Capstone Project - The Battle of Neighborhoods (Week 2)

Finding out the best neighbor for delivery company to send DMs

1. Introduction (The Problem and The Background)

An existing bicycle rental company in Manhattan, NYC is struggling tourists are now using Citi Bike, a privately owned public bicycle sharing system serving New York City. In Manhattan, you can travel by train, car, bus, but you often face to traffic, delayed trains. Since Manhattan is a small island, you can travel by bike.

The client company used to run a bike rental company, but they decided to a delivery company for local people. Luckily, they have enough employees to do the new business. However, the company is new in the delivery business, they need to run some ads.

With a limited budget, they wanted to find out which neighbors (Zipcode) are the best to send DMs. Their possible new customers are food-related businesses like restaurants and coffee shops, but the client doesn't want to deliver alcohol because it is time-consuming to check customers' ID.

2. Data

To find out the good area to send DMs, this project will use the following data.

1) Zip code Data

To send DMs, the data with zip code is easy to work with for the client. In this project, I used zip code data.

Data source: https://www.zipdatamaps.com/zipcodes-new-york-ny

2)Latitude and Longitude Data

To work with Foursquare, this project needs latitude and longitude data. By using this data, we can plot each zip code on map.

Data source: https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/download/?format=csv&refine.state=NY&timezone=America/New_York&lang=en&use_labels_for_header=true&csv_separator=%3B

3)Venue Data

To see which type of food-related business is common in each area, I used foursquare developer API to get venue data.

4)Population Data

In Manhattan, each area is different. Some areas are mainly for business buildings. Other areas are mainly for residential buildings. The population data will be one of the good data to deeply understand each area’s characteristics.

