Introduction

New York City is a collection of many neighbourhoods scattered among the city's five boroughs—Manhattan, Brooklyn, the Bronx, Queens, and Staten Island—each exhibiting its own lifestyle. It is the most populous and the most international city in the country. Furthermore, New York is the most ethnically diverse, religiously varied, commercially driven, famously congested, and, in the eyes of many, the most attractive urban centre in the country.

Hence, opening any new business should be carefully studied and a strategic plan should be developped in order to guarantee the long term succes. Indeed, this requires a careful understanding of the diffrent neighbourhoods of the city, the ethnicity of the people and most important their habits and tastes.

Business problem

We can say that healthcare is one of the most valuable industries in the World. This is due to the spread of the fast food culture which leads to many diseases and health problems. Hence, many people are trying to find methods in order to avoid falling into these problems. We can say that the preventive cure exists in two major categories: gym and healthy food.

In terms of healthy food, vegetrian restaurants come in the top of the list. Starting a vegterian restaurant can be a great business opportunity, but we need to distinguish ourselves from others to enjoy long-term success. We think that by exploring the neighbourhoods and their different venues and by emphesizing on the locations where gym centers are popular will lead us to the right place of our presumed vegetarian restaurants. This is because people who frequently visit gym are likely interested in eating healthy food.

We will content ourselves to provide this study for the city of New York and in particular to Manhattan neighbourhood since our client is interested in this area.

Data

We will rely on the publicly available data of New York city available online: https://geo.nyu.edu/catalog/nyu_2451_34572

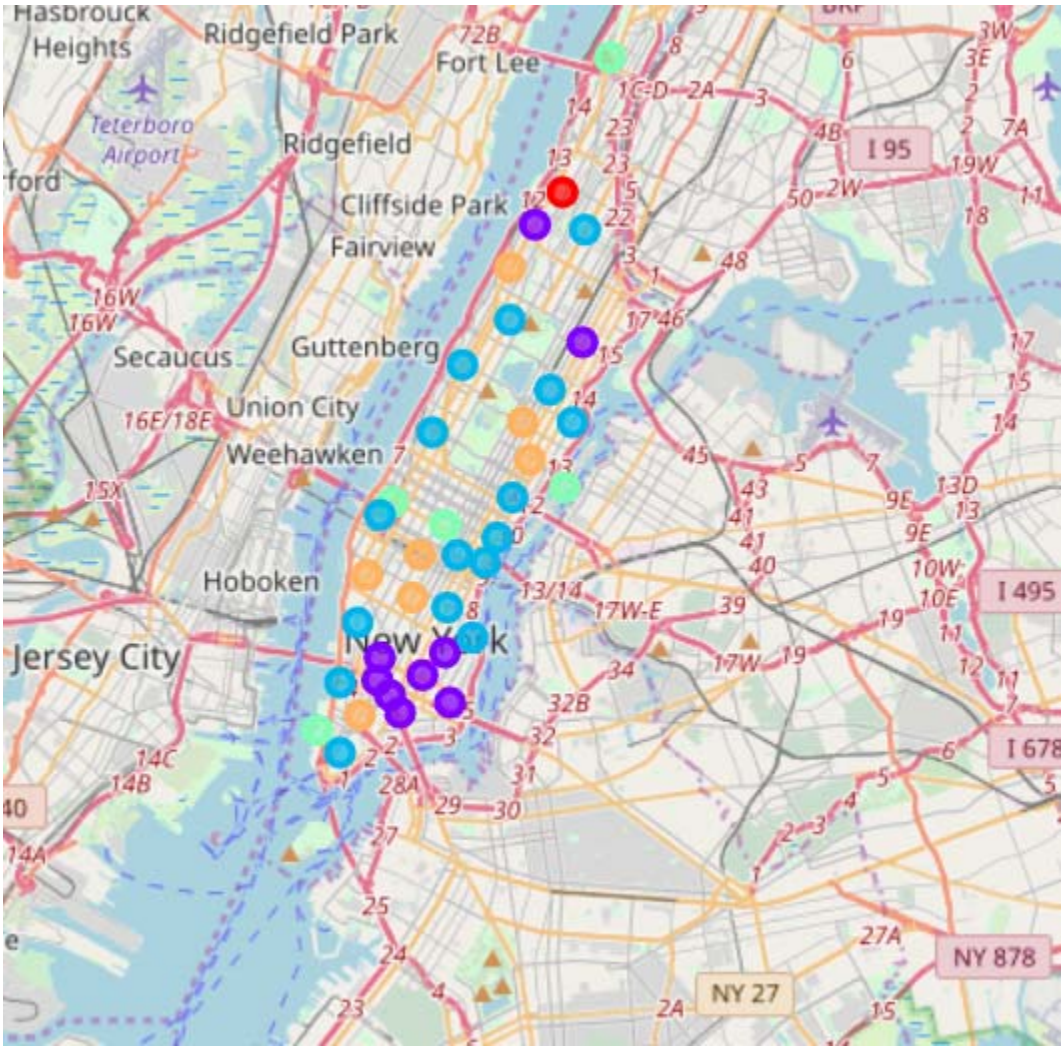
This dataset contains 5 boroughs and 306 neighborhoods. Now since our methodology is to search for gym/fitness center venues and try to plug our vegeterian restaurant nearby, we will use this website:https://developer.foursquare.com/docs/resources/categories in order to find gym/fitness center category id. We found that this id = 4bf58dd8d48988d175941735

