

# Find the best Neighborhood in Manhattan to open a Restaurant

---

ROOPENE HARIHARAN

13-JUNE-2020

# Entrepreneur trying to launch a new restaurant in Manhattan, New York (US)

---

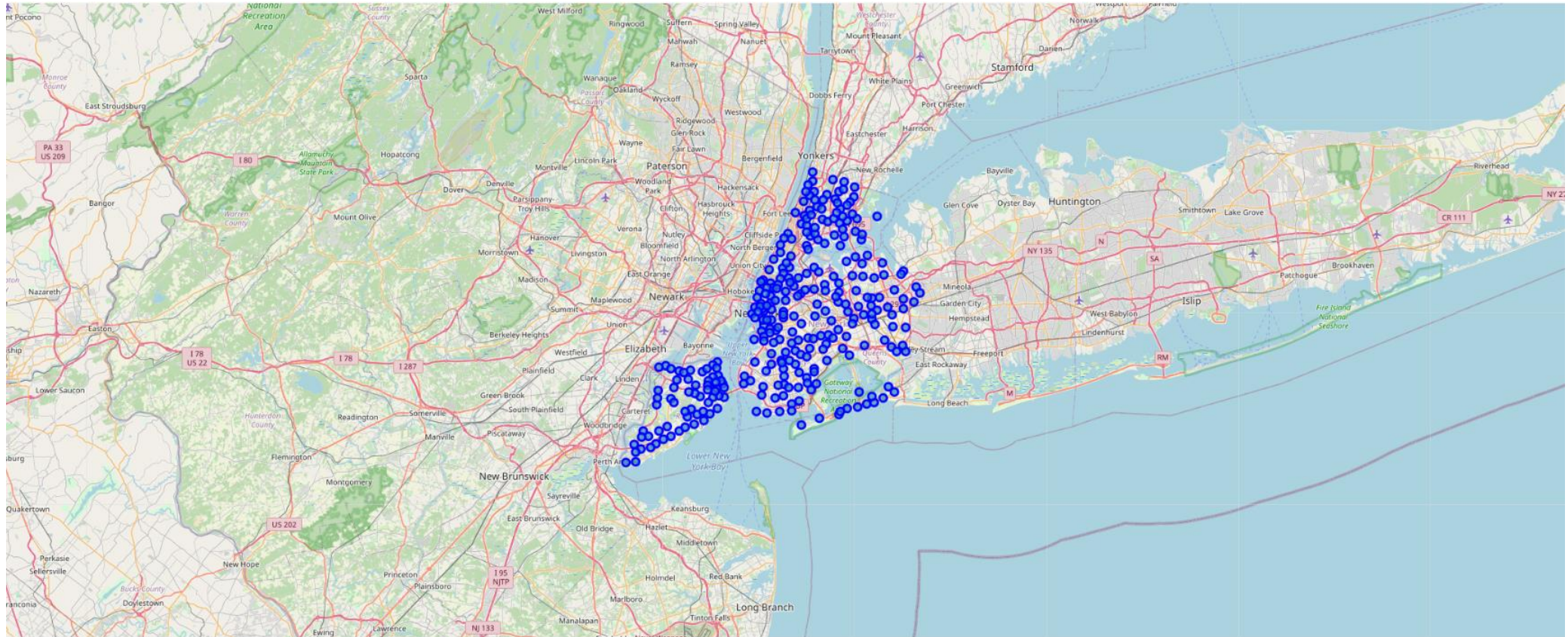
- New York City's food culture includes an array of international cuisines influenced by the city's immigrant history. The city is home to "nearly one thousand of the finest and most diverse haute cuisine restaurants in the world".
- We got to locate the similar neighborhoods with most number of restaurants and suggest that as the best place
- Borough with most number of restaurants in their top 10 most common venues
- Also, considering similar Neighborhoods with same neighborhoods where there are less number of Restaurants, this will be the place where a Restaurant is likely to have better reception.