Data source: https://www.zipdatamaps.com/zipcodes-new-york-ny

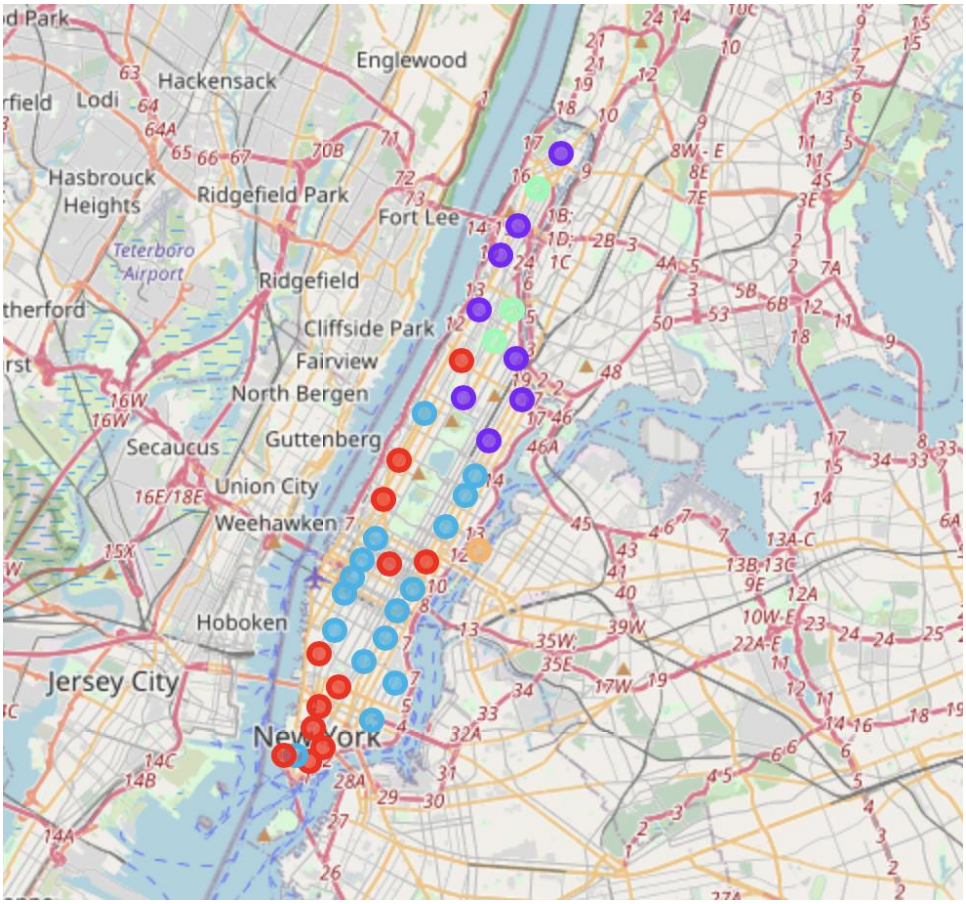
3. Methodology

There are In Manhattan, there are 47 zip codes. The goal of this project is to find out best neighbors (zip codes) to send DMs for the new delivery company. Therefore, it is not efficient to count every single food-related business in each zip code. Instead, I explored what types of food-related business are common in each zip code and did grouping by using K-mean clustering. To do so, if bodega, which is not the client's ideal customer, is the most common in a certain group, the client can decide not not send DMs to the group.

After getting venue data from Foursquare, I found out that some of the zip code areas don't have food-related businesses and 40 out of 47 zip code is the possible areas to send DMs.

For clustering, I set k = 5 and the result shows the characteristics for each group.

4. Results



Cluster 0 (Red): 12 zip codes, the number of populations is 33,7937

	ZIP Code	Population	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
3	10005	7135	Salad Place	American Restaurant	Food Truck	Pizza Place	Mexican Restaurant
5	10007	6988	Sandwich Place	Italian Restaurant	Deli / Bodega	Food Truck	Sushi Restaurant
9	10012	24090	Italian Restaurant	Café	American Restaurant	Sushi Restaurant	French Restaurant
10	10013	27700	Italian Restaurant	American Restaurant	French Restaurant	Café	Deli / Bodega
11	10014	31959	Italian Restaurant	American Restaurant	New American Restaurant	Café	French Restaurant
16	10020	0	Steakhouse	Italian Restaurant	American Restaurant	French Restaurant	Sandwich Place
18	10022	31924	Italian Restaurant	French Restaurant	American Restaurant	Bakery	Steakhouse
19	10023	60998	Italian Restaurant	Café	Food Truck	Pizza Place	American Restaurant
20	10024	59283	Italian Restaurant	Café	Indian Restaurant	American Restaurant	Mediterranean Restaurant
23	10027	59707	Seafood Restaurant	Italian Restaurant	Chinese Restaurant	Deli / Bodega	Café
34	10038	20300	Italian Restaurant	Sandwich Place	Café	Donut Shop	Pizza Place
39	10280	7853	Food Truck	Pizza Place	Sandwich Place	Chinese Restaurant	American Restaurant

The cluster 0, it is the mixture of residential areas and business areas (less population). However, if you see common venues, Italian restaurant and American restaurant are common. It is the good area to send DMs.

Cluster 1 (Purple): 8 zip codes, the number of populations is 36,7994

	ZIP Code	Population	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
22	10026	34003	Deli / Bodega	African Restaurant	Caribbean Restaurant	Café	Seafood Restaurant
25	10029	76003	Deli / Bodega	Mexican Restaurant	Pizza Place	Thai Restaurant	Café
27	10031	56438	Deli / Bodega	Pizza Place	Mexican Restaurant	Chinese Restaurant	Caribbean Restaurant
28	10032	57331	Pizza Place	Bakery	Mexican Restaurant	Chinese Restaurant	Latin American Restaurant
29	10033	53926	Deli / Bodega	Bakery	Spanish Restaurant	Pizza Place	Chinese Restaurant
30	10034	38908	Restaurant	Deli / Bodega	Pizza Place	Mexican Restaurant	Café
31	10035	33969	Deli / Bodega	Pizza Place	Chinese Restaurant	American Restaurant	Fast Food Restaurant
33	10037	17416	Café	Latin American Restaurant	Southern / Soul Food Restaurant	Cafeteria	Burger Joint

In Cluster 1 areas, Deli/Bodega is the most common. Even though they are residential areas, this cluster is not ideal for the client.

