Introduction/ Business problem:

Our client, a well-established meat supplier in New York, NY area. His clients are hotels, restaurants and bars. Currently, the various meats are stored in some big central cold storages outside New York city and distributed in the different venues on a daily basis.

The main problem of this business plan is that the distribution of the meat product becomes time consuming, inefficient, and costly. Our client wants to reduce the costs of distribution of meat by building 5 smaller cold storage factories in New York city to serve locally his clients. This approach will reduce the time which is required to serve his clients, the fuel costs, and will also transform his business to a green chain.

To do so, our client asked from us to find the 5 best locations in New York city at which the cold storage must be built to create smaller distribution clusters. The storage locations must be at the centers of these clusters to minimize the relative distance from each of the venues.

Data:

The data required for this task are the locations of the hotels, bars and restaurants in New York. To gather the data, we 'll use the locations of all neighborhoods in New York gathered from uploaded IBM Cloud storage as part of one of the previous exercises. Based on these locations we'll gather the locations of all venues in these neighborhoods from Foursquare. We will filter the data to acquire the locations of the targeted venues. To inspect the data, we'll use the folium library to extract the map of New York city and visualize the locations of the venues on the map. A k-means algorithm will be applied on the location features to define the 5 clusters of venues. The locations of the cold storages will be defined as the centroids of the clusters. Again, to visualize the map of New York city, the 5 clusters and the locations of the cold storage we'll use the folium library.

Methodology

To define the location of each cold storage, we choose to use the k-means algorithm. By default, this algorithm minimizes the distance of each point in a cluster from the centroid of the cluster. As a result, the output of the k-mean algorithm is a set of clusters whose points are lying at the minimum distance from the determined centroids. As an input, we used the set of locations (lat, long) of the venues of interest. In this example the category of each venue is not required in the algorithm. All venues belonging to hotels, bars, restaurants are included in the list. The locations of hotels, restaurants, bars are shown as they were gathered from the Four Square database in Figure 1, 2, 3. A total of 8198 restaurants, 1619 bars and 186 hotels were found and used in the k-means algorithm.

```
) hotels.head(10)
```

9]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Latitude	Longitude	Category
151	Co-op City	40.874294	-73.829939	Residence Inn by Marriott New York The Bronx a...	40.849325	-73.842534	Hotel
1018	Baychester	40.866858	-73.835798	Residence Inn by Marriott New York The Bronx a...	40.849325	-73.842534	Hotel
1113	Pelham Parkway	40.857413	-73.854756	Residence Inn by Marriott New York The Bronx a...	40.849325	-73.842534	Hotel
2011	Melrose	40.819754	-73.909422	Opera House Hotel	40.815250	-73.916090	Hotel
2322	Longwood	40.815099	-73.895788	Opera House Hotel	40.815250	-73.916090	Hotel
2438	Hunts Point	40.809730	-73.883315	Opera House Hotel	40.815250	-73.916090	Hotel
2525	Morrisania	40.823592	-73.901506	Opera House Hotel	40.815250	-73.916090	Hotel
2676	Soundview	40.821012	-73.865746	Residence Inn by Marriott New York The Bronx a...	40.849325	-73.842534	Hotel
2913	Country Club	40.844246	-73.824099	Residence Inn by Marriott New York The Bronx a...	40.849325	-73.842534	Hotel
3020	Parkchester	40.837938	-73.856003	Residence Inn by Marriott New York The Bronx a...	40.849325	-73.842534	Hotel

Fig 1: Hotel locations in NY

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Latitude	Longitude	Category
1	Wakefield	40.894705	-73.847201	Ripe Kitchen & Bar	40.898152	-73.838875	Caribbean Restaurant
2	Wakefield	40.894705	-73.847201	Ali's Roti Shop	40.894036	-73.856935	Caribbean Restaurant
3	Wakefield	40.894705	-73.847201	Jackie's West Indian Bakery	40.889283	-73.843310	Caribbean Restaurant
6	Wakefield	40.894705	-73.847201	Bx Uptown Caribbean Taste	40.890178	-73.859224	Caribbean Restaurant
11	Wakefield	40.894705	-73.847201	Fish & Ting	40.885539	-73.829151	Caribbean Restaurant
15	Wakefield	40.894705	-73.847201	Roc A Tone Restaurant	40.882620	-73.863130	Seafood Restaurant
16	Wakefield	40.894705	-73.847201	Patrizia's Of Woodlawn	40.900638	-73.867724	Italian Restaurant
20	Wakefield	40.894705	-73.847201	Gold Star Jerk And Seafood Center	40.883433	-73.862813	Caribbean Restaurant
24	Wakefield	40.894705	-73.847201	Dyre Fish Market	40.889318	-73.831453	Seafood Restaurant
28	Wakefield	40.894705	-73.847201	Country Thyme Cuisine	40.876880	-73.837734	Caribbean Restaurant

