Data Science Capstone Project week 2

Battle of the Neighborhoods

By Teresa Lau

---

1. Introduction/Business Problem

Background

There is a growing trend of plant-based diets in USA. New York City is one of the top cities in USA in terms of the number of vegetarian businesses in the city. This presents a good opportunity for vegetarian restaurant's owner to expand their business.

Problem Description

This project will focus on finding the best neighborhood(s) in New York to open up a new vegetarian restaurant.

First, we analyze data of vegetarian restaurants in New York City to determine neighborhoods that are currently successful in doing vegetarian business. Next, we collect information on popular venues from all neighborhoods to determine neighborhoods that are suitable for opening new vegetarian restaurants. Lastly, we use visualizations to help us make recommendations to the client.

Target Audience

The target audience is for a client who wants to open a vegetarian restaurant in New York City. This report will recommend the best neighborhood(s) to open the new vegetarian restaurant and also give insights on important factors to consider when opening the new restaurant.

2. Data Sources

Neighborhood Location Data

We use data from NYU Spatial Data Repository https://geo.nyu.edu/catalog/nyu_2451_34572) to determine the neighborhoods and location that we are interested in.
File: Raw_NYC_BLY_Location.csv

Borough *Type: String*	Neighborhood *Type: String*	Latitude *Type: String*	Longitude *Type: String*
Manhattan	Marble Hill	40.87655077879964	-73.91065965862981
Brooklyn	Greenpoint	40.7302009848647	-73.95424093127393

Venue Data

Based on each neighborhood location, we will use Four Square crowdsource data to get the following data.

- Popular Venues by Neighborhood using Explore API
    - https://api.foursquare.com/v2/venues/explore
    - Parameters: Section=food
    - File: Raw_NYC_BLY_Food_Venues.csv

Neighborhood *Type: String*	Neighborhood Latitude *Type: String*	Neighborhood Longitude *Type: String*	Id *Type: String*	Venue *Type: String*	Venue Latitude *Type: String*	Venue Longitude *Type: String*	Venue Category *Type: String*
Marble Hill	40.87655077879964	-73.91065965862981	4b4429abf964a52(	Arturo's	40.8744117711023	-73.9102710098157	Pizza Place
Marble Hill	40.87655077879964	-73.91065965862981	4b79cc46f964a52(	Tibbett Diner	40.8804044222466	-73.9089373800640	Diner
Marble Hill	40.87655077879964	-73.91065965862981	4b5357adf964a52(	Dunkin'	40.8771358420158	-73.9066655070141	Donut Shop

- Vegetarian / Vegan Venues by Neighborhood using Search API
    - https://api.foursquare.com/v2/venues/search
    - Parameters: Category= '4bf58dd8d48988d1d3941735' *# Vegatarian*
    - File: Raw_NYC_BLY_Veg_Venues.csv

Neighborhood *Type: String*	Neighborhood Latitude *Type: String*	Neighborhood Longitude *Type: String*	Id *Type: String*	Venue *Type: String*	Venue Latitude *Type: String*	Venue Longitude *Type: String*	Venue Category *Type: String*
Marble Hill	40.87655077879964	-73.91065965862981	52411ca98bbd75ed	Kingsbridge-Riverdal	40.87939442134257	-73.90712512314228	Vegetarian / Vegan Re
Greenpoint	40.7302009848647	-73.95424093127393	4b9709fcf964a520c	Paulie Gee's	40.72980124698718	-73.95852046352437	Pizza Place
Greenpoint	40.7302009848647	-73.95424093127393	53dd1ead498e9ef3f	Jungle Cafe	40.730201	-73.954761	Vegetarian / Vegan Re
Greenpoint	40.7302009848647	-73.95424093127393	4ff37cb9e4b0b8fda8	Adelina's	40.73030130113847	-73.95335193287592	Italian Restaurant

- For each. Vegetarian / Vegan Venues, we use the Venues Details endpoint to get Rating and Price info
    - https://api.foursquare.com/v2/venues/VENUE_ID
    - File: Raw_Veg_Details.csv

Id *Type: String*	Price *Type: String*	Rating *Type: String*
3fd66200f964a52056ee1ee3	Cheap	8.8
3fd66200f964a52088e91ee3	Moderate	8.7
3fd66200f964a520b1e91ee3	Expensive	8.8
3fd66200f964a520cae41ee3	Cheap	7.5

