

# Battle of the Neighborhood

## Introduction

Co-living has been the new buzz word in recent years. Traditionally, hunting down an apartment requires a real estate agent and then furnishing the place by renters themselves. Co-living companies are now offering a streamlined alternative in the form of move-in ready rooms in shared apartments with basic services such as utilities and housekeeping covered. Some companies focus on communal life while others emphasize getting out into the neighborhood. With co-living businesses rising – in particular, the city of New York – it is of paramount important to understand the area and the needs of their audience regarding a neighborhood.

## Business Problem

New York is the world's financial hub, home to more than 8.4 million people with a long history of international immigration. Yet, the bustling city is notorious for its sky-high rents. According to "Mapping the World's Prices" report by Deutsche Bank published in 2019, people spent a monthly average of \$2,909 for a two-bedroom apartment, ranked third after Hong Kong and San Francisco.

Co-living rises as an answer to sustainable living, counter-reacting high rents and offering more flexibility as well as affordability to premium living. With many new co-living businesses muscling into the city, it is essential to understand the neighborhood and its audience to create a successful business.

The objective of this capstone project is to analyze and select the best locations in the city of New York as a market research for companies intend to open multiple co-living spaces around the city to better target its audience. It aims to answer the question: If a co-living start-up is looking for the best locations for its premises, where would you recommend so that it can leverage its competitive advantages and tailor its services for its target audience?

## Data

The analysis will cover two main factors from choosing a residential area: safety and culture in the neighborhood. To come up with the model, the following data would be necessary:

1. New York City Neighborhoods The data set contains the geographical information of New York City.
2. New York Neighbourhood GeoJSON
3. New York City Crime Data The data set contains the crime data of New York City over the past decades. Only data collected for 2019 will be used in the analysis.
4. Food Venue data of New York City from Foursquare API As food is an important part of a culture, it helps understand the diversity and the vibes of a ghetto

The data can be downloaded from the above links while venue data can be parsed through posting a GET request to foursquare API. The data is formatted using one hot encoding with the categories of each venue.

## Methodology

### Crime Data

#### Data Wrangling

The data was in .csv format, which can be easily loaded to a pandas data frame. With a preview of the table, irrelevant data columns can be disposed using methods from the pandas library. Only data involving incident category, geographical information, date & time remained for further analysis.

#### Data Analysis

The data were grouped by categories and by neighbourhoods to find out the frequency of occurrence of crime in each category as well as each neighbourhood.

#### Visualization

The data is then put on a map using GPS coordinates using GeoPandas and folium libraries. The pandas data frame was converted to a geo data frame with Shapely geometry object formed. The crime reports was then binned on the NYC neighbourhoods GeoJSON using GeoPandas' sjoin function. The results were then grouped by neighbourhoods and visualised using a choropleth map generated from folium packages.

## Food Venues Data

### Data Wrangling

Neighbourhood data was first transformed into a pandas data frame from NYC neighbourhood data set for later use. Its latitude and longitude will be used to search for venues within the 500m radius.

In order to get the food venues in NYC, it was necessary to fetch Foursquare venue category hierarchy. A function was created to return the IDs, names of food and its sub-categories. Another function was created to loop through neighbourhoods to post the GET request and unfold the list of relevant information for each nearby venue. To counter any redundant requests to the Foursquare API, pickle library was used to serialise the information retrieved.

### Data Analysis

The categorical data retrieved was then turned to numerical data using One-Hot encoding. As we would like to understand the neighbourhood, it is best to cluster them using k-means clustering. The optimal number of clusters was determined from the Silhouette Method – measuring how similar a point is to its own cluster compared to other clusters. Then KMeans.fit method was applied to analyse the venue data.

### Visualisation

The results were visualised using folium library with coloured markers indicating different clustered neighbourhood.

