The battle of Neighborhoods

1. Introduction & Business Problem

1.1. Background

The third-party delivery industry has been growing exponentially in recent years due to technology adoption. Nowadays, this service industry is based on mobile apps in which a client orders a product from the business listed and a delivery man picks up the items from a store and transports it to the client's address while receiving part of the fee charged for the service.

1.2. Problem

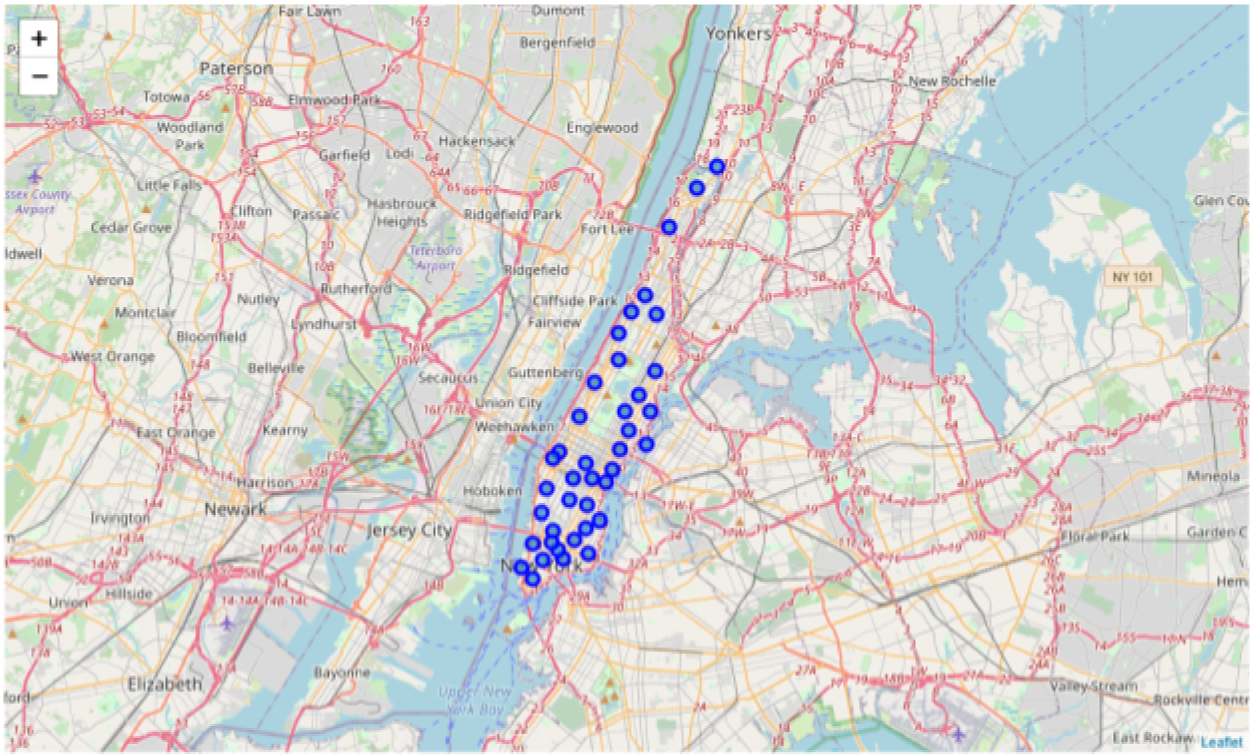
Orders are mostly accepted by the delivery man closest to the business selected by the client, because this means he has to invest the least amount of resources to get to the item, hence, he keeps a bigger profit from the fee. In order to have a substantial income from this activity, the delivery man has to carefully select the area in which he should be located to accept more orders without having to spend too many resources to get to the selected items. This project will propose which areas and times could be the most profitable for a delivery man based on the venue density and top visit hours, because it's inferred that these are also the top hours for take out orders, all this for venues in Manhattan.

2. Data

Only Manhattan will be analyzed in this project. We will analyze its neighborhoods' venue density by using the coordinates of each neighborhood, available here https://cocl.us/new_york_dataset, and use them as input to request the venues information available through Foursquare API. The data will be cleansed and wrangled in order to provide a dataframe that will be used for segmentation and clustering of the neighborhoods. The clusters will show which neighborhoods have more food business which will indicate a higher opportunity for the delivery man.