Neighborhood Demographics & Economic Data

Using NYC planning population fact finder we will get the median age for neighborhoods. This data comes from the 2013-2017 American Community Survey (ACS).

https://popfactfinder.planning.nyc.gov/profile/12005/demographic

File: Raw_Neighborhoods_Demographics.csv

Neighborhood	Median Age
East Village	32.3
North Side (Brooklyn)	32.1
South Side (Brooklyn)	32.1
Midtown	35.5

3. Methodology

Step 1: Exploratory Data Analysis using venue data

Our study is based on vegetarian restaurants in Manhattan Borough and selected neighborhoods in Brooklyn Borough. Using Four Square API, we will get list of Vegetarian/Vegan restaurants from these neighborhoods. We will get details such as rating and price-range of each restaurant. From that we use visualization techniques to determine neighborhoods that has most number of popular vegetarian restaurants (based on rating). Those will be the neighborhoods we are interested to study.

Step 2: Machine Learning Technique : K-means clustering

Using FourSquare API, we will get a list of all popular food venues from Manhattan and selected Brooklyn neighborhoods. Using one-hot encoding,we generate a Panda DataFrame that contains all these neighborhoods, with the different type of venues listed as columns. These columns do not have a predefined label and are considered unsupervised data. In order to see the similarity between neighborhoods, we can use a machine learning technique called K-means clustering algorithm on unsupervised data to label similar neighborhoods into same cluster.

Step 3: Visualization

Once we have the clusters generated, we can see neighborhoods that are similar to the vegetarian-popular neighborhoods that we determined in step 1. We will visualize these candidates in a Folium map. These candidate neighborhood should have similar clientele as the vegetarian-friendly neighborhood.

Step 4: Analysis & Recommendation

We will present the visualizations to the client and and let them make their decision by weighing the factors that are important to them. Here are things to consider :

- Neighborhood currently with too few vegetarian restaurants may not have much demand for vegetarian diet.
- Neighborhoods that are too close to the existing successful neighborhood may not be good as competition will be great.
- Neighborhoods already with too many vegetarian venues should be ruled out as the demand could be saturated.
- Age group of the neighborhood

The best neighborhood are ones that already have some vegetarian venues but not too saturated. Also, they should similar age group as the vegetarian-friendly neighborhood.

4. Results

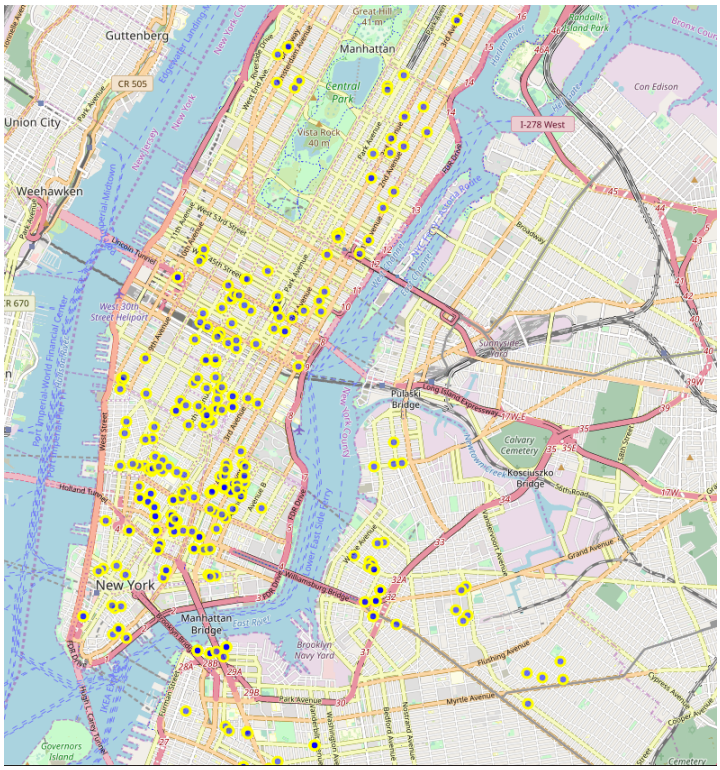
Summary of Data Used

- 54 neighborhoods (In Manhattan and Brooklyn Borough)
- 261 vegetarian venues
- 4018 food venues for clustering
- Please see VegRestaurantAnalysis.ipynb for the Jupyter Notebook I used for this analysis
- Please note that I also published a separate notebook Capstone_GetData.ipynb which is used to get data from Four Square and save it to file. These files are then used by VegRestaurantAnalysis.ipynb