## Results

### Crime Data

The 10 most common type of offense in NYC and its frequency:

|    | Offense                        | Count |
|----|--------------------------------|-------|
| 51 | PETIT LARCENY                  | 91648 |
| 25 | HARRASSMENT 2                  | 74322 |
| 5  | ASSAULT 3 & RELATED OFFENSES   | 55213 |
| 9  | CRIMINAL MISCHIEF & RELATED OF | 48637 |
| 23 | GRAND LARCENY                  | 42713 |
| 17 | FELONY ASSAULT                 | 21534 |
| 39 | OFF. AGNST PUB ORD SENSBLTY &  | 19618 |
| 35 | MISCELLANEOUS PENAL LAW        | 14719 |
| 11 | DANGEROUS DRUGS                | 14130 |
| 56 | ROBBERY                        | 13983 |

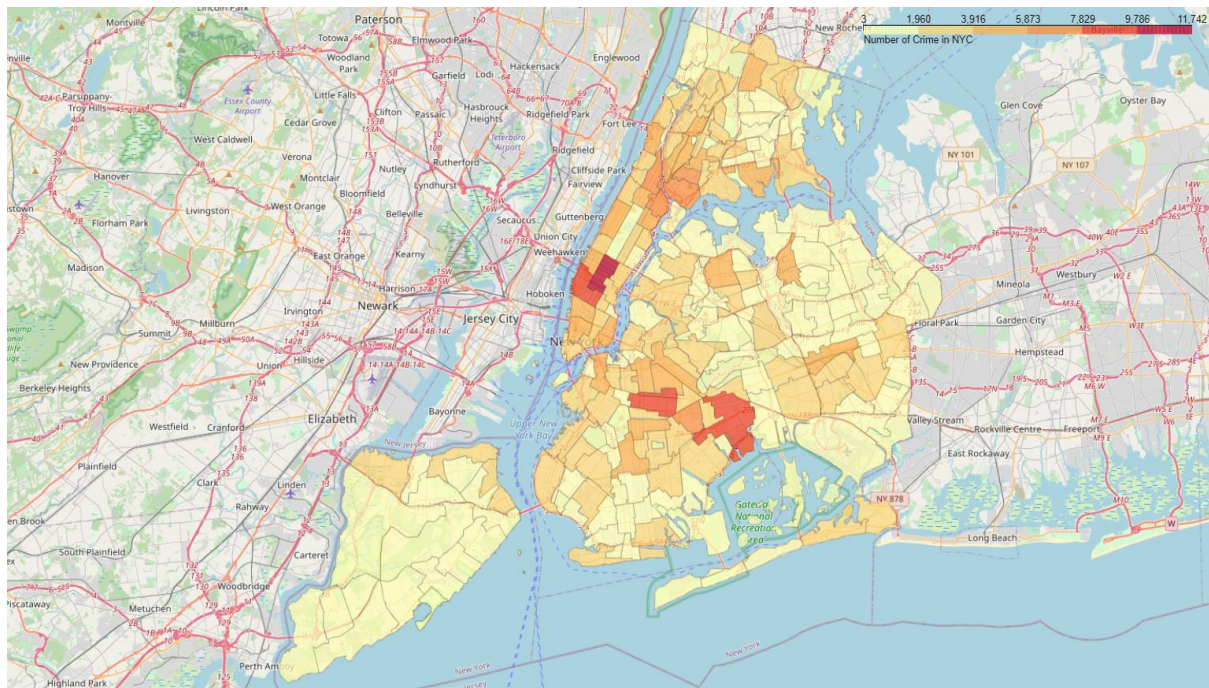
The 5 most common neighborhoods for crimes in NYC:

|     | Neighborhood                    | Count |
|-----|---------------------------------|-------|
| 194 | park-cemetery-etc-Staten Island | 3     |
| 0   | Airport                         | 22    |
| 163 | Stuyvesant Town-Cooper Village  | 181   |
| 3   | Arden Heights                   | 338   |
| 145 | Rossville-Woodrow               | 344   |

The 5 least common neighborhoods for crimes in NYC:

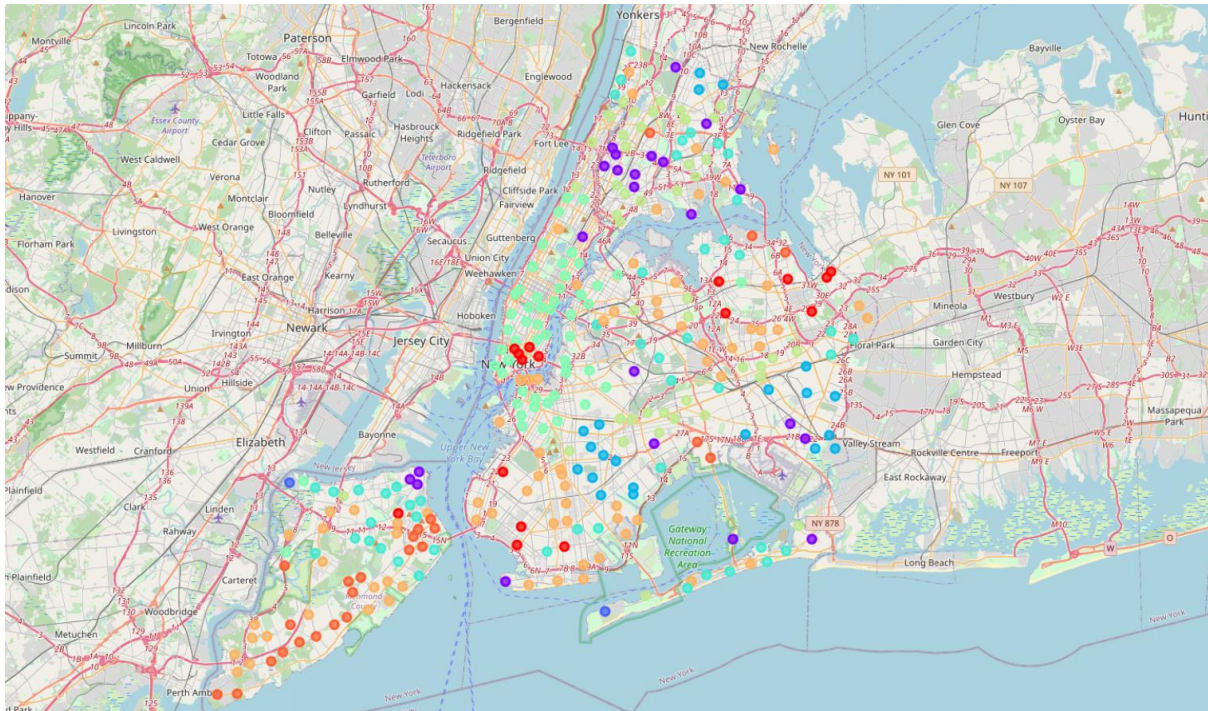
|   | Neighborhood                               | Count |
|---|--|-------|
| 0 | Airport                                    | 22    |
| 1 | Allerton-Pelham Gardens                    | 1163  |
| 2 | Annadale-Huguenot-Prince's Bay-Eltingville | 576   |
| 3 | Arden Heights                              | 338   |
| 4 | Astoria                                    | 3509  |

Visualizing crime frequency in neighborhoods:



## Food Venue Data

Visualizing k-means clustering results:



Cluster 0:

|                    |   |
|--------------------|---|
| Chinese Restaurant | 6 |
| Bakery             | 5 |
| Korean Restaurant  | 1 |
| Bubble Tea Shop    | 1 |
| Coffee Shop        | 1 |
| Mexican Restaurant | 1 |
| Sushi Restaurant   | 1 |

Name: 1st Most Common Venue, dtype: int64

---

|                      |   |
|----------------------|---|
| Bakery               | 4 |
| Coffee Shop          | 3 |
| Chinese Restaurant   | 2 |
| Italian Restaurant   | 2 |
| Donut Shop           | 1 |
| Deli / Bodega        | 1 |
| Cantonese Restaurant | 1 |
| Pizza Place          | 1 |
| American Restaurant  | 1 |

Name: 2nd Most Common Venue, dtype: int64

---

|               |   |
|---------------|---|
| Queens        | 6 |
| Manhattan     | 5 |
| Brooklyn      | 4 |
| Staten Island | 1 |

Cluster 1:

```

Deli / Bodega      23
Name: 1st Most Common Venue, dtype: int64
-----
Fast Food Restaurant    6
Donut Shop              4
Pizza Place             3
Italian Restaurant      3
Chinese Restaurant      2
Food                   2
Sandwich Place          1
Caribbean Restaurant    1
American Restaurant     1
Name: 2nd Most Common Venue, dtype: int64
-----
Bronx                  12
Queens                 5
Staten Island         3
Brooklyn              2
Manhattan             1
Name: Borough, dtype: int64
-----

```

Cluster 2:

```

Food Truck          1
Pizza Place         1
Name: 1st Most Common Venue, dtype: int64
-----
Italian Restaurant   1
American Restaurant   1
Name: 2nd Most Common Venue, dtype: int64
-----
Staten Island       1
Queens              1
Name: Borough, dtype: int64
-----