3. Methodology

We downloaded, cleaned and wrangled the neighborhoods data which is mainly the location by coordinates. Then we visualized the dataframe into a map, to review what we were dealing with.



Map of the selected New York area with its neighborhoods.

Next we used Foursquare API to search for nearby venues and we merged the requested information with the neighborhoods dataframe we already had.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop

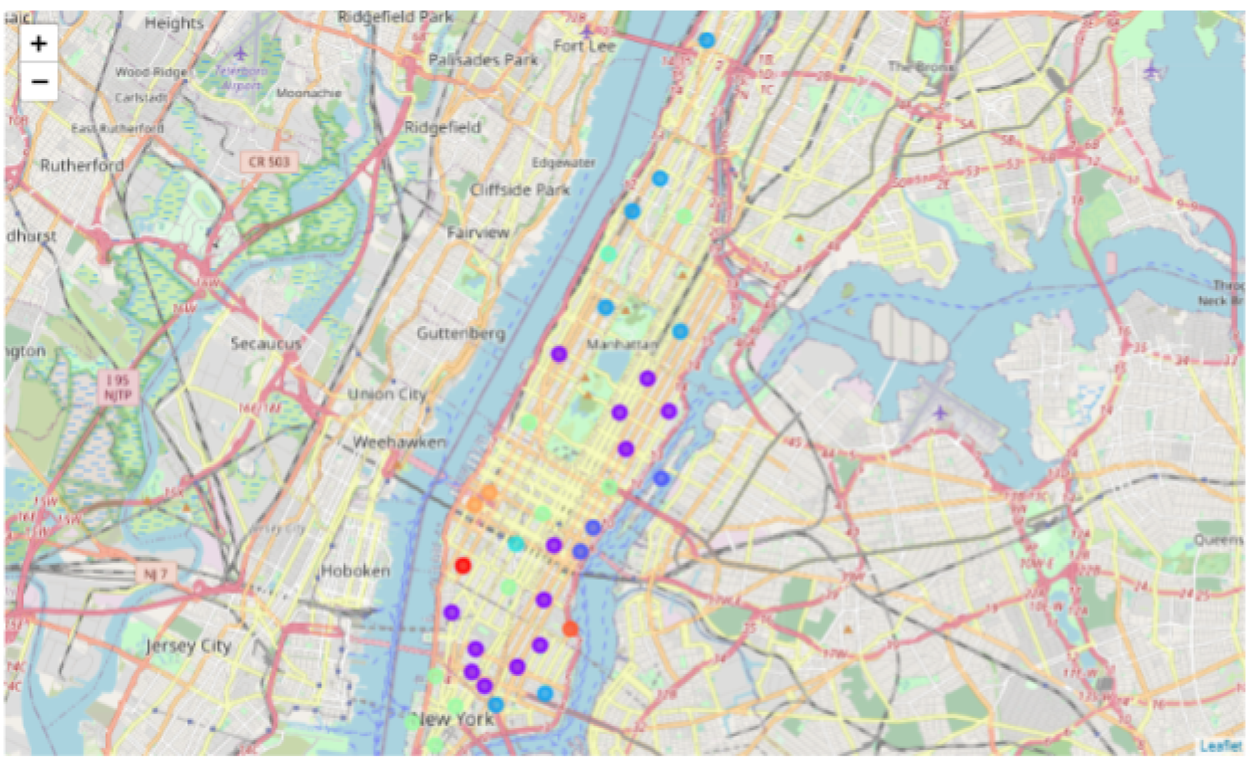
Neighborhood and venues dataframe head.

We then reviewed the frequency of the venues category in each neighborhood to determine clusters with venues similarity. We transferred the results into a new dataframe which included the top 10 venues category per 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	Park	Hotel	Gym	Boat or Ferry	Memorial Site	Italian Restaurant	Gourmet Shop	Burger Joint	Food Court	Shopping Mall
1	Carnegie Hill	Coffee Shop	Café	Yoga Studio	Pizza Place	Bar	Bookstore	Gym	Gym / Fitness Center	Japanese Restaurant	Italian Restaurant
2	Central Harlem	Chinese Restaurant	Gym / Fitness Center	African Restaurant	American Restaurant	Bar	Cosmetics Shop	French Restaurant	Seafood Restaurant	Café	Market
3	Chelsea	Art Gallery	Coffee Shop	Italian Restaurant	Seafood Restaurant	Ice Cream Shop	Cupcake Shop	Café	Bakery	Market	Juice Bar
4	Chinatown	Bakery	Chinese Restaurant	Cocktail Bar	Optical Shop	Salon / Barbershop	American Restaurant	Spa	Bar	Coffee Shop	Pizza Place

Venues frequency per neighborhood dataframe head.

