# The battle of Neighborhoods

Best place to settle a new Food Truck in NY during Covid19 pandemy

Introduction

- During the Covid pandemy, lockdown has heavily impacted how and where jobs are being performed. Facilities like restaurant and bars have been closed and only Drive-in or Food-truck are still allowed to pursue their activity.

- In this project, we'll take the role a new entrepreneurs who is looking after opening a Foodtruck in city of New York. We'll focus on how to choose the best place to settle and start this new activity, by looking after potential competitors.

Borough	Neighborhood	Latitude	Longitude
Bronx	52	52	52
Brooklyn	70	70	70
Manhattan	40	40	40
Queens	81	81	81
Staten Island	63	63	63
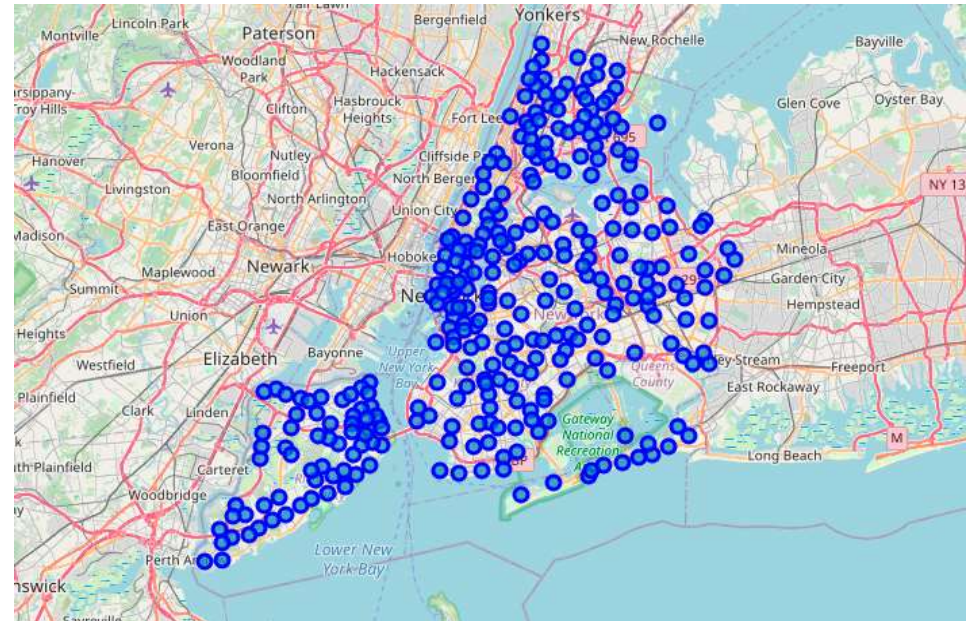
There are 5 Boroughs in NY and 306 Neighborhoods : how to choose the best place to settle a new Food Truck activity ?

Datasets

- Assuming that the lockdown is still in place, the main competitors are Drive-in and other Take away store in the same area. Using Foursquare API we'll be able to locate them by Neighborhood.

- Using the Category type, we'll classify them as standard Restaurant or Fast Food.
  - Standard Restaurant will be closed but still provide a good idea about the market.
  - Fast Food will be our main competitors, sot let's try to find a place where there are only few of them.

- Then using a JSON file on New York Neighborhoods, we'll be able to plot them on a map

	Unnamed: 0	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	FoodType	Customer	Borough
0	0	Wakefield	40.894705	-73.847201	Subway	40.890468	-73.849152	Fast Food Restaurant	Fast Food	NotCustomer	Bronx
1	1	Co-op City	40.874294	-73.829939	Arby's	40.870411	-73.828606	Fast Food Restaurant	Fast Food	NotCustomer	Bronx
2	2	Co-op City	40.874294	-73.829939	Kennedy's	40.876807	-73.829627	Fast Food Restaurant	Fast Food	NotCustomer	Bronx
3	3	Eastchester	40.887556	-73.827806	McDonald's	40.885636	-73.826840	Fast Food Restaurant	Fast Food	NotCustomer	Bronx

```
File for competitor_Bronx_tocsv.csv has (23, 11) lines.
File for competitor_Brooklyn_tocsv.csv has (34, 11) lines.
File for competitor_Manhattan_tocsv.csv has (89, 11) lines.
File for competitor_Queens_tocsv.csv has (33, 11) lines.
File for competitor_Staten Island_tocsv.csv has (15, 11) lines.

(194, 11)
```

Using Foursquare API to find out the implementation of competitors by Borough : now time to work on the consolidated result.