Methodology

Our methodology consists of loading and cleaning the dataset, then performing k-means statisical analysis in two stages. First, we consider venues with category = gym/fitness. Second, we do clustering regardless of the category type. Then we decide if a neighborhood needs a vegetarian restaurant, by inspecting if it appears in a cluster with high concentation in gym/fitness venues (i.e. by looking for clustering in Stage 1) and low concentration in restaurants, especially vegetarian restaurants (i.e. by looking for clustering in Stage 2).

Results

After following the above methodology we obtained clustering in the first stage as shown below:



Now let’s consider some of the resulting clusters. In particular let’s take cluster 5 :

Cluster 5

```
[34]: d.loc[manhattan_merged['Cluster Labels'] == 4, manhattan_merged.columns[[1] + list(range(5, manhattan_merged
```

t[34]:

	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
8	Upper East Side	Gym / Fitness Center	Gym	Yoga Studio	Pilates Studio	Cycle Studio	Doctor's Office	Women's Store	Martial Arts Dojo	Medical Center
10	Lenox Hill	Gym / Fitness Center	Gym	Pilates Studio	Yoga Studio	Cycle Studio	Non-Profit	Doctor's Office	Spa	Dance Studio
17	Chelsea	Gym / Fitness Center	Yoga Studio	Gym	Martial Arts Dojo	Cycle Studio	Recreation Center	Climbing Gym	Pilates Studio	Gymnastics Gym
26	Morningside Heights	Gym / Fitness Center	Gym	Yoga Studio	Track	Gym Pool	Cycle Studio	Basketball Court	Boxing Gym	Bridge
32	Civic Center	Gym / Fitness Center	Gym	Yoga Studio	Boxing Gym	Martial Arts Dojo	Athletics & Sports	Pilates Studio	Gym Pool	Women's Store
33	Midtown South	Gym / Fitness Center	Gym	Yoga Studio	Martial Arts Dojo	Pilates Studio	Cycle Studio	Building	Boxing Gym	Physical Therapist

Clearly, this cluster contains plenty of gym/fitness centers.

Considering one of the neighborhoods in this cluster, for example ‘Chelsea’, and looking for it in the second stage of clustering, we got the following:

Cluster 3

```
In [51]: tan_merged_all['Cluster Labels'] == 2, manhattan_merged_all.columns[[1] + list(range(5, manhattan_merged_all
```

Out[51]:

	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
0	Marble Hill	Deli / Bodega	Salon / Barbershop	Residential Building (Apartment / Condo)	High School	Laundry Service	Church	Furniture / Home Store	Nail Salon	Office
1	Chinatown	Chinese Restaurant	Miscellaneous Shop	Bus Station	Bakery	Bridge	Noodle House	Park	Arts & Crafts Store	Bus Line
3	Inwood	Laundry Service	Deli / Bodega	Salon / Barbershop	Nail Salon	American Restaurant	Bank	Mexican Restaurant	Pharmacy	Bus Line
4	Hamilton Heights	Residential Building (Apartment /	Salon / Barbershop	Building	Deli / Bodega	Non-Profit	Laundry Service	Church	Bar	Bus Station
13	Lincoln Square	Opera House	High School	Theater	Art Gallery	Residential Building (Apartment / Condo)	Building	Performing Arts Venue	Library	Event Space	Office
14	Clinton	Building	Bus Station	Theater	Bus Line	Lounge	Residential Building (Apartment / Condo)	Restaurant	Gym / Fitness Center	Deli / Bodega	Cocktail Bar
17	Chelsea	High School	Building	Residential Building (Apartment / Condo)	Deli / Bodega	Office	Art Gallery	Pet Store	Pharmacy	Music Venue	Laundry Service
18	Greenwich Village	Residential Building (Apartment /	Salon / Barbershop	Art Gallery	Furniture / Home Store	Cocktail Bar	Jewelry Store	Kids Store	Tech Startup	Ice Cream Shop	Clothing Store

We observe that Chelsea falls into cluster 3 which lacks of restaurants offering healthy food in general.

Hence we can consider that Chelsea is a good candidate for opening a new Vegetarian restaurant.

Discussion

Clearly, the methodology that we followed in this project can be beneficial for stakeholders seeking to open new food business respecting healthcare conditions. By using a simple and efficient clustering mechanism (kmeans), we showed an example of promoting a neighborhood (e.g. Chelsea) to be to most probably successful candidate for the business goal.

Conclusion

Nowadays, machine learning is considered a core leveraging technology which can be used in every field where data exists and plays an important role. In this project, we applied a very simple unsupervised machine learning algorithm (i.e. Kmeans) in order to solve a business problem consisting of opening a new vegetarian restaurant in the most appropriate neighborhood of New York city/ Manhattan borough. We showed, using one use case that our methodology is successful in finding a perfect location for our planned business.