```

Cluster 3:

|                      |    |
|----------------------|----|
| Caribbean Restaurant | 15 |
| Deli / Bodega        | 5  |
| Fried Chicken Joint  | 1  |

Name: 1st Most Common Venue, dtype: int64

---

|                      |   |
|----------------------|---|
| Deli / Bodega        | 9 |
| Caribbean Restaurant | 4 |
| Chinese Restaurant   | 3 |
| Fried Chicken Joint  | 2 |
| Pizza Place          | 1 |
| Food                 | 1 |
| Fast Food Restaurant | 1 |

Name: 2nd Most Common Venue, dtype: int64

---

|          |    |
|----------|----|
| Brooklyn | 10 |
| Queens   | 8  |
| Bronx    | 3  |

Name: Borough, dtype: int64

---

Cluster 4:

|                    |    |
|--------------------|----|
| Deli / Bodega      | 34 |
| Pizza Place        | 11 |
| Donut Shop         | 1  |
| Indian Restaurant  | 1  |
| Chinese Restaurant | 1  |

Name: 1st Most Common Venue, dtype: int64

---

|                    |    |
|--------------------|----|
| Pizza Place        | 21 |
| Deli / Bodega      | 13 |
| Chinese Restaurant | 5  |
| Sandwich Place     | 2  |
| Food Truck         | 2  |
| Italian Restaurant | 2  |
| Diner              | 1  |
| Greek Restaurant   | 1  |
| Dessert Shop       | 1  |

Name: 2nd Most Common Venue, dtype: int64

---

|               |    |
|---------------|----|
| Queens        | 19 |
| Staten Island | 16 |
| Bronx         | 9  |
| Brooklyn      | 4  |

Name: Borough, dtype: int64

---

Cluster 5:

|                    |    |
|--------------------|----|
| Coffee Shop        | 47 |
| Korean Restaurant  | 2  |
| Bagel Shop         | 1  |
| Pizza Place        | 1  |
| Bakery             | 1  |
| Italian Restaurant | 1  |

Name: 1st Most Common Venue, dtype: int64

---

|                                 |   |
|---------------------------------|---|
| Bagel Shop                      | 8 |
| Deli / Bodega                   | 7 |
| Bakery                          | 7 |
| Pizza Place                     | 7 |
| Coffee Shop                     | 5 |
| Fast Food Restaurant            | 3 |
| Café                            | 3 |
| Mexican Restaurant              | 2 |
| Sandwich Place                  | 2 |
| Italian Restaurant              | 2 |
| Caribbean Restaurant            | 1 |
| Salad Place                     | 1 |
| Southern / Soul Food Restaurant | 1 |
| Bar                             | 1 |
| Burger Joint                    | 1 |
| Food Court                      | 1 |
| American Restaurant             | 1 |

American Restaurant 1

Name: 2nd Most Common Venue, dtype: int64

---

|               |    |
|---------------|----|
| Manhattan     | 28 |
| Brooklyn      | 17 |
| Queens        | 6  |
| Staten Island | 2  |

Name: Borough, dtype: int64

---

Cluster 6:

|                      |    |
|----------------------|----|
| Deli / Bodega        | 19 |
| Pizza Place          | 10 |
| Fast Food Restaurant | 9  |
| Fried Chicken Joint  | 2  |
| Chinese Restaurant   | 1  |
| Coffee Shop          | 1  |
| Bakery               | 1  |
| Food                 | 1  |

Name: 1st Most Common Venue, dtype: int64

---

|                           |    |
|---------------------------|----|
| Pizza Place               | 10 |
| Deli / Bodega             | 9  |
| Fast Food Restaurant      | 6  |
| Fried Chicken Joint       | 3  |
| Chinese Restaurant        | 3  |
| Donut Shop                | 2  |
| Caribbean Restaurant      | 2  |
| Mexican Restaurant        | 2  |
| Taco Place                | 1  |
| Latin American Restaurant | 1  |
| Food                      | 1  |
| Spanish Restaurant        | 1  |
| Food Truck                | 1  |
| Bakery                    | 1  |

American Restaurant 1

Name: 2nd Most Common Venue, dtype: int64

---

|               |    |
|---------------|----|
| Bronx         | 19 |
| Queens        | 12 |
| Brooklyn      | 8  |
| Manhattan     | 4  |
| Staten Island | 1  |