Then we proceeded to run the k-means clustering tool, and we visualized the obtained information over a map.



Map of neighborhood clusters based on venue frequency similarity.

We also reviewed each cluster and picked the one with more restaurant frequency to focus on those venues for the next step of the analysis.

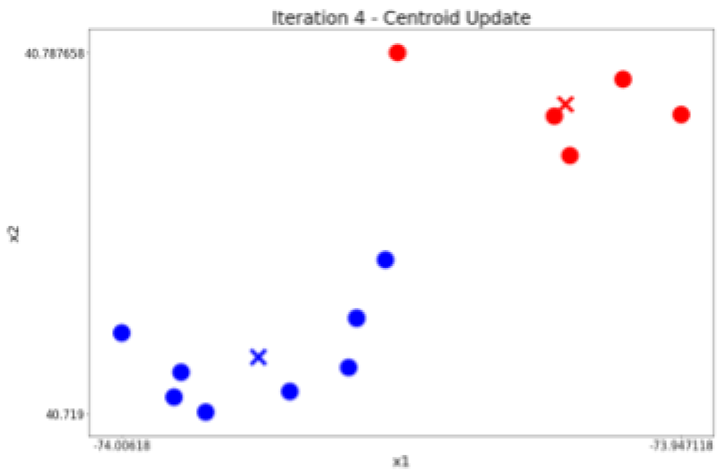
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
8	Upper East Side	Italian Restaurant	Bakery	Exhibit	Gym / Fitness Center	Spa	American Restaurant	Juice Bar	Pizza Place	Yoga Studio	Wine Shop
9	Yorkville	Coffee Shop	Italian Restaurant	Gym	Sushi Restaurant	Bar	Deli / Bodega	Wine Shop	Pizza Place	Mexican Restaurant	Diner
10	Lenox Hill	Italian Restaurant	Coffee Shop	Pizza Place	Sushi Restaurant	Café	Cocktail Bar	Burger Joint	Gym	Gym / Fitness Center	Bakery
12	Upper West Side	Italian Restaurant	Wine Bar	Coffee Shop	Bakery	Bar	Middle Eastern Restaurant	Bagel Shop	Mediterranean Restaurant	Indian Restaurant	Ice Cream Shop
16	Murray Hill	Sandwich Place	Chinese Restaurant	Gym / Fitness Center	Pizza Place	Sushi Restaurant	Hotel	Bar	Liquor Store	Burger Joint	Japanese Restaurant
18	Greenwich Village	Italian Restaurant	Coffee Shop	Café	Sushi Restaurant	Wine Bar	Dessert Shop	Chinese Restaurant	Gym	Jazz Club	Gourmet Shop
19	East Village	Pizza Place	Coffee Shop	Juice Bar	Cocktail Bar	Vietnamese Restaurant	Bar	Japanese Restaurant	Filipino Restaurant	Mexican Restaurant	Bagel Shop
22	Little Italy	Italian Restaurant	Chinese Restaurant	Spa	Pizza Place	Mediterranean Restaurant	Bakery	Thai Restaurant	Café	Bubble Tea Shop	Cosmetics Shop
23	Soho	Italian Restaurant	Coffee Shop	Mediterranean Restaurant	Dessert Shop	Art Gallery	Gym	French Restaurant	Café	Clothing Store	Jewelry Store
24	West Village	Wine Bar	Italian Restaurant	New American Restaurant	Park	Coffee Shop	American Restaurant	Seafood Restaurant	Cycle Studio	Steakhouse	Jazz Club
27	Gramercy	Italian Restaurant	Coffee Shop	Playground	Pizza Place	Bar	Grocery Store	Cocktail Bar	Spa	Salon / Barbershop	Taco Place
30	Carnegie Hill	Coffee Shop	Café	Yoga Studio	Pizza Place	Bar	Bookstore	Gym	Gym / Fitness Center	Japanese Restaurant	Italian Restaurant
31	Noho	Italian Restaurant	Coffee Shop	Sandwich Place	Japanese Restaurant	Grocery Store	Wine Shop	Wine Bar	Pizza Place	Seafood Restaurant	Asian Restaurant

Selected cluster dataframe.

