**Objective:**

This project aims at finding a suitable location for setting up a Warehouse for vegetables vendor.

The recommended location should be able to enable easy and faster supply of vegetables to the customers. Also should aim at a place where vegetable consumption is high so that gets good revenue.

**Business Problem:**

For a vegetable vendor the place where he starts the warehouse should be near to the locations he need to supply and should be able to reach without hassles. Also the requirement for vegetables in that area should be high.

Therefore to solve the current problem, we need to find a location where

* No. of restaurants are more
* No. of supermarkets are more
* Population is more

There are other factors which the current problem may depend on such as rent, traffic conditions, area type etc. There factors are currently out of scope for this project due to unavailability of such data.

We consider New York city in this project to find a suitable location for a vegetable vendor to setup warehouse for his business expansion.

**Data:**

To resolve the aforesaid problem, the basic data we need is the latitude and longitude of each area in NewYork city. I refer each location based on zip code of each area in New York. I get this data from

<https://gist.github.com/erichurst/7882666/>

Once we get the latitude and longitude, we need to find top 10 nearby places for each area.

I get this by clustering top venue for each of the areas in New York.

I use FourSquare api to get the nearby places <API: getNearbyVenues()

Another important factor we consider is the population in that area. I get the population data from

<https://www.zip-codes.com/city/ny-new-york.asp>

Sample Data:

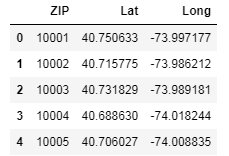
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ZIP Code** | **Type** | **County** | **Population** | **Area Code(s)** |
| [ZIP Code 10001](https://www.zip-codes.com/zip-code/10001/zip-code-10001.asp) | Standard | [New York](https://www.zip-codes.com/county/ny-new-york.asp) | 21,102 | [646](https://www.zip-codes.com/area-code/area-code-646.asp) / [718](https://www.zip-codes.com/area-code/area-code-718.asp) / [917](https://www.zip-codes.com/area-code/area-code-917.asp) |
| [ZIP Code 10002](https://www.zip-codes.com/zip-code/10002/zip-code-10002.asp) | Standard | [New York](https://www.zip-codes.com/county/ny-new-york.asp) | 81,410 | [718](https://www.zip-codes.com/area-code/area-code-718.asp) |
| [ZIP Code 10003](https://www.zip-codes.com/zip-code/10003/zip-code-10003.asp) | Standard | [New York](https://www.zip-codes.com/county/ny-new-york.asp) | 56,024 | [212](https://www.zip-codes.com/area-code/area-code-212.asp) / [646](https://www.zip-codes.com/area-code/area-code-646.asp) / [347](https://www.zip-codes.com/area-code/area-code-347.asp) / [917](https://www.zip-codes.com/area-code/area-code-917.asp) / [718](https://www.zip-codes.com/area-code/area-code-718.asp) |

Based on above data, I find the top 3 areas in which population, restaurants and supermarkets are more to setup a vegetable warehouse.

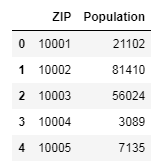
This concludes my data section.

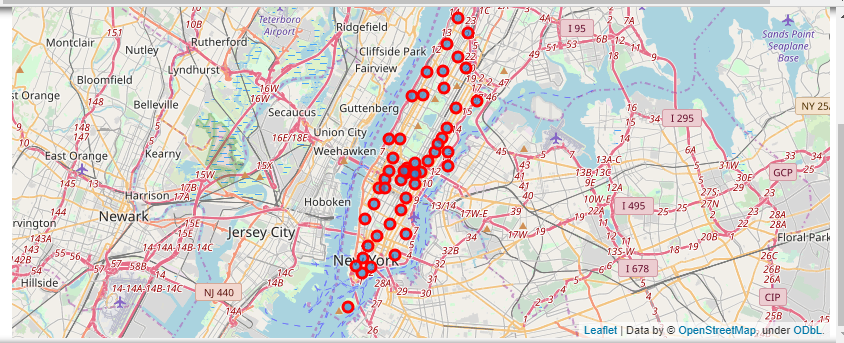
**Methodology**

* The dataset was populated with the latitude and longitude of each zip code in New York city

