Before starting :

Most of the elements in this report are figments of my imagination.

There is no C.O.W.R. aka consortium of worldwide restaurateurs. Neither NY is the #1 spot for restaurants in the world.

These were just assumption to create this data science project and try to give a sense to it.

REPORT: BATTLE OF THE RESTAURANTS

1) INTRODUCTION

1.1 Background

The C.O.W.R (Consortium Of Worldwide Restaurateurs) is the most important restaurant consortium in the world.

They nominated NY as the next #1 world spot for restaurants, and are planning to open more of them as soon as possible.

1.2 Problem

But before that, they want to understand in which neigborhoods of NY is more convenient to open more restaurants, or in which neighborhoods there are too many restaurants in order to consider to relocate some of their restaurants to another one.

1.3 Interest

The consortium is mostly interested in American, Italian and Chinese restaurants among the NY area.

2) DATA ACQUISITION AND CLEANING

2.1 Data sources

Datas regarding the NY neighborhood were collected through the link provided in one of the earlier labs (https://cocl.us/new_york_dataset).

Datas regarding the restaurants in NY were collected through the FOURSQUARE API

2.2 Data cleaning and Feature selection

All the datas were combined into one single data frame. All neighborhoods with no italian, american or chinese restaurant were removed from the dataset.

In the end. a data frame containing 196 rows remained. Each row contained one neighborhood , and AT LEAST one restaurant among the three categories.

	Neighborhood	American Restaurant	Italian Restaurant	Chinese Restaurant
0	Allerton	0	0	2
1	Annadale	2	0	0
2	Arlington	1	0	0
3	Arrochar	0	2	0
4	Astoria	0	2	0
5	Astoria Heights	0	1	1
6	Auburndale	1	1	0
7	Bath Beach	0	2	2
8	Battery Park City	1	3	1
9	Bay Ridge	3	7	2

3) Methodology section
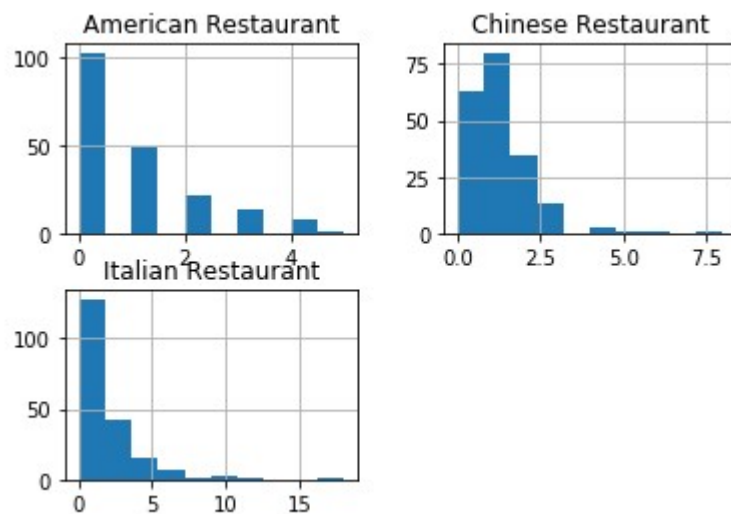
A preliminary idea about the data I collected suggested that there were few neighborhood with a large amount of restaurants, and many with just a few.

The American restaurants, indeed, seemed to be the smaller group among the three.

In order to gain some valuable data involving the restaurants and the neighborhoods in NY, I used the k-means clustering model included in the scikit-learn machine learning python libraries.

The dataset was divided into **FIVE** different clusters.

The resulting python series was merged with the original data frame (the one containing neighborhoods and restaurants), in order to proceed to create a map of NY showing at which cluster each neighborhood belonged.


In order to get to the map visualization, the latitude and longitude of every neighborhood was also added to the data frame.

