

# Battle of the Neighborhoods-Vegan Restaurants in New York City

## Project Motivation and Background

### New York City:

New York City is one of the most populous cities in the world. With an estimated population of over 8 million people and over 65 million visitors in 2018, New York is very diverse and has many different offerings for its many tourists and residents. New York City has over 24,000 restaurants, of many different varieties and cuisines, with approximately 10,000 located in Manhattan alone.

### Vegan/Vegetarian Restaurants:

Vegan and vegetarian diets have grown significantly in popularity in the last few years. Many people may choose to follow plant-based diets for a variety of different reasons, such as concern for animal welfare, positive health benefits or environmental concerns. There are currently over 9 million Americans who adhere to a plant-based diet and many of those people will inevitably travel to locations around the world and will look for places to eat that may accommodate dietary needs. In 2018, the global vegan food market was estimated to be over \$12 billion, which shows that there is a definite market for plant-based foods around the world.

### Target Audience:

The audience for this project could be either a person or company looking to open a plant-based restaurant, or a tourist coming to NYC and looking to find plant-based food options while there. As seen above, with a number of visitors coming to New York City each year and with a significant number of Americans now adhering to plant-based diets, there could certainly be a market for vegan or vegetarian cuisine in New York.

### Data and Resources:

This project utilizes publicly available data regarding New York City neighborhoods & boroughs from [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset). We will also be utilizing the Foursquare API to find locations and information regarding existing vegan/vegetarian restaurants already in NYC (specifically Manhattan borough).

## Data Usage & Project Plan-Project Steps:

### Data Collection & Cleaning

As mentioned above, the data regarding New York City neighborhoods and boroughs was gathered from [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset). The json file was loaded and labeled ny\_data.

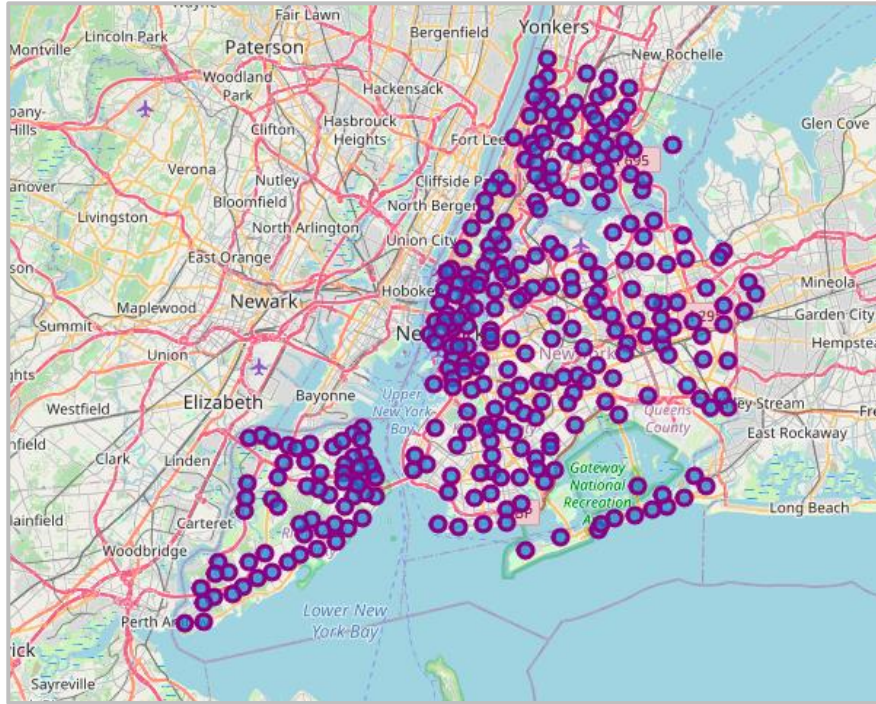
```
In [4]: ny_data

Out[4]: {'type': 'FeatureCollection',
        'totalFeatures': 306,
        'features': [{'type': 'Feature',
                        'id': 'nyu_2451_34572.1',
                        'geometry': {'type': 'Point',
                                    'coordinates': [-73.84720052054902, 40.89470517661]},
                        'geometry_name': 'geom',
                        'properties': {'name': 'Wakefield',
                                      'stacked': 1,
                                      'annoline1': 'Wakefield',
                                      'annoline2': None,
                                      'annoline3': None,
                                      'annoangle': 0.0,
                                      'borough': 'Bronx',
                                      'bbox': [-73.84720052054902,
                                              40.89470517661,
                                              -73.84720052054902,
                                              40.89470517661]}},
                      {'type': 'Feature',
                        'id': 'nyu_2451_34573.1',
                        'geometry': {'type': 'Point',
                                    'coordinates': [-73.84720052054902, 40.89470517661]},
                        'geometry_name': 'geom',
                        'properties': {'name': 'Wakefield',
                                      'stacked': 1,
                                      'annoline1': 'Wakefield',
                                      'annoline2': None,
                                      'annoline3': None,
                                      'annoangle': 0.0,
                                      'borough': 'Bronx',
                                      'bbox': [-73.84720052054902,
                                              40.89470517661,
                                              -73.84720052054902,
                                              40.89470517661]}}
```

