## **Choosing a location for a distribution center for a food supplier to maximize business.**

**Capstone Project**

**Melvin Sebastian**

**May 22 2019**

### **1.Introduction**

In this project I aim to find location co-ordinates for a distribution center for a food supplier to maximize business.We choose the city of New York to set up a distribution centre for a food supplier which is expanding. The limitation of each distribution center is that we have a maximum radius for distribution of 4km. So our goal is to choose a location which would give the outreach.

A good solution to this problem would help many distributors to select an optimum location for its center which would in turn help reduce the transportation costs which is an important consideration for any supply chain. This method can not only be used for distribution centers but also various links in any supply chain for a range of industries.

The only constraint considered in this problem is the range of the distribution centre. The quantity supplied, the number of supermarkets, the scale of each supermarket etc aren't considered taking into consideration the scope of our project. All of this can be added in a later project which can be applied to real world scenarios.

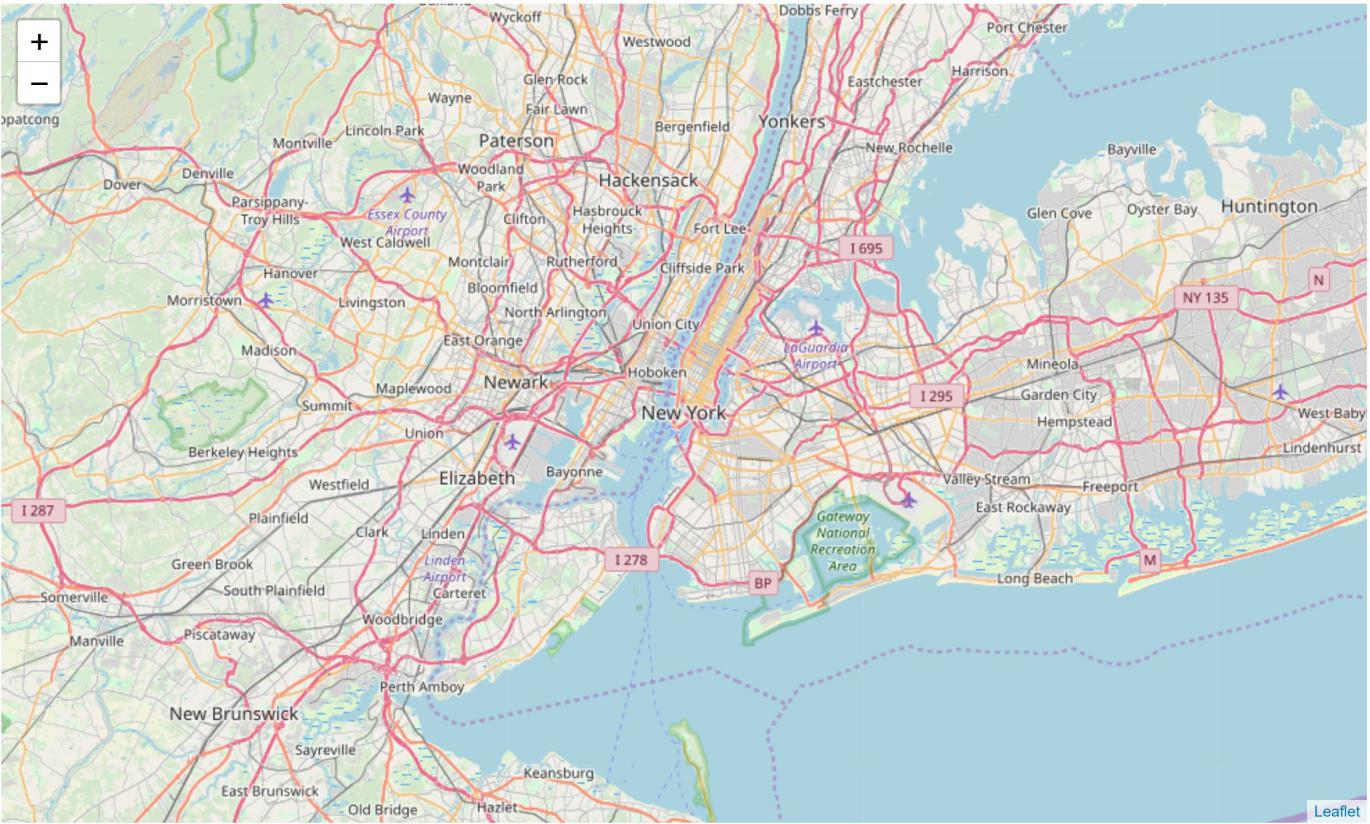
### **2.Data to be Acquired**

For this project we need to obtain the map of New York. What we are going to do is use clustering to obtain clusters of supermarkets in New York based on their proximity to one another. Hence to complete this project. We also need the coordinates of the various supermarkets in NY which can be easily obtained on Foursquare using its API.

1. **Methodology**
   1. Data acquired and cleaning:

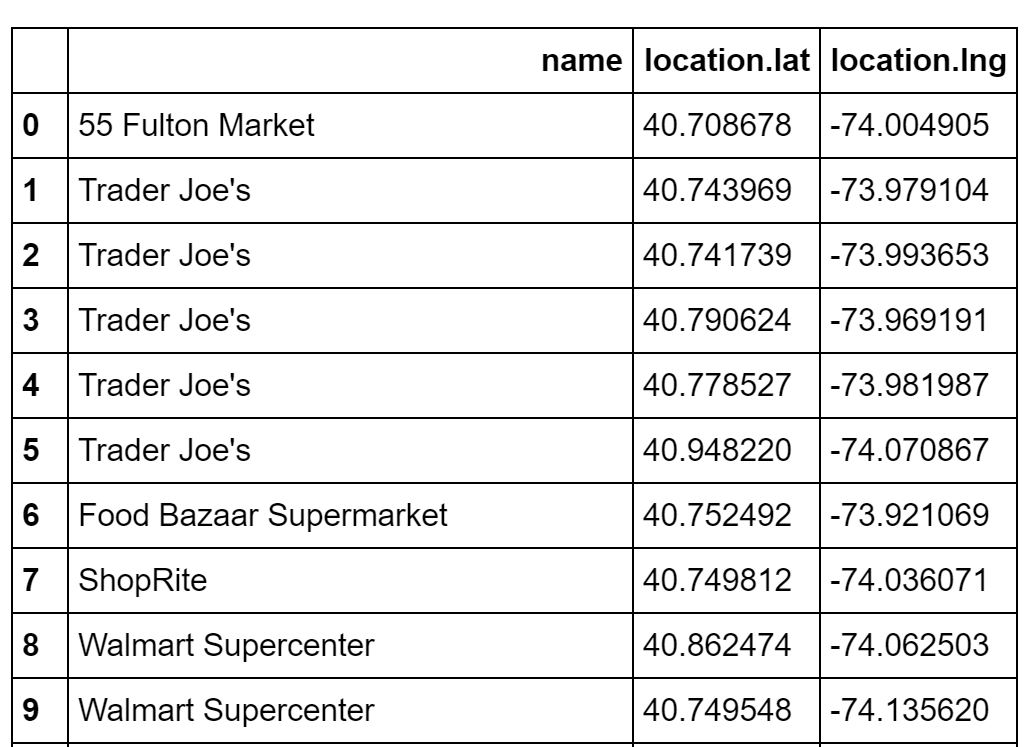
Firstly, we need a map of New York. This is easily obtained by obtaining the coordinates of NY using Geolocator API on Python.