Methodology

- As we are limited in the number of free call to Foursquare API, we'll use it 5 times (one per Borough) and store the result in a CSV file.

- Those 5 new entries will be merged and categorized according to our focus of interest : whether they are Regular Restaurant or Fast Food.

- Using One-hot encoding method those new category will help us to implement KNN algorithm to cluster the Neighborhoods.

- Finally, we will be able to get a great vizualizations of those cluster on a map, and hopefully make the best choice for our FoodTruck implementation.

	Neighborhood	C_Fast Food	C_Restaurant	C_Sandwich
0	Wakefield	1	0	0
1	Co-op City	1	0	0
2	Co-op City	1	0	0
3	Eastchester	1	0	0
4	Eastchester	1	0	0

Focus on the Main Category :
Restaurant, Sandwich, Fast Food
and apply One-hotencoding
Method to come up with new fields.

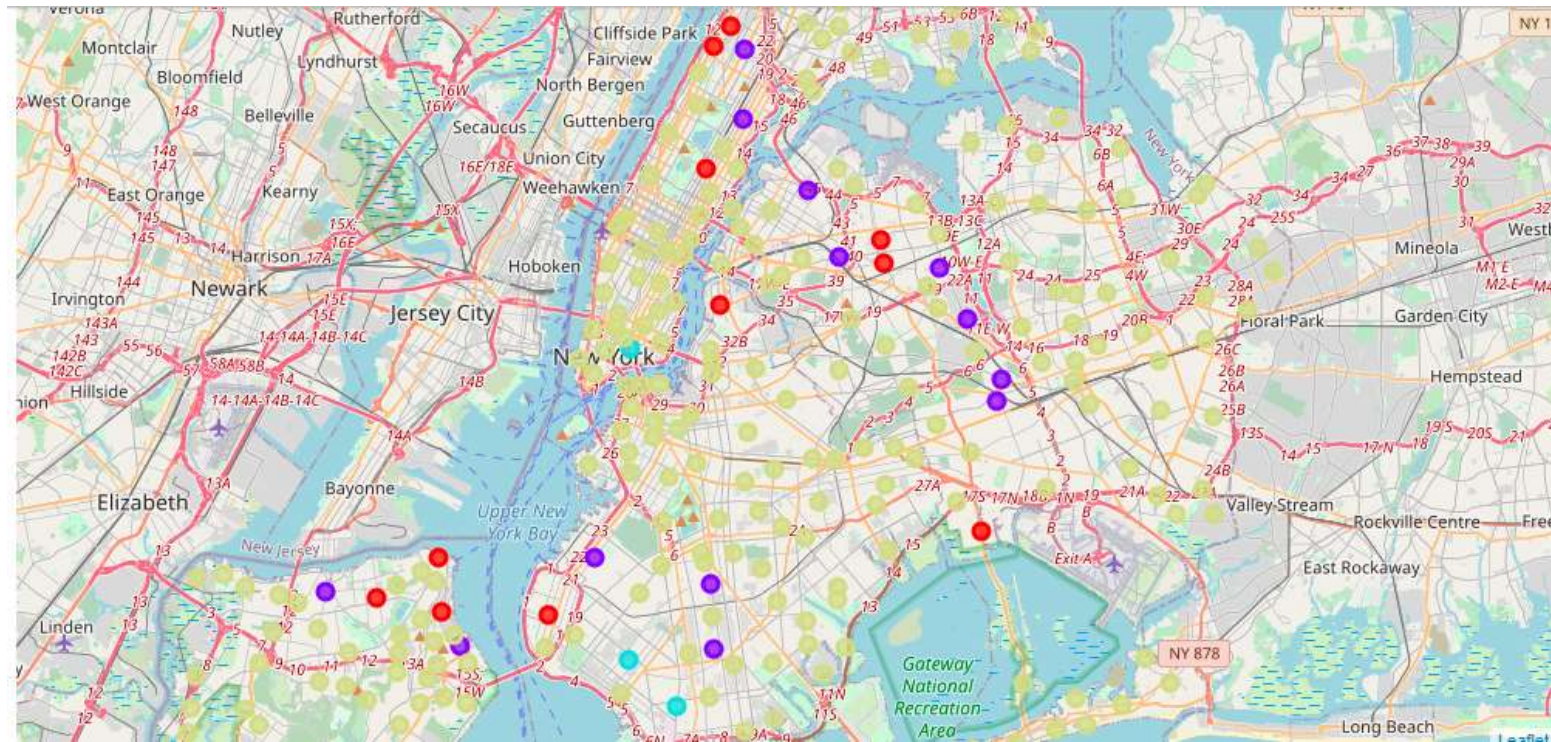|	Neighborhood	C_Fast Food	C_Restaurant	C_Sandwich
0	Astoria	0.833333	0.166667	0.000000
1	Bay Ridge	0.800000	0.200000	0.000000
2	Bensonhurst	0.125000	0.875000	0.000000
3	Brighton Beach	0.750000	0.250000	0.000000
4	Central Harlem	0.875000	0.125000	0.000000
5	Chinatown	0.285714	0.678571	0.035714