Fig 2: Restaurant locations in NY

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Latitude	Longitude	Category
12	Wakefield	40.894705	-73.847201	Rory Dolan's	40.904888	-73.868792	Bar
52	Wakefield	40.894705	-73.847201	Dave & Buster's	40.892236	-73.821055	Sports Bar
76	Wakefield	40.894705	-73.847201	JP Clarke's Saloon	40.903455	-73.866695	Bar
158	Co-op City	40.874294	-73.829939	Dave & Buster's	40.892236	-73.821055	Sports Bar
216	Eastchester	40.887556	-73.827806	Dave & Buster's	40.892236	-73.821055	Sports Bar
283	Eastchester	40.887556	-73.827806	Rory Dolan's	40.904888	-73.868792	Bar
304	Fieldston	40.895437	-73.905643	An Beal Bocht Cafe	40.887303	-73.905025	Bar
309	Fieldston	40.895437	-73.905643	Bronx Alehouse	40.884689	-73.899707	Beer Bar
319	Fieldston	40.895437	-73.905643	Tilila	40.883872	-73.898209	Bar
370	Fieldston	40.895437	-73.905643	Rory Dolan's	40.904888	-73.868792	Bar

Fig 3: Bar locations in NY

Having extracted the clusters, descriptive statistics were used to determine the radius of each cluster and therefore the area in which the venues of interest will be served by the specific cold storages. Moreover, in order to examine the result, the density of venues in each area was calculated. Mean values were used to determine the exact location of each cold storage.

Results:

In Figure 5 we present the map of New York city. The locations of the venues of interest are marked on the map with different colors according to their category.
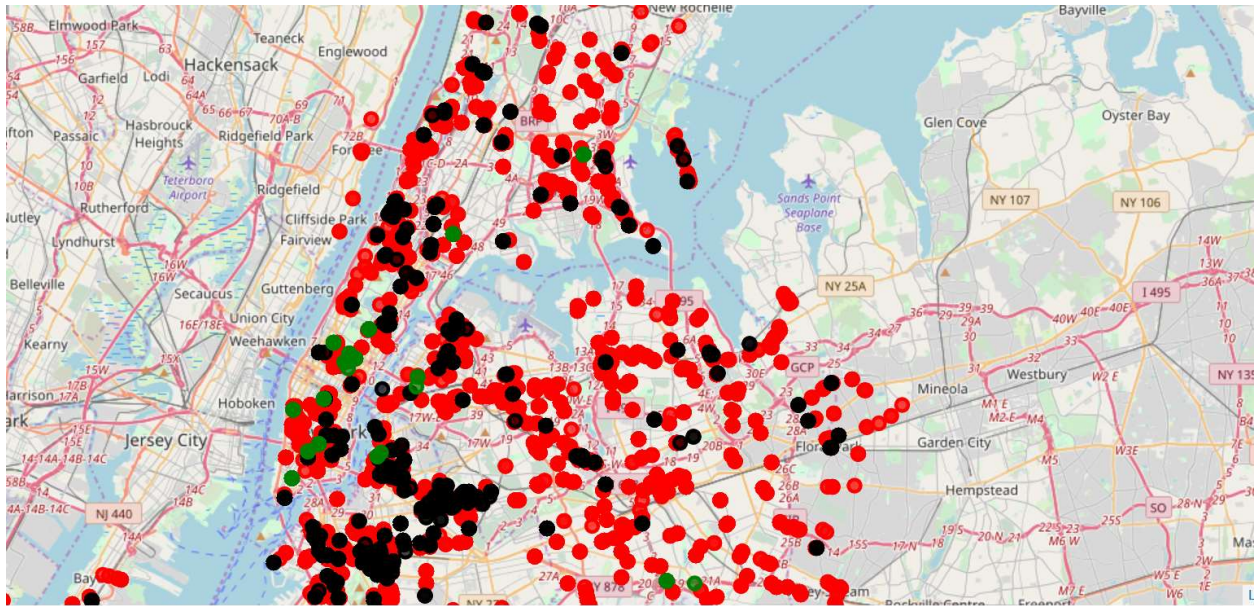

Figure 5: Hotels (Green dots), Restaurants (Red dots) and Bars (Black dots) in New York.