Name: Borough, dtype: int64

---

Cluster 7:



|                             |    |
|-----------------------------|----|
| Pizza Place                 | 34 |
| Deli / Bodega               | 8  |
| Italian Restaurant          | 3  |
| Donut Shop                  | 3  |
| Sushi Restaurant            | 3  |
| Fast Food Restaurant        | 2  |
| Café                        | 2  |
| Thai Restaurant             | 2  |
| Ice Cream Shop              | 2  |
| Food Truck                  | 2  |
| Seafood Restaurant          | 2  |
| Bagel Shop                  | 2  |
| Bakery                      | 2  |
| American Restaurant         | 1  |
| Fish & Chips Shop           | 1  |
| Dessert Shop                | 1  |
| Indian Restaurant           | 1  |
| Eastern European Restaurant | 1  |
| Coffee Shop                 | 1  |
| Korean Restaurant           | 1  |

Name: 1st Most Common Venue, dtype: int64

-----

|                           |    |
|---------------------------|----|
| Deli / Bodega             | 10 |
| Pizza Place               | 9  |
| Coffee Shop               | 7  |
| Donut Shop                | 6  |
| Bakery                    | 6  |
| Chinese Restaurant        | 5  |
| Italian Restaurant        | 4  |
| Bagel Shop                | 4  |
| American Restaurant       | 2  |
| Ice Cream Shop            | 2  |
| Sandwich Place            | 2  |
| Asian Restaurant          | 2  |
| Hot Dog Joint             | 2  |
| Café                      | 2  |
| Middle Eastern Restaurant | 1  |
| Fast Food Restaurant      | 1  |
| Fish & Chips Shop         | 1  |
| Indian Restaurant         | 1  |
| Food                      | 1  |
| Greek Restaurant          | 1  |
| Thai Restaurant           | 1  |
| Food Truck                | 1  |
| Spanish Restaurant        | 1  |
| Japanese Restaurant       | 1  |
| Mexican Restaurant        | 1  |

Name: 2nd Most Common Venue, dtype: int64

-----

|               |    |
|---------------|----|
| Brooklyn      | 25 |
| Queens        | 20 |
| Staten Island | 20 |
| Bronx         | 7  |
| Manhattan     | 2  |

Name: Borough, dtype: int64

-----

Cluster 8:

|                    |    |
|--------------------|----|
| Italian Restaurant | 16 |
| Pizza Place        | 8  |

Name: 1st Most Common Venue, dtype: int64

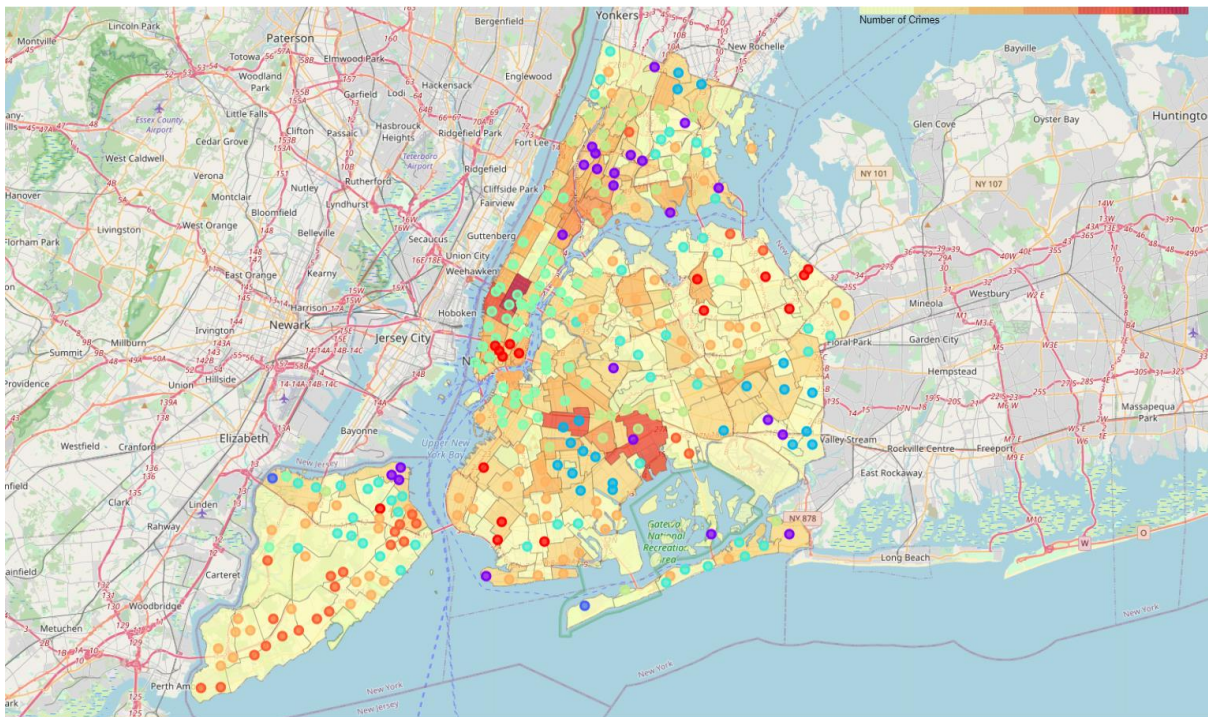
|                     |   |
|---------------------|---|
| Pizza Place         | 7 |
| Italian Restaurant  | 6 |
| Deli / Bodega       | 5 |
| Ice Cream Shop      | 2 |
| Asian Restaurant    | 1 |
| Bakery              | 1 |
| American Restaurant | 1 |
| Café                | 1 |