Find out Occurency for
each one of them by
Borough.

Design Cluster using KNN algorithm.

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Cluster
0	Bronx	Wakefield	40.894705	-73.847201	1.0	C_Fast Food	C_Sandwich	C_Restaurant	1
1	Bronx	Co-op City	40.874294	-73.829939	1.0	C_Fast Food	C_Sandwich	C_Restaurant	1
2	Bronx	Eastchester	40.887556	-73.827806	1.0	C_Fast Food	C_Sandwich	C_Restaurant	1
3	Bronx	Fieldston	40.895437	-73.905643	1.0	C_Fast Food	C_Sandwich	C_Restaurant	1
4	Bronx	Riverdale	40.890834	-73.912585	NaN	NaN	NaN	NaN	3

Let's cluster the Neighborhoods, focusing on how Restaurant and
Fast Food are implemented.

There are 4 Clusters identity on NY for Food implementation.

Let's cluster the Neighborhoods, focusing on how Restaurant and Fast Food are implemented.

Results

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Cluster
Woodlawn	C_Restaurant	C_Sandwich	C_Fast Food	2
Bensonhurst	C_Restaurant	C_Fast Food	C_Sandwich	2
Gravesend	C_Restaurant	C_Sandwich	C_Fast Food	2
Sheepshead Bay	C_Sandwich	C_Restaurant	C_Fast Food	2
Chinatown	C_Restaurant	C_Fast Food	C_Sandwich	2

- Since Restaurant are closed because of the Lockdown, the best place to settle a Food Truck activity looks like the Cluster 2.

- From a foreign perspective, I'll probably skip Bronx and Chinatown, which let us with Brooklyn location to settle : either in Gravesend or Sheepshead Bay.

Discussion

Foursquare API provides a limited number of free call and number of result. Results provided during this exercises might be biaised by the number of line found in the results set.

The model could be improved by adding other features to get a better understanding of the neighborhoods :

- Customer : find out office and building where workers are when working on-site instead of remotely.
- Safety : looking after criminalty could also have a huge impact on the final decision.

The analysis was made for a standard foodtruck. We could have focus on a specific type of food : american, asian, european... and find different results.

Conclusion

This Capstone project was a great exercise to apply the different aspects of Data Science discussed during the Training.

From a functional perspective : to actually step-in the full process from a Business problem to the analysis phase.

From a technical perspective : using some new librairies and understand how to include them in the resolution phase.

From global perspective : to understand the power as well as the limitation of Datascience. **The findings are as good as the input which you use to fuel your model.**