All the above points were used in the k-mean algorithm. The resulted clusters are shown in Figure 6.
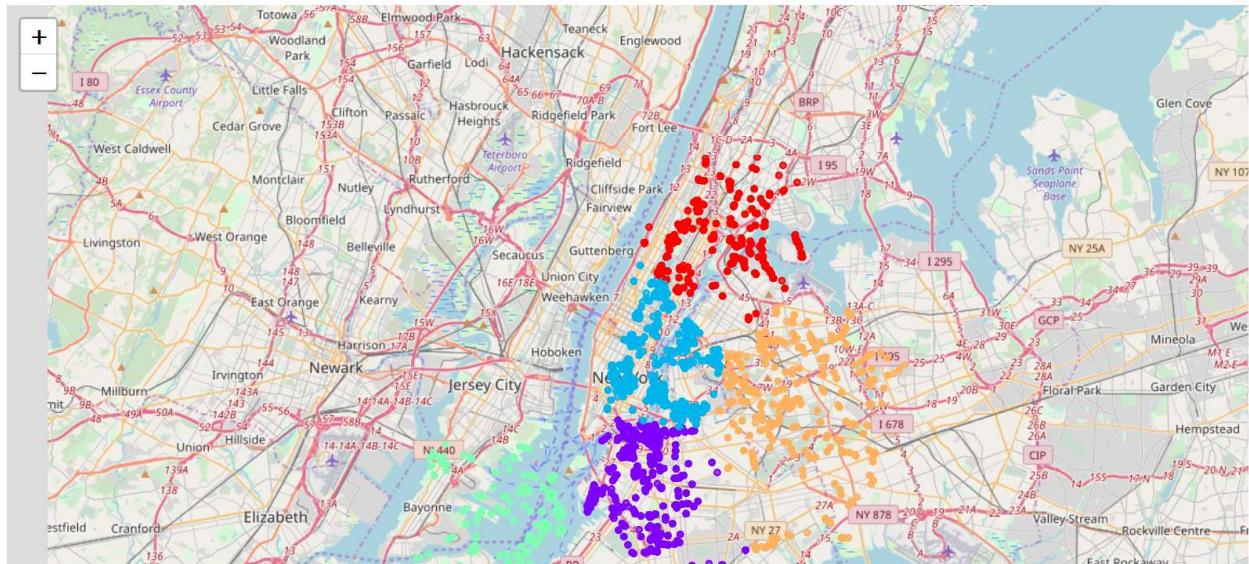
Fig 6: Shows 5 clusters of the venues indicated by the different colors.

The mean radius was calculated for each cluster. The location of the cold storage was defined by the clusters centroids calculated in the k-means algorithm. Moreover, the number of venues in each cluster was extracted and used to define the density in each area.

Discussion:

From the results we can see the variability in the distribution of venues in the New York city. As seen the two central clusters in the middle of the city contain at least a double number of hotels, restaurants, bars than the three-other cluster. For this reason, their areas are smaller than those of the other three clusters. Based on this observation we suggest:

• the size of the cold storages placed at the orange and purple areas be at least double than the size of the other cold storages.

• Similarly, employees employed in the central cold storages be at least double than those employed in the other three cold storages,

• the number of trucks in each cold storage to chosen regarding the number of venues in each cluster.

Conclusions

The main problem solved in this work is the definition of the locations where 5 cold storages can be built to improve the process of meat distribution in the New York city. Our approach was to divide the New York city in 5 different sub regions. The criterion used for this division was the distribution (density) of the venues of interest in the entire city. The center of each sub region was estimated using the k-means algorithm. The standard deviation of each cluster was used to define the area containing the venues to which the cold storage must provide the products. Based on the results, we can define the location of the warehouses, their size, the number of employees in each cold storage and the number of trucks required.