Name: 2nd Most Common Venue, dtype: int64

|               |    |
|---------------|----|
| Staten Island | 19 |
| Queens        | 4  |
| Bronx         | 1  |

Name: Borough, dtype: int64

Clustering results superimposed on crime map:



## Discussion and Recommendations\*\*

### Crime Data

Park and cemetery area in the Staten Island, the airport, Stuyvesant Town and Cooper Village has the least crime occurred over 2019. On the contrary, Mid-town South, East New York and Union Square areas recorded the most crimes.

Petite larceny is the most common crime followed by second degree harassment and third degree assault. Second degree is the harassment in obscene language with intention, while third degree assault is physical injury attempted with intention.

### **Food Venues**

Cluster 0 has a lot of Chinese, Korean and Japanese restaurants, which shows an East Asian vibes. These neighbourhoods are mostly in Queens and Manhattan.

Cluster 1 is most common for Deli and mostly in Bronx. It might be the school areas and a more get-go neighbourhood. Yet, the features are ambiguous and more information is needed to arrive at a conclusion.

Cluster 2 has little to almost no food venues. It can be concluded that these are rather suburb neighbourhoods of only residential use.

Cluster 3 presents a number of Caribbean restaurants. It shows a Caribbean vibes on mostly Brooklyn and Queens.

Cluster 4 is most common for Deli and mostly in Queens. Yet, the features are ambiguous and more information is needed to arrive at a conclusion.

Cluster 5 has a dominating 47 coffee shops over any other categories, as well as in Manhattan over any other boroughs. It can be concluded that this cluster represents the ghetto of the higher-end.

Cluster 6: is most common for Deli and mostly in Bronx. Yet, the features are ambiguous and more information is needed to arrive at a conclusion.

Cluster 7 has a diverse restaurant type. It is difficult to come to a conclusion of a specific culture that is present in all these neighbourhoods, adding to the fact that a lot of them can be found in Brooklyn, Queens and Staten Island. Thus, it is possible to say that these are the neighbourhoods where all the shopping malls, offices/business activities and centres of activities are concentrated.

Cluster 8 has mostly Italian restaurants and pizza places which can easily induce the conclusion – an Italian focused neighbourhood. Almost all of these neighbourhoods are found in the Staten Island.

For future analysis, it is recommended to add more venue data such as demographics and property price to reveal more valuable information on the neighbourhoods hence bring more insights to the topic.

## **Conclusion**

Crime committed in NYC are petite larceny, 2nd degree harassment and 3rd degree assault. It happened around the neighbourhood of Mid-town south, East New York and Union Square. These neighbourhoods happen to be in Cluster 1, 3, 5 and 6.

On a cultural perspective as promised from the food venue data, each cluster represents a different vibe, yet some might not be conclusive enough as the limited data taken into account (only food venues were assessed while more non-quantitative analysis should also be done to arrive at a more accurate analysis).

This project serves as a preliminary insights report for the neighbourhoods and cultural diversity of NYC.