The final data frame appeared to be like this:

	Cluster Labels	Neighborhood	American Restaurant	Italian Restaurant	Chinese Restaurant	Latitude	Longitude
0	4	Allerton	1	0	2	40.8658	-73.8593
1	0	Annadale	1	0	0	40.5381	-74.1785
2	0	Arrochar	0	2	0	40.5963	-74.0671
3	0	Astoria	0	2	0	40.7685	-73.9157
4	0	Astoria Heights	0	1	1	40.7703	-73.8947

For the map visualization, the 'Folium' library was used.
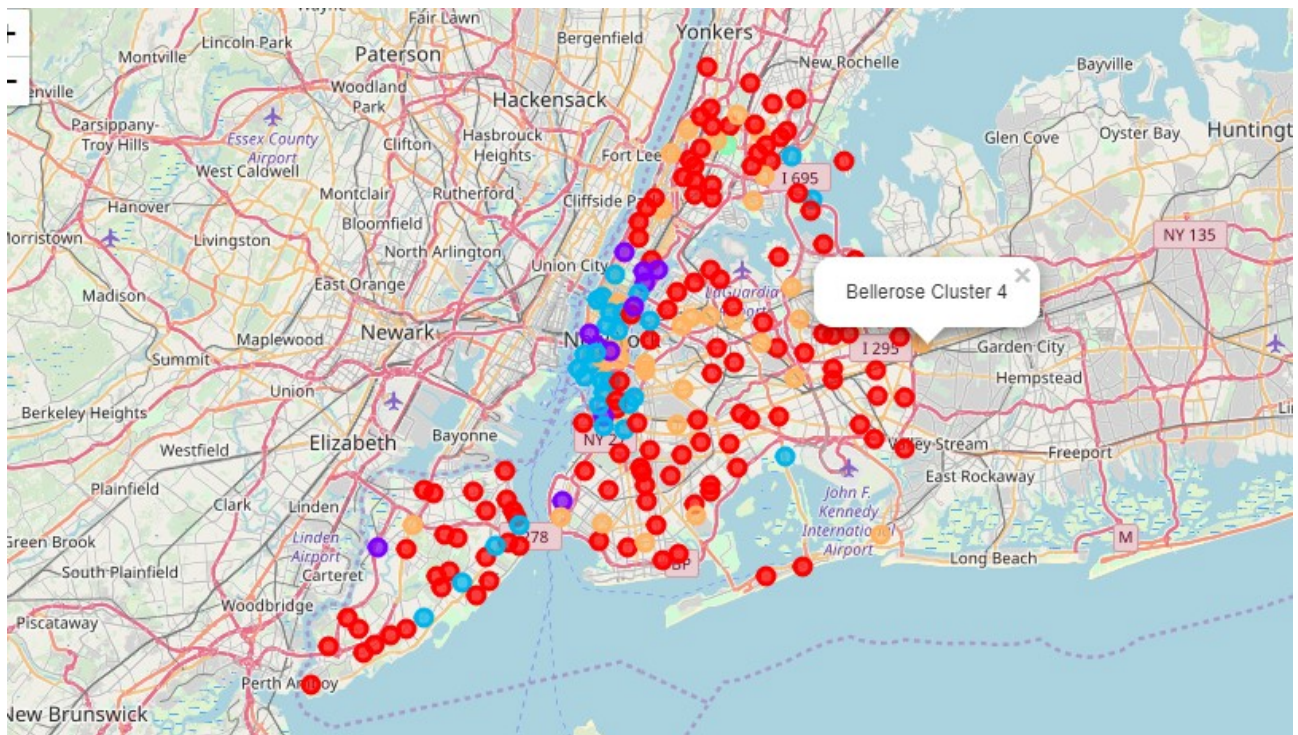
4) Results / Cluster analysis

Here follows an analysis of the four clusters:

Cluster zero

Cluster Zero has a total of 221 restaurants and 122 neighborhoods.