# Data Acquisition and Data Cleaning

---

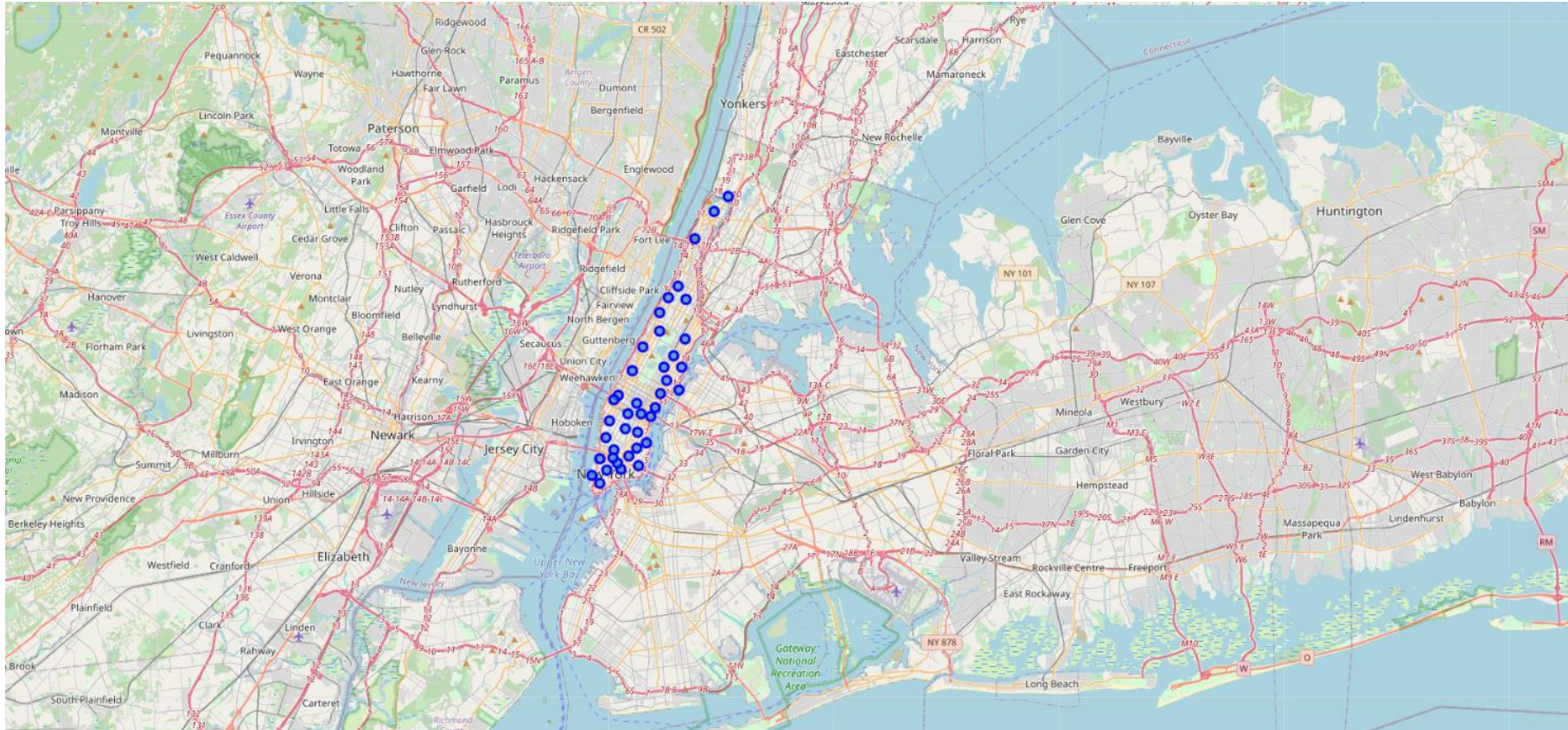
- Our initial Borough and Neighborhood details for entire New York City can be taken from the link [here](#)
- Use the Four Square API service to extract the list of venue categories around each neighborhood.
- Geocoder package can be used to extract the Latitude and Longitude data of each Neighborhood