Using these coordinates we obtain the map of NY using the folium as shown below:

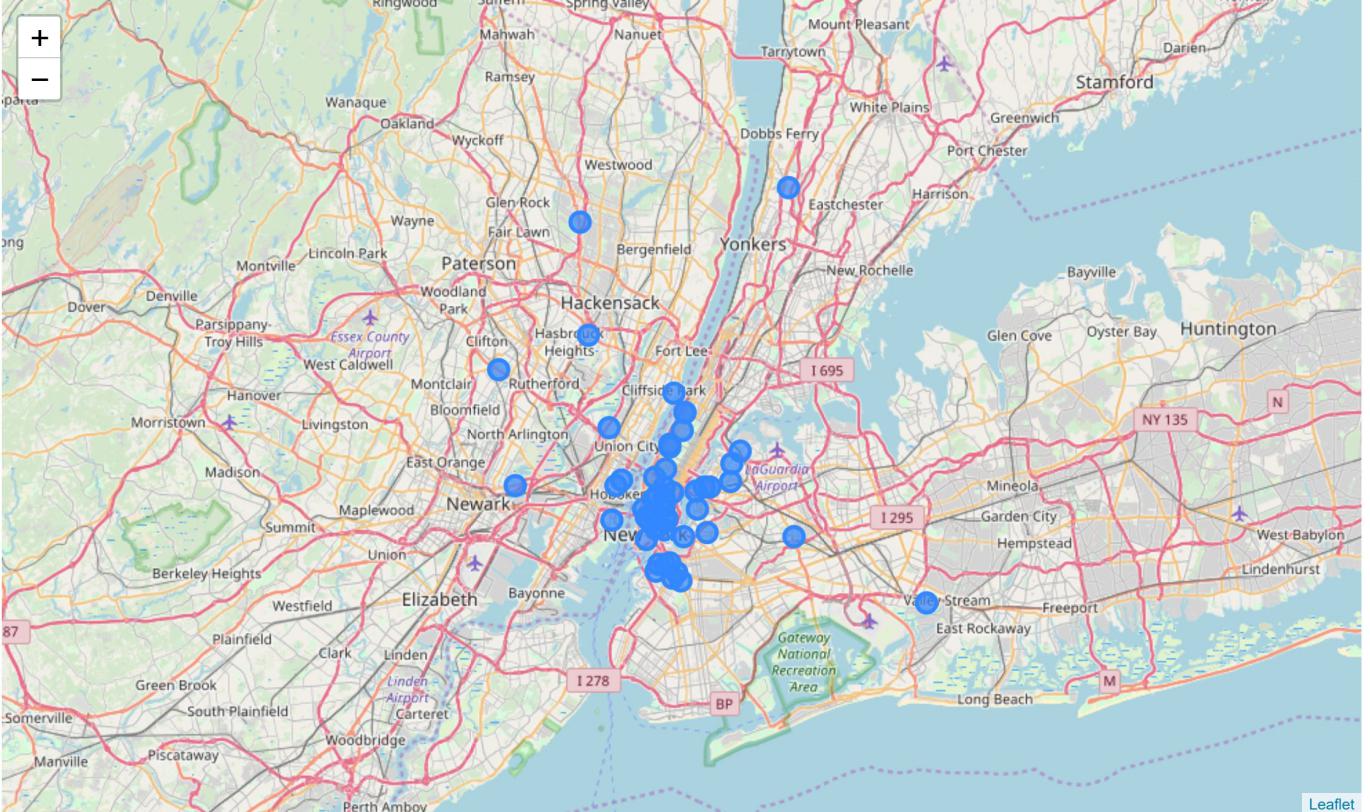


Now that we have obtained our map. Our next step is to obtain the locations and details of various supermarkets in NY. This is done using the Foursquare API. We should note however, due to our limited account, the number of supermarkets obtained is restricted to 50. Hence the size of our dataset remains limited to 50. However, if a larger number is required, it is easily obtainable by paying for the data on Foursquare.