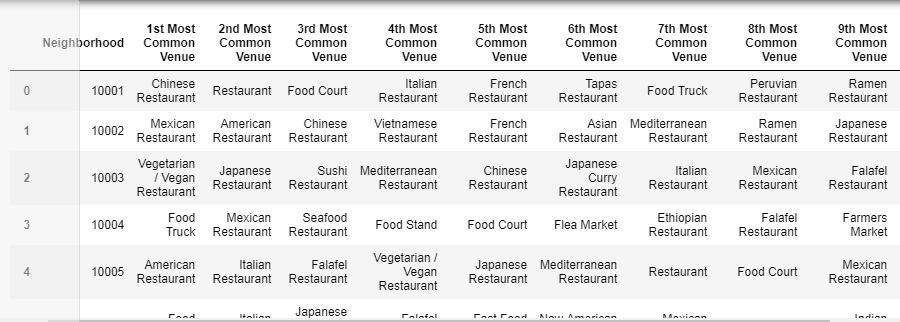
****

* Population data for each zip code was obtained from the url given in the data section.

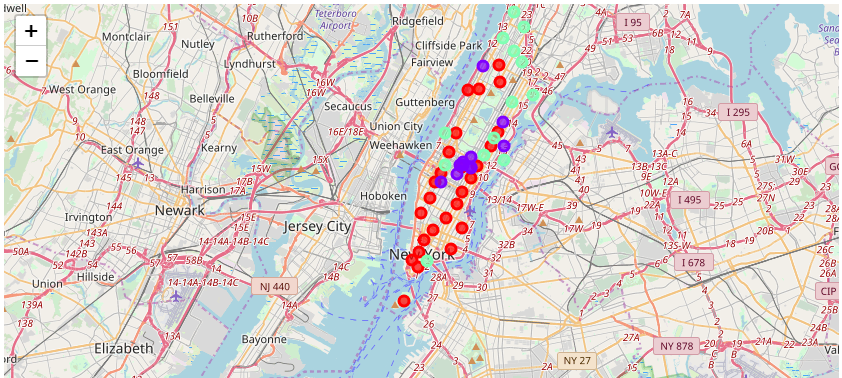
****

****

* Both the above data was merged to form a dataframe which was used for further processing
* For each zip code in New York City, the nearby neighborhood venues were obtained from FourSquare.
* Further the dataframe was processed to obtain only the venue which are of Restaurant Category, Food Category and Market Catgory

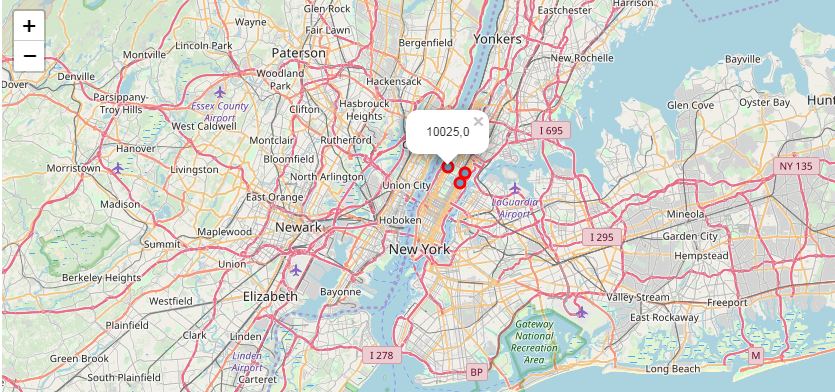
****

The obtained data was clustered into 3 groups so that the vendor can start his warehouses one in each group to diversify investment across different markets (Eg., Indian restaurant could be popular in some markets whereas American restaurant could be popular in some markets. Accordingly vegetable consumption varies across clusters)****

****

• From this clustered data, top zip area across each cluster with highest population is picked up to suggest vendor for setting up warehouse.

****



**Conclusion:**

From the obtained results, we can see that the zip areas with codes 10025, 10029 and 10128 are the most suitable places to set up a warehouse across various markets having restaurants, super markets and population ;considering the parameters analyzed in this project scope.

The same has been indicated in the maps with zip code and cluster label.

**Recommendation**

The following analysis can be improved with following extensions:

• Dataset can be improved by including the parameters like rent , maintenance cost for the warehouse in each zip area.

* Warehouse may not be allowed to be setup in mid of a residential area. Hence the type of the area can be included as one more parameter.
* Other clustering techniques like Dbscan can also be analyzed to get more pruned clusters for avoiding outliers.