# Plot of New York Neighborhoods





# Plot the Neighborhoods of Manhattan Borough



# Venue Frequencies for different Neighborhood

| Neighborhood         | Accessories Store | Adult Boutique | Afghan Restaurant | African Restaurant | American Restaurant | Antique Shop | Arcade   | Arepa Restaurant | Argentinian Restaurant | Art Gallery | Art Museum | Arts & Crafts Store | Asian Restaurant | Athletics & Sports | Auditorium | Australian Restaurant |
|----------------------|-------------------|----------------|-------------------|--------------------|---------------------|--------------|----------|------------------|------------------------|-------------|------------|---------------------|------------------|--------------------|------------|-----------------------|
| 0 Battery Park City  | 0.000000          | 0.00           | 0.00              | 0.000000           | 0.015385            | 0.000000     | 0.000000 | 0.000000         | 0.000000               | 0.000000    | 0.000000   | 0.000000            | 0.000000         | 0.015385           | 0.015385   | 0.000000              |
| 1 Carnegie Hill      | 0.000000          | 0.00           | 0.00              | 0.000000           | 0.011494            | 0.000000     | 0.000000 | 0.000000         | 0.011494               | 0.000000    | 0.011494   | 0.000000            | 0.000000         | 0.000000           | 0.000000   | 0.000000              |
| 2 Central Harlem     | 0.000000          | 0.00           | 0.00              | 0.066667           | 0.044444            | 0.000000     | 0.000000 | 0.000000         | 0.000000               | 0.022222    | 0.000000   | 0.000000            | 0.000000         | 0.000000           | 0.000000   | 0.000000              |
| 3 Chelsea            | 0.000000          | 0.00           | 0.00              | 0.000000           | 0.030000            | 0.000000     | 0.000000 | 0.000000         | 0.000000               | 0.070000    | 0.000000   | 0.000000            | 0.000000         | 0.000000           | 0.000000   | 0.000000              |
| 4 Chinatown          | 0.000000          | 0.00           | 0.00              | 0.000000           | 0.030000            | 0.000000     | 0.000000 | 0.000000         | 0.000000               | 0.000000    | 0.000000   | 0.000000            | 0.020000         | 0.000000           | 0.000000   | 0.000000              |
| 5 Civic Center       | 0.000000          | 0.00           | 0.00              | 0.000000           | 0.040404            | 0.010101     | 0.000000 | 0.000000         | 0.000000               | 0.000000    | 0.000000   | 0.000000            | 0.010101         | 0.000000           | 0.000000   | 0.010101              |
| 6 Clinton            | 0.000000          | 0.00           | 0.00              | 0.000000           | 0.030000            | 0.000000     | 0.000000 | 0.000000         | 0.000000               | 0.010000    | 0.000000   | 0.000000            | 0.000000         | 0.000000           | 0.000000   | 0.000000              |
| 7 East Harlem        | 0.000000          | 0.00           | 0.00              | 0.000000           | 0.000000            | 0.000000     | 0.000000 | 0.000000         | 0.000000               | 0.000000    | 0.000000   | 0.000000            | 0.000000         | 0.000000           | 0.000000   | 0.000000              |
| 8 East Village       | 0.000000          | 0.00           | 0.00              | 0.000000           | 0.010000            | 0.000000     | 0.000000 | 0.010000         | 0.010000               | 0.010000    | 0.000000   | 0.010000            | 0.000000         | 0.000000           | 0.000000   | 0.000000              |
| 9 Financial District | 0.000000          | 0.00           | 0.00              | 0.000000           | 0.040000            | 0.000000     | 0.000000 | 0.000000         | 0.000000               | 0.000000    | 0.000000   | 0.000000            | 0.000000         | 0.000000           | 0.000000   | 0.000000              |
| 10 Flatiron          | 0.000000          | 0.00           | 0.00              | 0.000000           | 0.010000            | 0.000000     | 0.000000 | 0.000000         | 0.000000               | 0.010000    | 0.000000   | 0.020000            | 0.000000         | 0.000000           | 0.000000   | 0.000000              |
| 11 Gramercy          | 0.000000          | 0.00           | 0.00              | 0.000000           | 0.036585            | 0.000000     | 0.012195 | 0.000000         | 0.000000               | 0.012195    | 0.000000   | 0.000000            | 0.000000         | 0.000000           | 0.000000   | 0.000000              |
| 12 Greenwich Village | 0.000000          | 0.00           | 0.00              | 0.000000           | 0.010000            | 0.000000     | 0.000000 | 0.000000         | 0.000000               | 0.010000    | 0.000000   | 0.000000            | 0.000000         | 0.000000           | 0.000000   | 0.000000              |
| 13 Hamilton Heights  | 0.000000          | 0.00           | 0.00              | 0.000000           | 0.000000            | 0.000000     | 0.000000 | 0.000000         | 0.000000               | 0.000000    | 0.000000   | 0.000000            | 0.000000         | 0.000000           | 0.000000   | 0.000000              |
| 14 Hudson Yards      | 0.000000          | 0.00           | 0.00              | 0.000000           | 0.067797            | 0.000000     | 0.000000 | 0.000000         | 0.000000               | 0.000000    | 0.000000   | 0.000000            | 0.000000         | 0.000000           | 0.000000   | 0.000000              |

