	Average Rating
10	Chinatown	9.20
51	Noho	9.10
34	Greenwich Village	9.10
11	Clinton	9.10
44	Lenox Hill	9.05

Fig 2.13 Average Ratings of Thai Restaurants in each Neighbourhood

Problem Questions

1. Which is/are the best venues for Thai Cuisine in New York?
2. Which areas have Thai Restaurant Market potential?
3. Which is the best area to stay if a tourist prefers Thai Cuisine?
4. Indicate the areas that have badly rated or no Thai Restaurants.
5. Which areas have no Thai Restaurant Market potential?

Answers to Problem Questions:

1. The neighbourhoods with some of the best Thai restaurants with ratings of 9.0 and over are:
 - Boerum Hill
 - Chinatown
 - Clinton
 - Downtown
 - Greenwich Village
 - Lenox Hill
 - Noho
 - Upper East Side
2. Manhattan and Brooklyn have the highest potential for Thai Restaurant market compared to all the other boroughs. While Greenwich Village has the highest potential for Thai Restaurant market compared to all the other neighbourhoods.
3. Manhattan and Brooklyn are the best boroughs to live in for a tourist if they love to eat out in Thai Restaurants.
4. The neighbourhoods with some of the best Thai restaurants with ratings of 9.0 and over are:
 - Rockaway Beach
 - Utopia
 - Tompkinsvilles
 - Kew Gardens Hills
 - St. George
5. Bronx and Staten Island have the least potential for Thai Restaurant market compared to all the other boroughs. While Rockaway Beach, Utopia and St. George all have the least potential for Thai Restaurant market, just to name a few.

Results and Discussion

Now that the data had been acquired and cleaned, exploratory data analysis was performed and some useful insight on Thai restaurants around New York was obtained. We will now express these insights onto three different maps to visualize our findings.

The map below shows us the neighbourhoods across New York that have an average rating of 9.0 and above for Thai restaurants (see **Fig 2.14**). The map can be viewed in this ([link](#)). The neighbourhoods with their ratings and boroughs are as follows:

- Chinatown, Manhattan (9.20)
- Noho, Manhattan (9.10)
- Greenwich Village, Manhattan (9.10)
- Clinton, Manhattan (9.10)
- Lenox Hill, Manhattan (9.05)
- Boerum Hill, Brooklyn (9.00)
- Downtown, Brooklyn (9.00)
- Upper East Side (9.00)