Our next step was to propose the perfect spots for our delivery man to wait for new orders opportunities based on what venues were trending in the area. So we ran a function from Foursquare API to review which venues of the selected cluster were trending, unfortunately, due to COVID19 emergency venues were closed and no data was available, so we changed our approach and we now propose spots that are the center or new clusters formed by the available venues in this cluster.

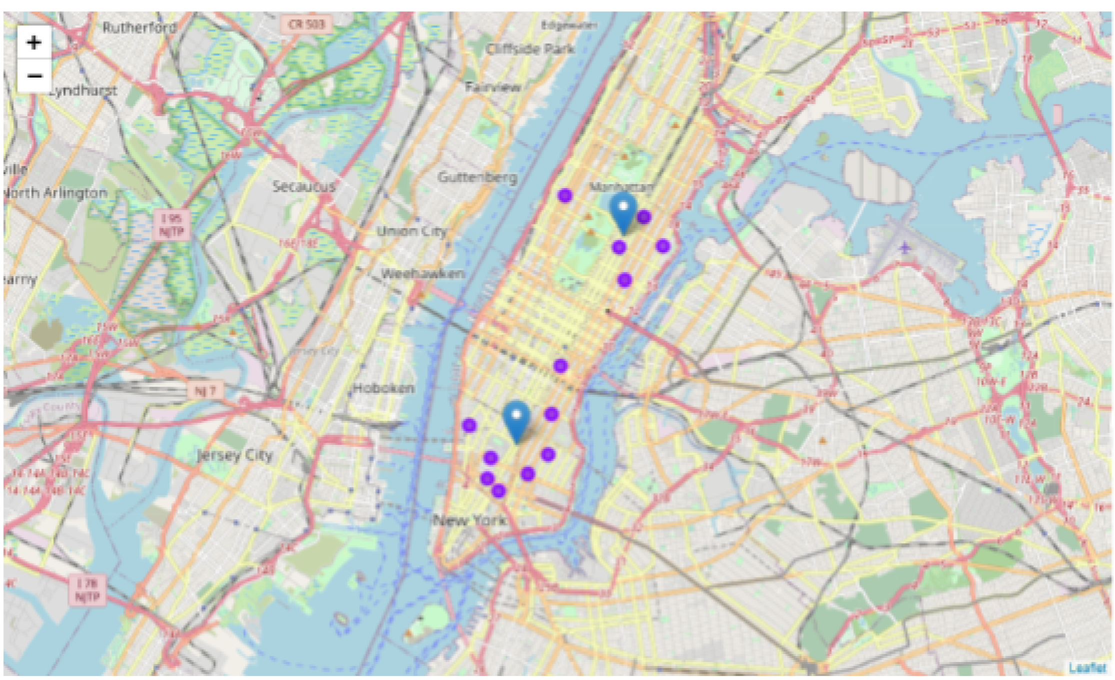
4. Results

We iterate k-means until the centers of the new cluster wouldn't change with the following result.



Clusters centers by K-means iterations.

Then we passed these proposed centers into the neighborhood map we deployed earlier with icons showing the standby spots for our delivery man.



Map with proposed standby spots for our delivery man.

Using geolocator we found the addresses for the proposed standby spots as:

- FedEx, Astor Place, NoHo Historic District, NoHo, Manhattan Community Board 2, New York, Manhattan, New York County, New York, 10003, United States of America
- 69, East 82nd Street, Yorkville, New York, Manhattan Community Board 8, Manhattan, New York County, New York, 10028, United States of America

The analysis resulted in 2 spots for the delivery man to wait for new orders. These 2 spots are placed within the closest distance of 2 neighborhood clusters that have the highest density of restaurants according to the data obtained from Foursquare. However, there's a lack of confidence in the Foursquare data, it looks outdated, and it's dependent on its users feedback, if Foursquare is not being used anymore the results of this project could be outperformed by any other proposal based on more accurate data.

5. Conclusion

This project reviewed the problem a third party delivery man faces when selecting the area with more orders opportunities. Based on data extracted from New York area, focused on Manhattan, we determined a cluster of neighborhoods with the most restaurants density according to Foursquare API, then we tried to propose the stand by spots based on those trending venues, unfortunately, due to Covid19 restaurants are temporarily closed and no trending data could be obtained, hence, an alternative approach is defined, standby spots were determined by K-means iterations of the selected cluster's venues. The result is 2 Standby spots that represent the perfect center of 2 restaurant clusters.