For each Neighborhood there may be many repeated venue categories, for example: “Central Harlem” might have more than 1 American Restaurant.

So to find the most common venues and to have the simplified version of data, I found the frequencies of the venues within a neighborhood

# Top 10 common Venues in Neighborhoods

I sorted the table in descending order of the frequency mean of venues occurrence in each neighborhood.

Created a Dataframe with 10 levels of columns to display top 10 common venues in neighborhoods

|   | Neighborhood      | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|-------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|
| 0 | Battery Park City | Park                  | Hotel                 | Memorial Site         | Gym                   | Coffee Shop           | Playground            | Food Court            | Mexican Restaurant    | Shopping Mall         | Gourmet Shop           |
| 1 | Carnegie Hill     | Coffee Shop           | Café                  | Yoga Studio           | Bookstore             | Gym / Fitness Center  | Gym                   | Italian Restaurant    | Pizza Place           | Wine Shop             | Vietnamese Restaurant  |
| 2 | Central Harlem    | African Restaurant    | Seafood Restaurant    | Gym / Fitness Center  | American Restaurant   | Bar                   | French Restaurant     | Chinese Restaurant    | Café                  | Boutique              | Market                 |
| 3 | Chelsea           | Coffee Shop           | Art Gallery           | Ice Cream Shop        | Café                  | American Restaurant   | Bakery                | Pizza Place           | Cocktail Bar          | Market                | Seafood Restaurant     |
| 4 | Chinatown         | Chinese Restaurant    | Bakery                | Cocktail Bar          | Bubble Tea Shop       | Coffee Shop           | Optical Shop          | Bar                   | Spa                   | American Restaurant   | Ice Cream Shop         |