Exploratory Analysis Results

(a) Vegetarian Venues by Location

We produced a Folium map to show existing vegetarian restaurants. We note a high concentration of them in lower Manhattan. We also noticed some area of Brooklyn also have reasonable concentration. Upper Manhattan is more sparsely populated with vegetarian restaurants.



(b)Top Vegetarian Neighborhoods Selection

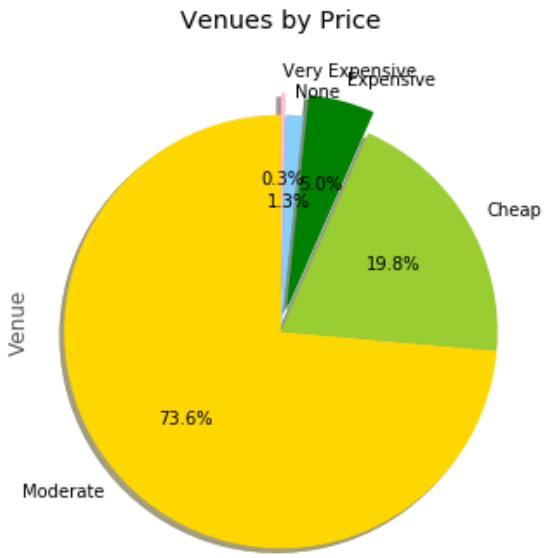
We used 2 criteria to find the neighborhoods of interest

- Neighborhood with most number of vegetarian restaurants
- Neighborhood with most number of high-rated (rating > 7.5) vegetarian restaurants

Based on above criteria, we found the top 3 neighborhoods for successful vegetarian venues are: **Flatiron, East Village** and **Greenwich Village.** We will try to find neighborhoods which are similar to these three neighborhoods.

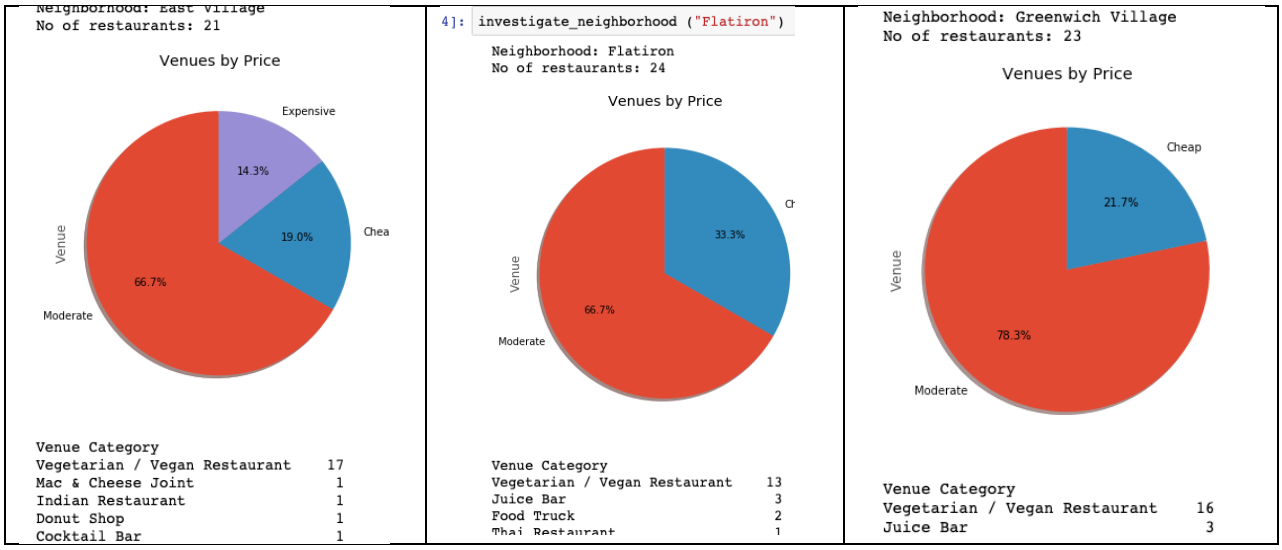
(c) Vegetarian Restaurants by Price Range

This analysis is useful for the client when determining the price range of the restaurant. 75% of current vegetarian restaurants are moderately priced, ehile 20% are in the cheap priced range.