The data acquired from Foursquare is extremely unstructured. Hence we clean it and form a Pandas dataframe with three columns: Name ,Longitude and Latitude as shown:



Using this data we obtain the map of NY with the locations of the supermarkets marked on it using Folium as shown below:



* 1. **Machine Learning method:**

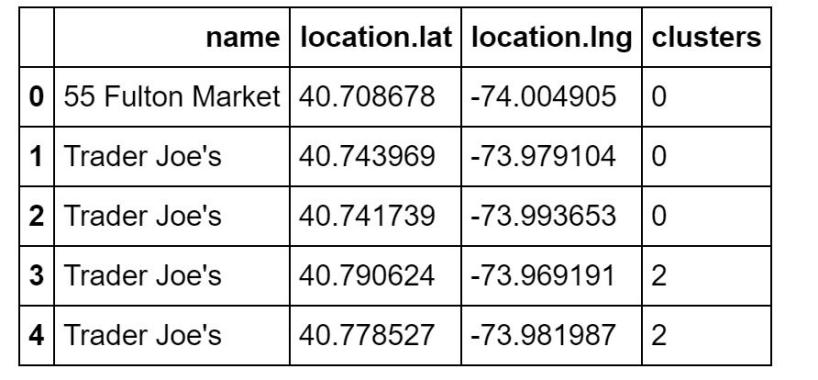
Now that we have got all the data required, cleaned and structured. We need to locate the actual point of the distribution centre to be build. We do that using a clustering method called **mean-shift algorithm.**

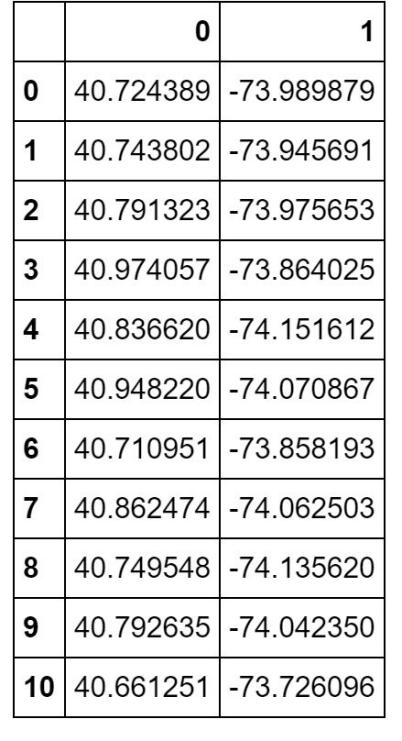
Mean shift clustering aims to discover “blobs” in a smooth density of samples. It is a centroid-based algorithm, which works by updating candidates for centroids to be the mean of the points within a given region. These candidates are then filtered in a post-processing stage to eliminate near-duplicates to form the final set of centroids.

It is ideal for our project as we need to find the highest density area with a specified radius. The number of centers need not be specified either and a radius of our choice can be input.

The radius otherwise called bandwidth is given as 4Km is found to be 0.0363636 in terms of coordinates.

Using this clustering technique we find that we find 10 clusters with the maximum frequency the any of the clusters being 25.The value of the cluster for each supermarket is giving in the form of a dataframe below:



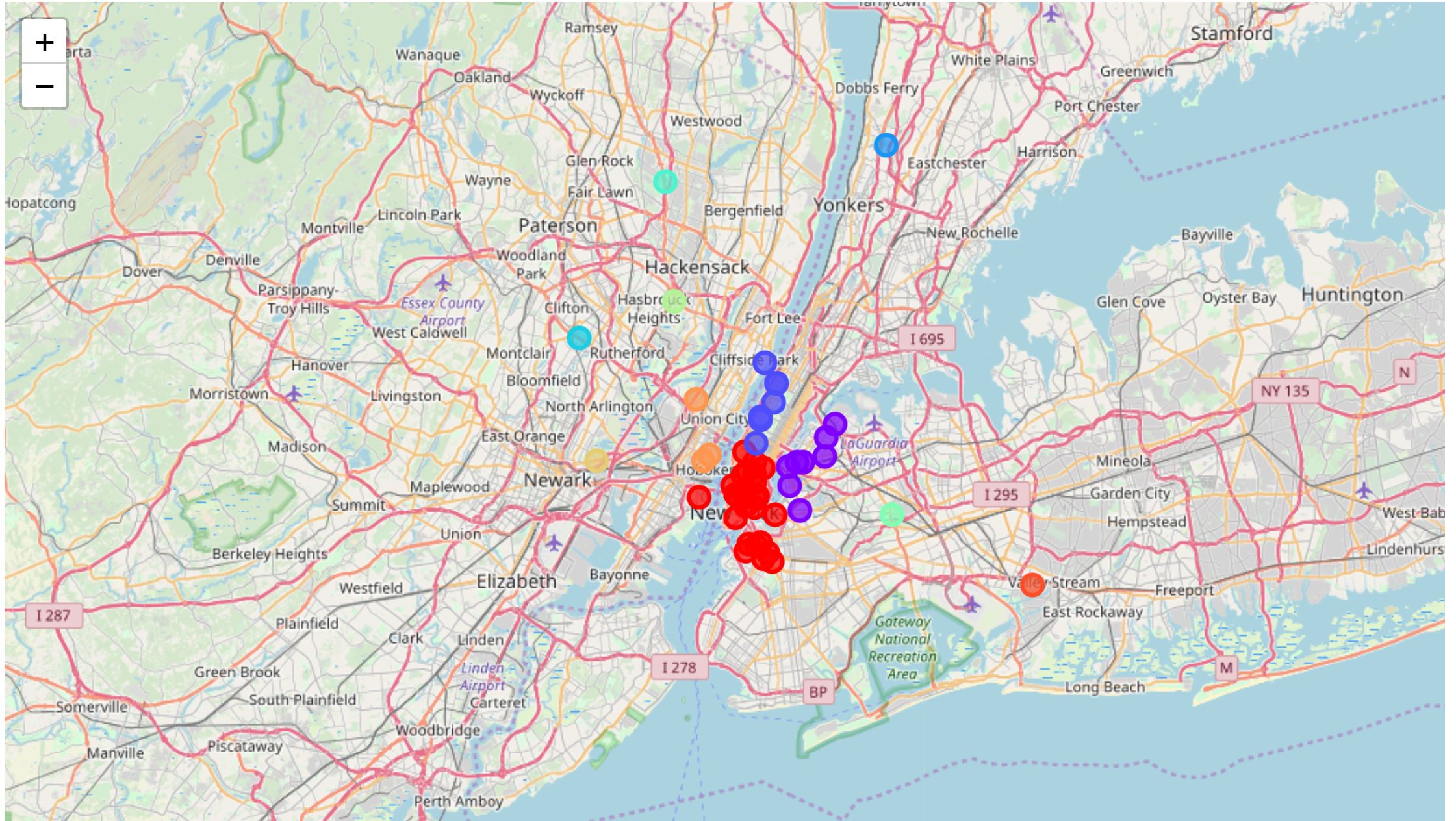
The centers of all the clusters are also found and given in the table shown: 

The coordinates of the cluster with the maximum frequency is found to be:

**40.72438942 , -73.98987903.**

1. **Results and Discussion:**

We have found the exact coordinates of our location required for the maximum outreach of supermarkets and this was found to be at **40.72438942 , -73.98987903.**

The various supermarkets that comes under this distribution center is marked in red in the figure given below.

As we can see the algorithm has found the highest density regions on the map.

The difference between K-means and shift-means algorithm is understood.

The drawbacks of this model include the lack of data on the supermarkets.

1. **Conclusions and future work:**

This project was executed successfully albeit a few drawbacks and improvements

that can be done in the future. These include including more data on the shops,

eliminating the smaller supermarkets, including the scale of each supermarket,

Considering other factors such as locality, resources such as logistical services near

by etc., the terrain etc.

A firm can be greatly helped using this model with the above modifications which

is unfortunately outside the scope of this capstone.