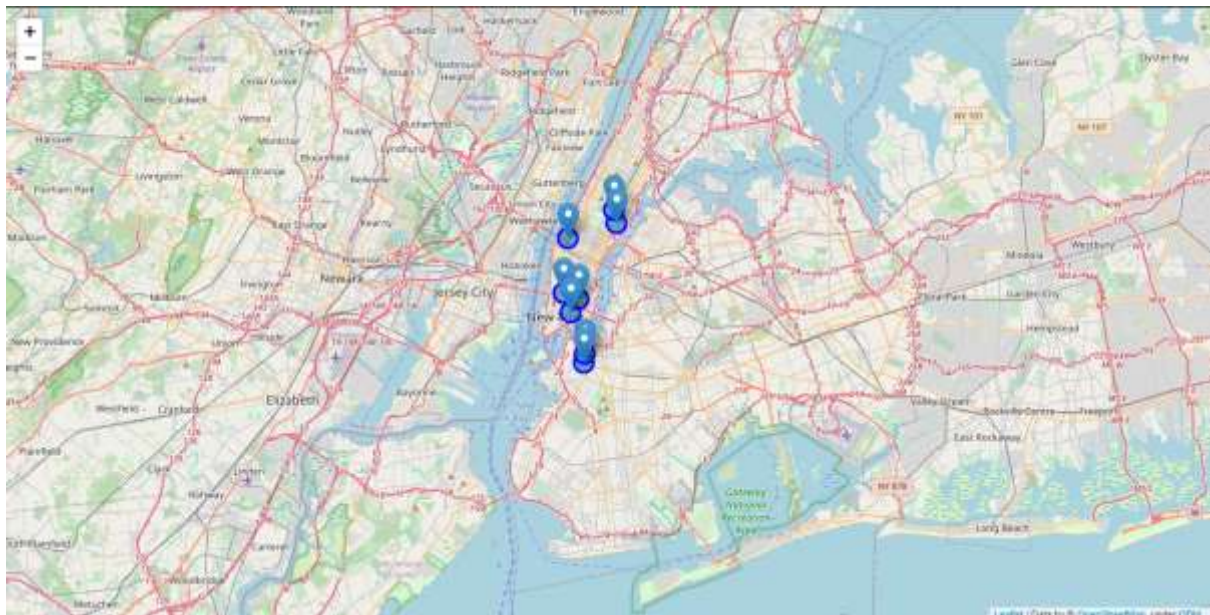


Fig 2.14 Neighbourhoods with a Rating higher than 9.0

The map below shows us the neighbourhoods across New York that have an average rating of 1.0 and below for Thai restaurants (see **Fig 2.15**). The map can be viewed in this ([link](#)). The rating of 0.0 might also indicate that there are no Thai restaurants in that neighbourhood. The neighbourhoods with their ratings and boroughs are as follows:

- Rockaway Beach, Queens (0.0)
- Utopia, Queens (0.0)
- Tompkinsville, Staten Island (0.0)
- Kew Gardens Hills (0.0)
- St. George, Staten Island (0.0)

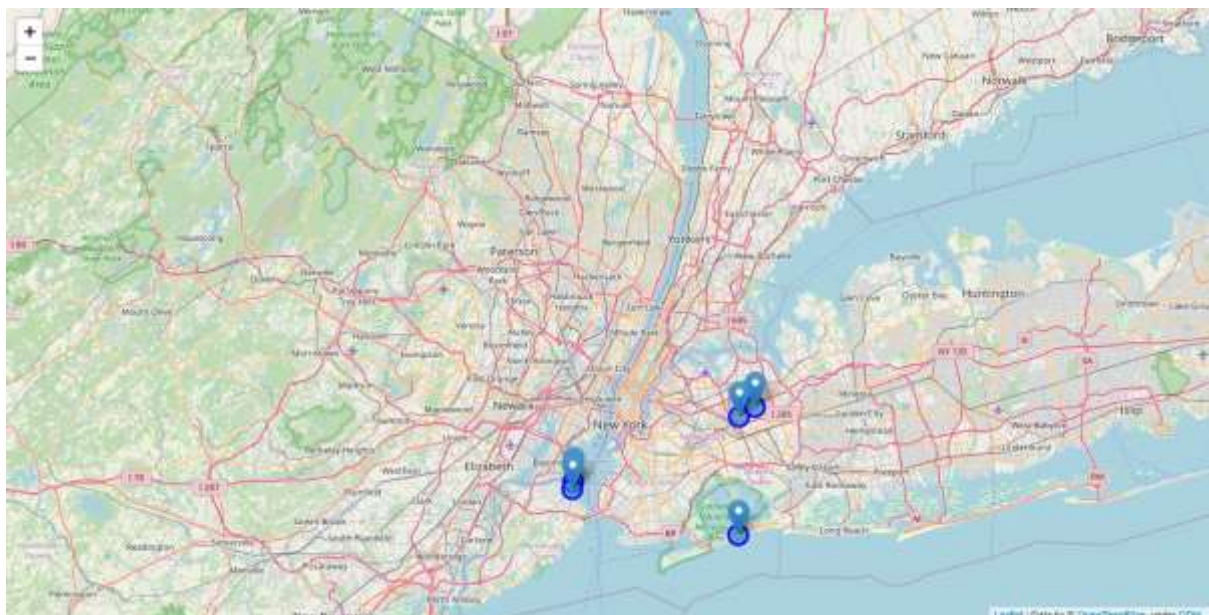


Fig 2.15 Neighbourhoods with a Rating lower than 1.0

The map below shows us the average ratings of Thai restaurants of each borough across New York (see **Fig 2.16**). The map can be viewed in this ([link](#)). The range of the average ratings of each borough are as follows:

- Manhattan (Range of average rating: 7.9 – 8.4)
- Brooklyn (Range of average rating: 7.9 – 8.4)
- Queens (Range of average rating: 7.3 – 7.9)
- Bronx (Range of average rating: 6.1 – 6.7)
- Staten Island (Range of average rating: 5.0 – 5.6)

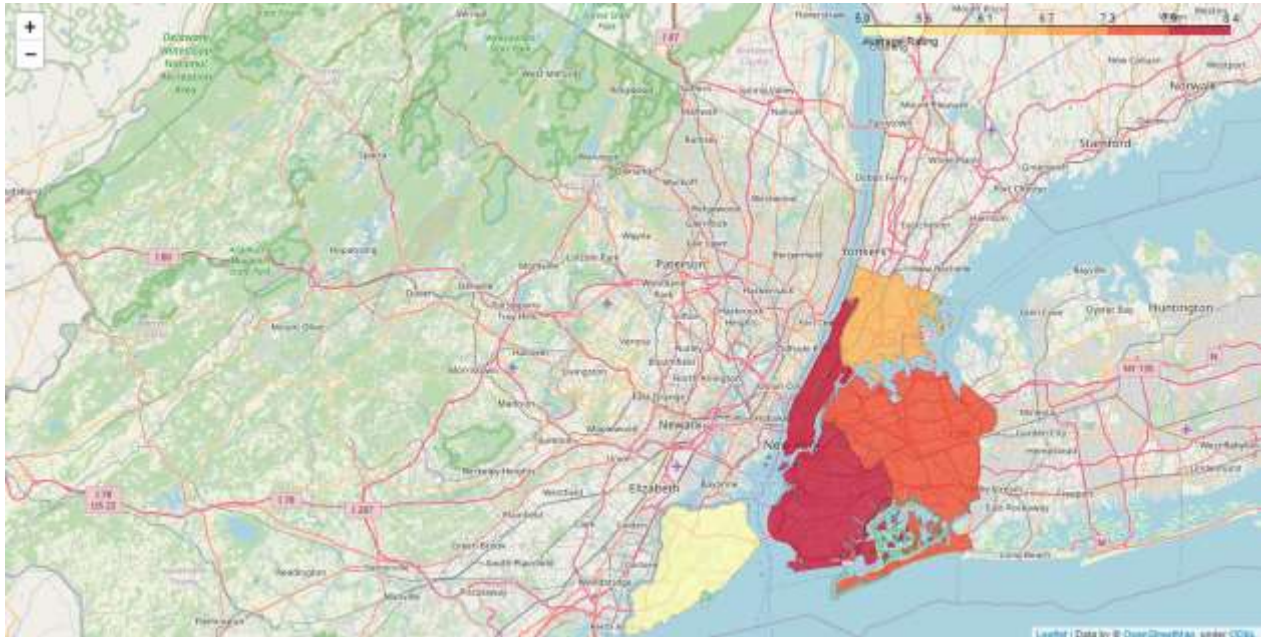


Fig 2.16 Boroughs with Different Average Ratings

The aim of this project is to help stakeholders find areas with the best Thai restaurants for dining purposes. For example, stakeholders such as tourists will be interested in the results of this project. The results indicate that Manhattan and Brooklyn are a hotspot for highly rated Thai restaurants and boroughs such as Staten Island and Bronx should be avoided as they have lowly rated Thai restaurants. Entrepreneurs who are interested in opening their own Thai restaurants will be looking into areas such as Staten Island who have low ratings as the barrier of entry into the Thai restaurant market will be low, therefore making it a great opportunity for business.

Conclusion

This project helps stakeholders get a better understanding of the restaurant market, especially restaurants that offer Thai cuisines in New York. Not only does it give them a better understanding of the market but also the demographics of all the Thai restaurants across New York. This knowledge will help them find the best Thai restaurant where they will be able to dine and enjoy the best Thai food New York has to offer. The results of this project will also aid stakeholders such as entrepreneurs who are interested in starting their own chain of Thai restaurants in New York.