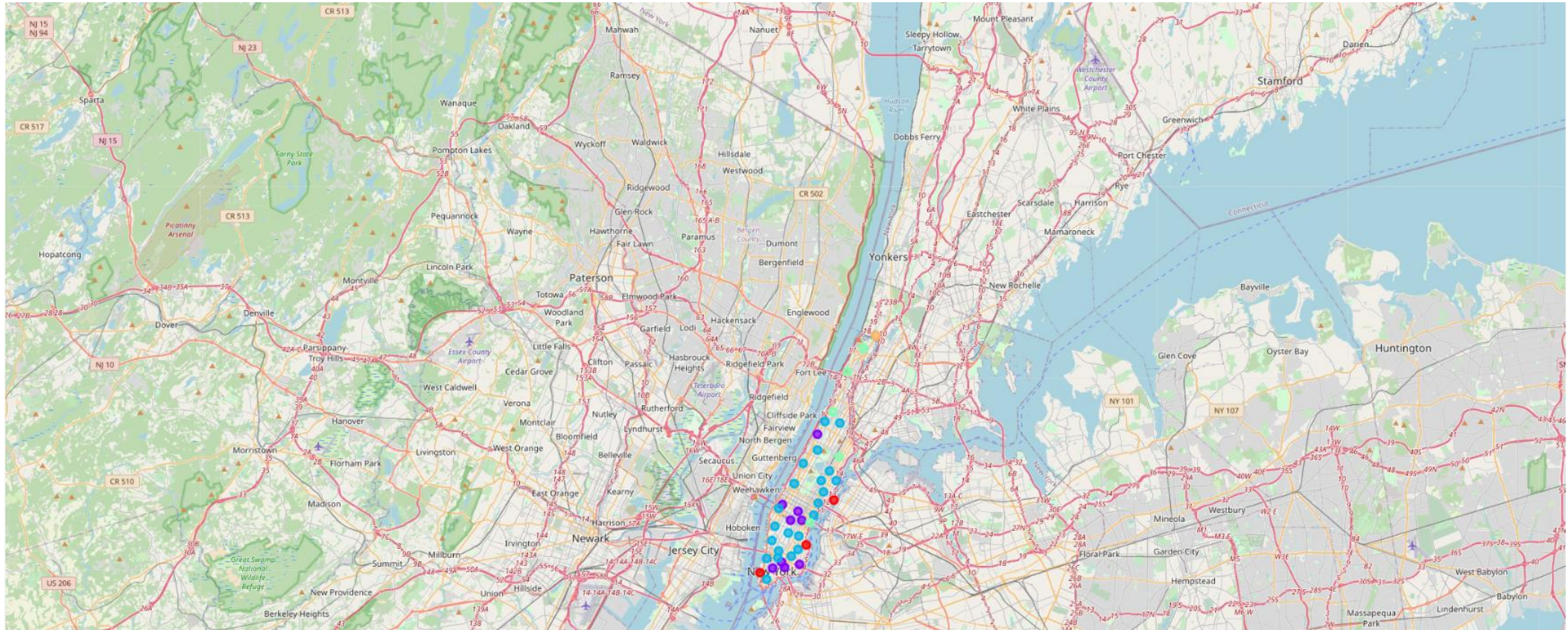
# k-Means Clustering

---

- k-Means clustering is a method of vector quantization, originally from signal processing that aims to partition  $n$  observations into  $k$  number clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.
- I plan to specify the number of clusters to be 5 as initial setup, which seems to be optimum or the New York model. My random state for the k-Means would be 0. This would give me a picture of which neighborhoods share the same pattern of lifestyles.