Cluster 2 (Blue): 16 zip codes, the number of populations is 69,3105

	ZIP Code	Population	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	10001	21102	Salad Place	Café	Burger Joint	Deli / Bodega	American Restaurant
1	10002	81410	Mexican Restaurant	Asian Restaurant	Pizza Place	Bakery	Ramen Restaurant
2	10003	56024	Japanese Restaurant	Mediterranean Restaurant	Pizza Place	Sushi Restaurant	Chinese Restaurant
4	10006	3011	Pizza Place	Mexican Restaurant	American Restaurant	Steakhouse	Café
6	10009	61347	Mexican Restaurant	Pizza Place	Italian Restaurant	Korean Restaurant	Vegetarian / Vegan Restaurant
7	10010	31834	Indian Restaurant	Italian Restaurant	Bagel Shop	Pizza Place	American Restaurant
8	10011	50984	American Restaurant	Italian Restaurant	Sushi Restaurant	Bakery	Mexican Restaurant
12	10016	54183	Italian Restaurant	Mexican Restaurant	Deli / Bodega	Sandwich Place	Greek Restaurant
13	10017	16575	Japanese Restaurant	Italian Restaurant	Seafood Restaurant	Sushi Restaurant	Deli / Bodega
14	10018	5229	American Restaurant	Italian Restaurant	Pizza Place	Thai Restaurant	Sandwich Place
15	10019	42870	Italian Restaurant	Mexican Restaurant	Pizza Place	Sandwich Place	Thai Restaurant
17	10021	43631	Italian Restaurant	Pizza Place	Burger Joint	Café	Sushi Restaurant
21	10025	94600	Mexican Restaurant	Pizza Place	Indian Restaurant	Thai Restaurant	Latin American Restaurant
24	10028	45141	Italian Restaurant	Sushi Restaurant	Mexican Restaurant	Bakery	Thai Restaurant
32	10036	24711	Italian Restaurant	Mexican Restaurant	Pizza Place	Sandwich Place	Thai Restaurant
38	10128	60453	Pizza Place	Italian Restaurant	Café	Sushi Restaurant	Deli / Bodega

Cluster 2 is the biggest group. Like Cluster 0, residential and business areas are fell into this group. In this cluster, bodega is not common like cluster 1, and Mexican, Japanese, American, Pizza restaurants are common. This is the ideal group for the client.

Cluster 3 (Green): 3 zip code areas, the number of populations is 93,413

	ZIP Code	Population	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
26	10030	26999	Deli / Bodega	Southern / Soul Food Restaurant	Pizza Place	Chinese Restaurant	Sandwich Place
35	10039	24527	Deli / Bodega	Chinese Restaurant	Caribbean Restaurant	Indian Restaurant	Pizza Place
36	10040	41905	Deli / Bodega	Pizza Place	Chinese Restaurant	Seafood Restaurant	Donut Shop

Deli/Bodega is the most common in this group. Other than Deli/Bodega, other common venues are diverse in terms of food compared to other clusters.

Cluster 4 (Orange): 1 zip code, the number of 11,661

	ZIP Code	Population	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
37	10044	11661	Deli / Bodega	Kosher Restaurant	Restaurant	Sandwich Place	Pizza Place

Cluster 4 is the smallest group. This zip code is for Roosevelt Island. Deli/Bodega is the most common venue and Kosher Restaurant is the second.

5. Discussion

Each cluster has different characteristics, and it is the good presentation for the client to decide where to send DMs. However, due to the free account limitation of Foursquare, I can use up to 100 venue data for each zip code. Some of the zip code areas have more than 100 venues. I assume that 100 venues can represent the area's characteristics well, but more data might be necessary if the client needs.

6. Conclusion

Based on the results, I would recommend to send DMs to cluster 0 and 2 areas. Those 2 areas cover about 60% of the populations in Manhattan. As the map shows, the upper side of Manhattan is not attractive for the client because Deli/Bodega is the most common.