(d)Detailed Analysis of the 3 Neighborhoods we are interested in

This gives an overview of the existing popular vegetarian neighborhood We can see the total number of vegetarian restaurants, their price range and their category in each neighborhood. Note that vegetarian restaurants are often categorized in multiple category, and Juice Bar, Salad Place are often categories we found when searching under the big category Vegetarian.



Cluster Results

Using K Means Elbow Method, we use Cluster = 15 to generate our cluster. The results are generated are saved to Result_NYC_FoodClusterTopVenue.csv. We can see how neighborhoods with similar top 10 popular venues are grouped together. The clusters we found of interest are:

a. Cluster 3 (where East Village is)

	Neighborhood	Cl	1st Most Common	2nd Most Common	3rd Most Common	4th Most Common	5th Most Common	6th Most Common	7th Most Common	8th Most Common	9th Most Common	10th Most Common
16	Manhattan Valley	3	Indian Restaurant	Pizza Place	Deli / Bodega	Mexican Restaurant	Thai Restaurant	Chinese Restaurant	Szechuan Restaurant	CafV©	French Restaurant	Latin American Rest
17	Midtown	3	Sandwich Place	American Restaurant	Bakery	Japanese Restaurant	Steakhouse	Pizza Place	French Restaurant	Burger Joint	Mexican Restaurant	Indian Restaurant
18	Lower East Side	3	Pizza Place	Bakery	Ramen Restaurant	Chinese Restaurant	CafV©	Deli / Bodega	Japanese Restaurant	Latin American Resta	Sandwich Place	Food Truck
19	Battery Park City	3	Pizza Place	Chinese Restaurant	Italian Restaurant	Sandwich Place	Donut Shop	Seafood Restaurant	BBQ Joint	Food Court	Burger Joint	Sushi Restaurant
20	North Side	3	Pizza Place	Burger Joint	American Restaurant	Mexican Restaurant	Italian Restaurant	Bakery	Vegetarian / Vegan F	Breakfast Spot	Asian Restaurant	Food Truck
21	Downtown	3	Pizza Place	Sandwich Place	Food Truck	Burger Joint	Caribbean Restauran	American Restaurant	Asian Restaurant	French Restaurant	Chinese Restaurant	Bakery
22	Chinatown	3	Chinese Restaurant	Bakery	Vietnamese Restaura	Dim Sum Restaurant	Dumpling Restaurant	Mexican Restaurant	Hotpot Restaurant	Cantonese Restauran	Noodle House	Sandwich Place
23	Chelsea	3	Bakery	CafV©	French Restaurant	Italian Restaurant	Pizza Place	American Restaurant	Seafood Restaurant	Mexican Restaurant	Tapas Restaurant	Sandwich Place
24	South Side	3	Pizza Place	American Restaurant	Burger Joint	Mexican Restaurant	Chinese Restaurant	Deli / Bodega	Japanese Restaurant	Latin American Resta	Taco Place	Indian Restaurant
25	East Village	3	Pizza Place	Vietnamese Restaura	Chinese Restaurant	Italian Restaurant	Japanese Restaurant	Mexican Restaurant	Vegetarian / Vegan F	Deli / Bodega	French Restaurant	Bagel Shop

b. Cluster 0 (where Flatiron and Greenwich Village are)