The data was transformed into a pandas dataframe called nyc\_neighborhoods which showed information for Borough, Neighborhood, Latitude and Longitude and let us know that there are 5 boroughs and 306 neighborhoods in New York City.

	Borough	Neighborhood	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

In order to gain accurate geographic data for use later in the analysis, the geopy library was used via Nominatum and Folium. This showed us that New York City’s geographical coordinates are 40.7127281, -74.0060152. A map was subsequently produced with the New York City neighborhoods superimposed as a reference.



As New York City has a significant amount of data due to its large size, for the purposes of this project, I decided to focus on Manhattan specifically. As seen above, a new dataframe was created and we obtained the borough's geographical coordinates (40.7896239,-73.9598939).

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900

In addition to the coordinates, a map was created specifically for Manhattan, once again with the neighborhoods superimposed for reference.







The rows of our data frame were then grouped by neighborhood and I took the mean of the frequency of occurrence for each venue category. While many venue categories serve vegan/vegetarian foods, the top venue frequency was, unsurprisingly, vegan/vegetarian specific restaurants.

	Neighborhood	American Restaurant	Asian Restaurant	Bagel Shop	Salad Place	Sandwich Place	Sushi Restaurant	Thai Restaurant	Vegetarian / Vegan Restaurant
0	Battery Park City	0.000000	0.000000	0.000000	0.153846	0.000000	0.000000	0.000000	0.461538
1	Carnegie Hill	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.875000
2	Central Harlem	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.833333
3	Chelsea	0.000000	0.029412	0.000000	0.058824	0.000000	0.000000	0.000000	0.500000
4	Chinatown	0.000000	0.022222	0.000000	0.022222	0.000000	0.000000	0.000000	0.688889
5	Civic Center	0.000000	0.028571	0.000000					

The dataframe was then sorted and a new dataframe was created and printed with the top 10 most common venues serving vegan/vegetarian food in each neighborhood.

	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	Vegetarian / Vegan Restaurant	Salad Place	Falafel Restaurant	Food Truck	Juice Bar	Burger Joint	Eastern European Restaurant	Gluten-free Restaurant	Filipino Restaurant	Fast Food Restaurant
1	Carnegie Hill	Vegetarian / Vegan Restaurant	Juice Bar	Eastern European Restaurant	Grocery Store	Gluten-free Restaurant	Food Truck	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Dessert Shop
2	Central Harlem	Vegetarian / Vegan Restaurant	Market	Grocery Store	Gluten-free Restaurant	Food Truck	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Eastern European Restaurant	Dessert Shop
3	Chelsea	Vegetarian / Vegan Restaurant	Food Truck	Salad Place	Juice Bar	Café	Mediterranean Restaurant	Ice Cream Shop	Gluten-free Restaurant	Health Food Store	Mexican Restaurant
4	Chinatown	Vegetarian / Vegan Restaurant	Pizza Place	Mediterranean Restaurant	Japanese Restaurant	Juice Bar	Dessert Shop	Latin American Restaurant	Deli / Bodega	Food Truck	Grocery Store

## K-Means Neighborhood Clustering Analysis on NYC & Manhattan Locations

In order to perform a more in-depth location analysis, K-Means Clustering was used to divide the many Manhattan neighborhoods into just five clusters. A new dataframe was subsequently created with the top 10 most common venues serving vegan/vegetarian food in each cluster.