# Clustering of Neighborhoods with different lifestyle



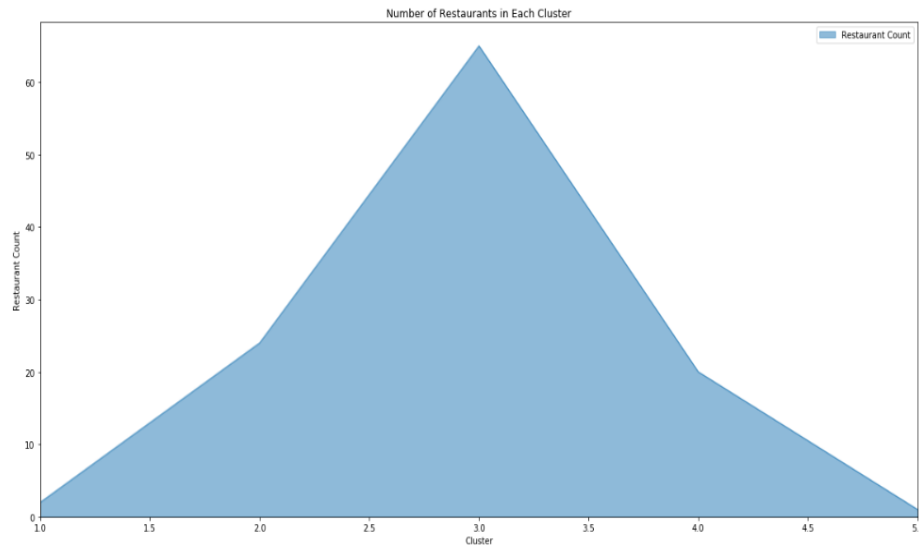
| Restaurant Count |    |
|------------------|----|
| Cluster          |    |
| 1                | 2  |
| 2                | 24 |
| 3                | 65 |
| 4                | 20 |
| 5                | 1  |

# Venue Frequencies for different Clusters

To find the count of restaurant in each Cluster, we need to find the venues with the category string as “Restaurant” and find the total count for each cluster

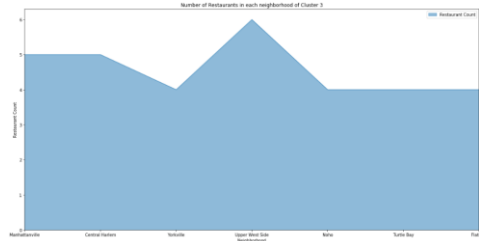
From this we find the Cluster with the greatest number of Restaurants i.e., Cluster 3 in this.

Most importantly we are referring to the DataFrame we created with top 10 common venues.



# Venue Frequencies for different Neighborhoods in Cluster 3

---



|                 | Restaurant Count |
|-----------------|------------------|
| Neighborhood    |                  |
| Manhattanville  | 5                |
| Central Harlem  | 5                |
| Yorkville       | 4                |
| Upper West Side | 6                |
| Noho            | 4                |
| Turtle Bay      | 4                |
| Flatiron        | 4                |

We found that Cluster 3 has more number of Restaurants.

We now need to analyze Neighborhoods of Cluster 3 to get the place where we can start the New Restaurant.

We are referring to the Dataframe created for Cluster 3 here.

# Insights

---

**Insight 1** – The first histogram gives a clear picture that only in Cluster 3 we have higher number of restaurants with completely different lifestyle in it.

**Insight 2** – In the second histogram, we come to know that “Upper West Side” has the highest restaurant count compared to other places

**Insight 3** – Also, regarding the other neighborhoods with same lifestyle, there are nearly less number of restaurants.



# Conclusion

---

**Strategy 1** – Wanting to Launch a Restaurant in a Neighborhood where majority of venues are Restaurants, considering they would catch the same growth as other venues

For this strategy, Insight 1,2 would help to decide as it explains the neighborhoods with most number of restaurants is cluster 3. So my recommendation would be to launch a new restaurant in Cluster 3 more across all boroughs, but specifically on Upper West Side, as they have the most number of restaurants as the top 10 common venues

**Strategy 2** - Wanting to Launch a Restaurant in a Neighborhood where people likely to visit Restaurant but they have moderate number of Restaurants

For this strategy, Insight 2 and 3 would help to decide as it explains the neighborhood with opportunity to have a greater number of restaurants but have moderate restaurants. So, my recommendation would be to launch a new restaurant in Cluster 3 in Sutton Place can be a good place to start a New Restaurant.

## Future directions

---

As we have the Venue Category as my most trusted data feature to determine the recommendation of the neighborhood for launching new restaurant. In future, we can also avail other data that Four Square API provides, that is we can also include User reviews and how frequent users visit these restaurants, to derive a rating of these restaurants or identifying most suitable neighborhood which has more number of visitors to the restaurants.