Neighborhood	Cl	1st Most Common	2nd Most Common	3rd Most Common	4th Most Common	5th Most Common	6th Most Common	7th Most Common	8th Most Common	9th Most Common	10th Most Comm
Little Italy	0	Italian Restaurant	Bakery	CafV©	Chinese Restaurant	Mediterranean Resta	Sandwich Place	Thai Restaurant	French Restaurant	Seafood Restaurant	Japanese Restaura
West Village	0	Italian Restaurant	American Restaurant	New American Resta	Gastropub	French Restaurant	Mexican Restaurant	Japanese Restaurant	Seafood Restaurant	Pizza Place	CafV©
Upper East Side	0	Italian Restaurant	American Restaurant	Pizza Place	Diner	French Restaurant	Sushi Restaurant	Bakery	Deli / Bodega	Bagel Shop	Salad Place
Greenwich Village	0	Italian Restaurant	CafV©	Sushi Restaurant	American Restaurant	Indian Restaurant	Pizza Place	Seafood Restaurant	Chinese Restaurant	French Restaurant	Sandwich Place
Flatiron	0	Italian Restaurant	American Restaurant	New American Resta	CafV©	Mediterranean Resta	Mexican Restaurant	Japanese Restaurant	Sandwich Place	Bakery	Vegetarian / Vegar
Soho	0	Italian Restaurant	French Restaurant	CafV©	Mediterranean Resta	American Restaurant	Seafood Restaurant	Bakery	Sandwich Place	Pizza Place	Asian Restaurant
Civic Center	0	Italian Restaurant	Sandwich Place	American Restaurant	French Restaurant	Mexican Restaurant	CafV©	Bakery	Diner	Pizza Place	Sushi Restaurant

5. Discussion

(a) Potential Neighborhoods Based on Clustering

Based on above clustering, we found that

- Greenwich Village and Flatiron are in a **Cluster 0** where popular venues are mostly Italian Restaurant, American Restaurant, Mediterranean Restaurant etc.
- East Village is in **Cluster 3** where popular venues are Pizza Place, Café, Asian, Mexican Restaurants.

All neighborhoods in these two clusters are candidates for our new vegetarian restaurants. Next we will consider some criteria to narrow down the candidates.

(b) Eliminating Neighborhoods by number of existing vegetarian restaurants

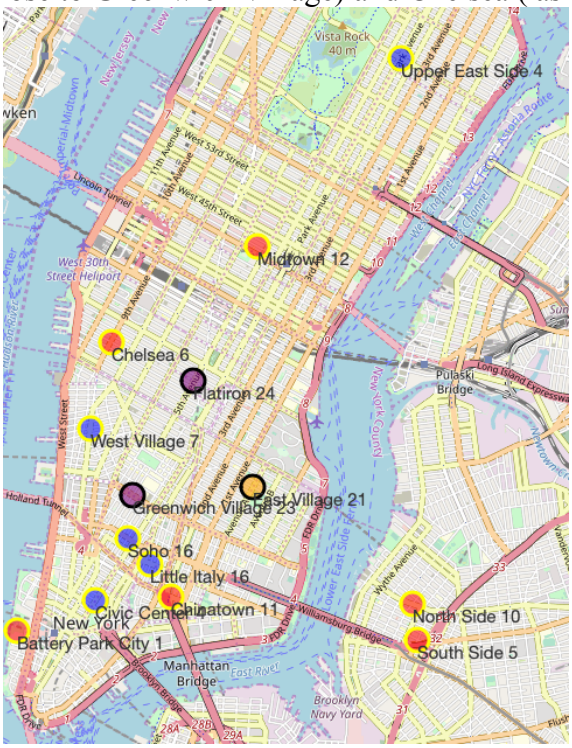
Neighborhoods that currently have zero vegetarian restaurants are too risky to open up a new one as there does not seem to be enough demand currently, so we eliminate them.

Cluster 0 (similar to Flatiron, Greenwich Village)	
Neighborhood	**Veg Restaurants**
Little Italy	16
Soho	16
Upper East Side	4
Greenwich Village	23
Flatiron	24
West Village	7
Civic Center	4

Cluster 3 (similar to East Village)	
Neighborhood	**Veg Restaurants**
Downtown (Brooklyn)	5
Chinatown	11
~~Lower East Side~~	~~0~~
~~Manhattan Valley~~	~~0~~
Midtown	12
North Side(Brooklyn)	10
Chelsea	6
South Side(Brooklyn)	5
East Village	21
~~Battery Park City~~	~~1~~

(c) Eliminating Neighborhoods by proximity to established vegetarian neighborhoods