Most the majority of the restaurants are chinese and italian ones. Given the large number of neighborhoods included, this doesn't seem to appear a congested cluster.

```
American Restaurant      45
Italian Restaurant       83
Chinese Restaurant       93
total neighborhoods     122
total restaurants       221
```

Cluster one

Cluster one has 115 restaurants and 11 neighborhoods.
The majority of the restaurants are italians.

```
American Restaurant      21
Italian Restaurant       85
Chinese Restaurant        9
total neighborhoods      11
total restaurants       115
```

Cluster two

Cluster two has 191 restaurants and 30 neighborhoods.
The majority of the restaurant are italians with a moderate
Presence of american ones.

```
American Restaurant      62
Italian Restaurant      106
Chinese Restaurant       23
total neighborhoods      30
total restaurants       191
```

Cluster Three

Cluster three has 19 restaurants and only one neighborhood.
Every restaurant's italian except an american one.
This will be further examined in the discussion chapter.

```
American Restaurant        1
Italian Restaurant        18
Chinese Restaurant         0
total neighborhoods        1
total restaurants         19
```

Cluster Four

Cluster four has 168 restaurants and 33 neighborhoods.
Most of the restaurants are chinese ones.

```
American Restaurant       43
Italian Restaurant        30
Chinese Restaurant        95
total neighborhoods       33
total restaurants        168
```

4) Discussion

4.1 Cluster zero

Cluster zero is the least congested cluster of the four. With just 221 restaurants per 122 neighborhoods, it lends itself very well to be considered for new restaurants.

It contains 45 American restaurants, 83 Italian restaurants and 93 Chinese restaurants.

The consortium could consider to open any type of restaurant since the large gamma of neighborhoods, with a special consideration for the american ones, since they are the less numerous group.

4.2 Cluster one

Cluster one has 155 restaurants per 11 neighborhoods.
It contains 21 American restaurants, 85 italian restaurants and 9 chinese restaurants.

The big number of italian restaurants weights on the total, making the idea of new italian restaurants less desirable.
The consortium could consider to relocate some of them in some less congested areas, other than open new chinese or american ones due their low presence in the group.

4.2 Cluster two

Cluster has 191 restaurants per 30 neighborhoods.
62 are American restaurants, 106 italian restaurants and 23 chinese ones.

Like in the previous clusters, the italian restaurants presence is heavy and open new ones is not really appealing.
The American restaurant presence is a bit stronger than the previous cluster, and the chinese one tends to be lower.
The best choiche for the consortium is, as in the previous cluster, to relocate some italian ones and consider opening new chinese.

4.2 Cluster three

Cluster has 19 restaurants per only 1 neighborhood.
Apart from an american restaurant, they are all italians.

This deserves further investigations.


```
[25]: final_df.loc[final_df['Cluster Labels'] == 3]
```

	Cluster Labels	Neighborhood	American Restaurant	Italian Restaurant	Chinese Restaurant	Latitude	Longitude
17	3	Belmont	1	18	0	40.8573	-73.8885

Seems like the name of this neighborhood is 'Belmont'.

A rapid search on Google reveals this area to be the new 'Little Italy' of the Bronx.

(https://fifiandhop.com/2017/03/24/arthur-avenue-bronx-new-york-italy/)

This conciliates itself very well with the data results.

Nevertheless, since the non-presence of chinese restaurants, the consortium could consider to open one to see how it performs in this kind of environment.

4.2 Cluster four

Cluster four has 168 restaurants per 33 neighborhoods.
95 restaurants are chinese, 30 italian and 43 american ones.

Given the strong presence of chinese restaurants, the consortium could opt to open new italian or american ones and relocate some of the chinese's.

5) Discussion

An overview of the restaurants presence in the NY's neighborhood was given.

The clustering of the neighborhoods gave an idea and some recommendation for the consortium to evaluate the options among opening or relocating restaurants in order to maximize the profit and make NY the leader restaurants city in the world.