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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	Manhattan	Marble Hill	40.876551	-73.910660	2.0	Vegetarian / Vegan Restaurant	Ice Cream Shop	Grocery Store	Gluten-free Restaurant	Food Truck	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Eastern European Restaurant	Dessert Shop
1	Manhattan	Chinatown	40.715618	-73.994279	1.0	Vegetarian / Vegan Restaurant	Pizza Place	Mediterranean Restaurant	Japanese Restaurant	Juice Bar	Dessert Shop	Latin American Restaurant	Deli / Bodega	Food Truck	Grocery Store
2	Manhattan	Washington Heights	40.851903	-73.936900	0.0	Lounge	Vegetarian / Vegan Restaurant	Eastern European Restaurant	Grocery Store	Gluten-free Restaurant	Food Truck	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Dessert Shop
3	Manhattan	Inwood	40.867684	-73.921210	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	Manhattan	Hamilton Heights	40.823604	-73.949688	2.0	Vegetarian / Vegan Restaurant	Ice Cream Shop	Grocery Store	Gluten-free Restaurant	Food Truck	Filipino Restaurant	Fast Food Restaurant	Falafel Restaurant	Eastern European Restaurant	Dessert Shop

After creating the above dataframe, I once again created a map to visualize the clusters that were created earlier.



### Analyze Results

In this section, each cluster was analyzed similar to previous examples, by neighborhood and top 10 most common venues. The first cluster contained the Washington Heights neighborhood and the top three venues were Lounge, Vegetarian/Vegan Restaurant and Eastern European restaurant.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
2	Washington Heights	Lounge	Vegetarian / Vegan Restaurant	Eastern European Restaurant

The second cluster in our analysis contained a variety of neighborhoods (Chinatown, Upper East Side, etc.) and the most common venue was Vegetarian/Vegan Restaurants, followed by Salad Places and Juice Bars.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
Chinatown	Vegetarian / Vegan Restaurant	Pizza Place	Mediterranean Restaurant
Upper East Side	Vegetarian / Vegan Restaurant	Salad Place	Juice Bar
Lenox Hill	Vegetarian / Vegan Restaurant	Salad Place	Bagel Shop
Roosevelt Island	Vegetarian / Vegan Restaurant	Salad Place	Dessert Shop
Lincoln Square	Vegetarian / Vegan Restaurant	Italian Restaurant	Salad Place

Clusters 3, 4 and 5 also all contain multiple neighborhoods and again, the most common venues are Vegetarian/Vegan Restaurants.

### Conclusion:

From our analysis of the food offerings in Manhattan, we can conclude that the venue containing Vegetarian/Vegan cuisines is Vegetarian/Vegan restaurants. Salad Places and Juice Bars also seem to also carry these items more than many of the other cuisines in Manhattan. From our onehot encoding, we can see that several neighborhoods such as Chelsea, SoHo and Lower East Side have multiple venues aside from purely Vegetarian/Vegan establishments that serve this cuisine. A vegan or vegetarian tourist coming to visit Manhattan will undoubtedly have many different restaurant choices to choose from.

From a business perspective, we can see that Vegetarian/Vegan restaurants have the market share on this cuisine, however there are still a number of Salad Places, Juice Bars and ethnic cuisine venues that carry more offerings. Someone who is looking to open a restaurant that serves Vegetarian/Vegan friendly cuisine, may want to look at opening a restaurant with a variety of cuisines but still having a large menu catering to plant-based eaters in order to improve variety and cater to multiple palettes. Further analysis could be completed on demographics in these neighborhoods to further see what kind of cuisine could be offered that would be most profitable and cater to vegetarians/vegans along with the general population.