Next we graph remaining candidates on the Folium map. The ones that are too close to the existing successful neighborhood should be ruled out as there are already many established vegetarian venues, and it may not be easy to compete with them. Here we further rule out West Village (as it is too close to Greenwich Village) and Chelsea (as it is too close to Flatiron)



Cluster 0 (similar to Flatiron, Greenwich Village)	
Neighborhood	**Veg Restaurants**
Little Italy	16
Soho	16
Upper East Side	4
Greenwich Village	**23**
Flatiron	**24**
~~West Village~~	~~7~~
Civic Center	4

Cluster 3 (similar to East Village)	
Neighborhood	**Veg Restaurants**
Downtown	5
Chinatown	11
Midtown	12
North Side	10
~~Chelsea~~	~~6~~
South Side	5
East Village	**21**

(d) Eliminating Neighborhoods by age differences

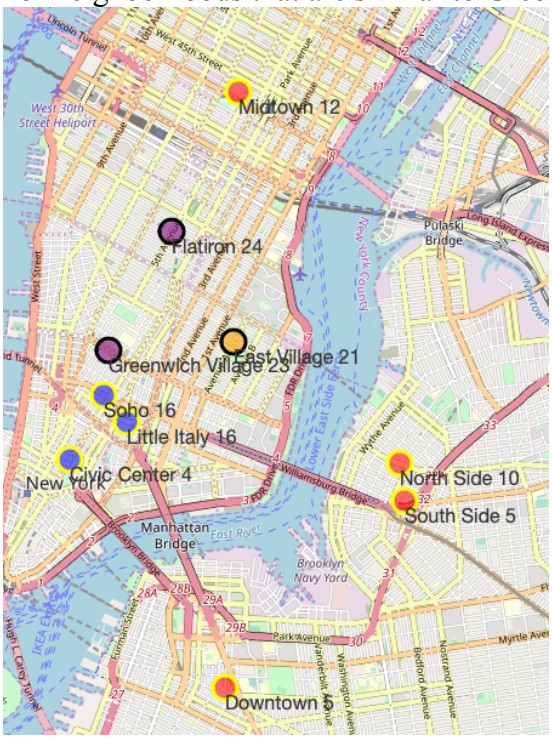
Next we found that the vegetarian-friendly neighborhood tend to have younger people (around age 35) as can be seen from the median age of these 3 popular neighborhoods. (East Village, Greenwich Village, Flatiron). Based on age, we further eliminated Chinatown and Upper East Side

Neighborhood	Median Age
East Village	32.3
North Side (Brooklyn)	32.1
South Side (Brooklyn)	32.1
Midtown	35.5
Battery Park City	33.4
Downtown (Brooklyn)	33.9
~~Chinatown~~	~~41~~

Neighborhood	Median Age
Greenwich Village	36.8
Flatiron	37.1
Soho	36.3
Civic Center	36.3
Little Italy	36.3
~~Upper East Side~~	~~49.5~~

(e) Plotting Remaining Candidates

We plot the final remaining candidate on the map with its current number of vegetarian restaurants. The red dots are location of neighborhoods that are similar to East Village. The blue dots are location of neighborhoods that are similar to Greenwich Village and Flatiron



6. Conclusion

Our recommendation to the client is to open a Vegetarian restaurants in one of these neighborhoods we listed below. These neighborhoods already have some vegetarian restaurants, so it is a desirable area.. Based on the clustering and also median age consideration, the people in neighborhoods who liked vegetarian also liked these neighborhoods. Further, since the number of vegetarian restaurants here are not yet saturated, there is room to get more businesses here by opening more vegetarian venues.

- Location to open a restaurant that is similar to a East Village vegetarian restaurant
    - North Side (Brooklyn)
    - South Side (Brooklyn)
    - Downtown (Brooklyn)
    - Midtown (Manhattan)
- Location to open a restaurant that is similar to a Flatiron or Greenwich Village restaurant
    - Civic Center (Manhattan)
    - Little Italy(Manhattan)
    - Soho (Manhattan)

The client can then further make his/her decision based on

- The kind of vegetarian restaurant they want to open
- Rental cost of the neighborhood and availability
- Foot traffic of the neighborhood