

# Choosing location for a new grocery store

Minh-Quan Dang

## Introduction

### Background

I am an investor with experience in retail business and seeking opportunities to open a new grocery store in New York City. New York is a big city with 306 neighbourhoods. Plenty of opportunities.

### Problem

Even though I have experience in running a store, I need to decide where is the best place to open my new store. I want to find out the neighbourhood where groceries is most likely welcomed by the people but on the other hand not so much competition.

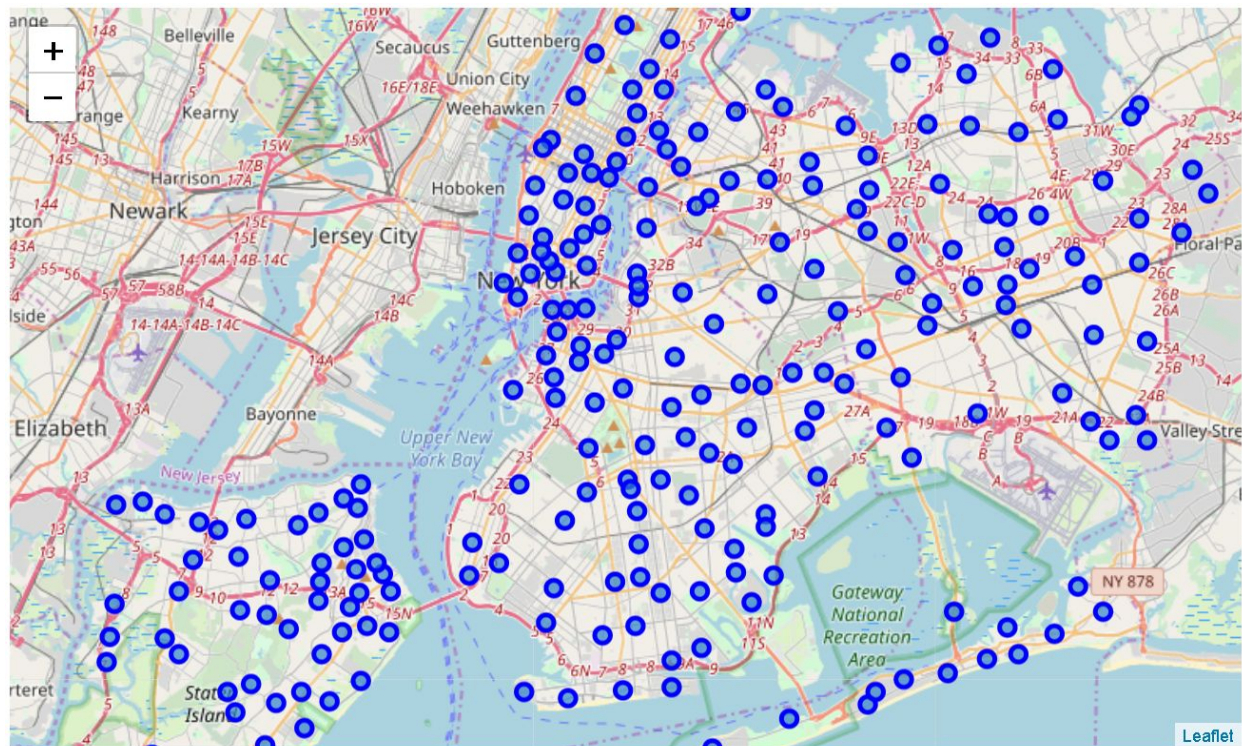
## Data description and sources

### Data source

Data set is aquiced from [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset) provided by Course creator. Data is under the json format and the interested information is extracted by customized function.

### Data exploratory

Map of New York and mark of neighbourhoods



## Methodology

Ideal is using the information about popularity of venues from Foursquare to identify the neighbourhood has high interest in Grocery stores of similar stores. The K means algorithm will be used to create group neighbourhoods to cluster according to their preference. The most appropriate cluster will be indicated for the most suitable location for opening the new store.

## Results

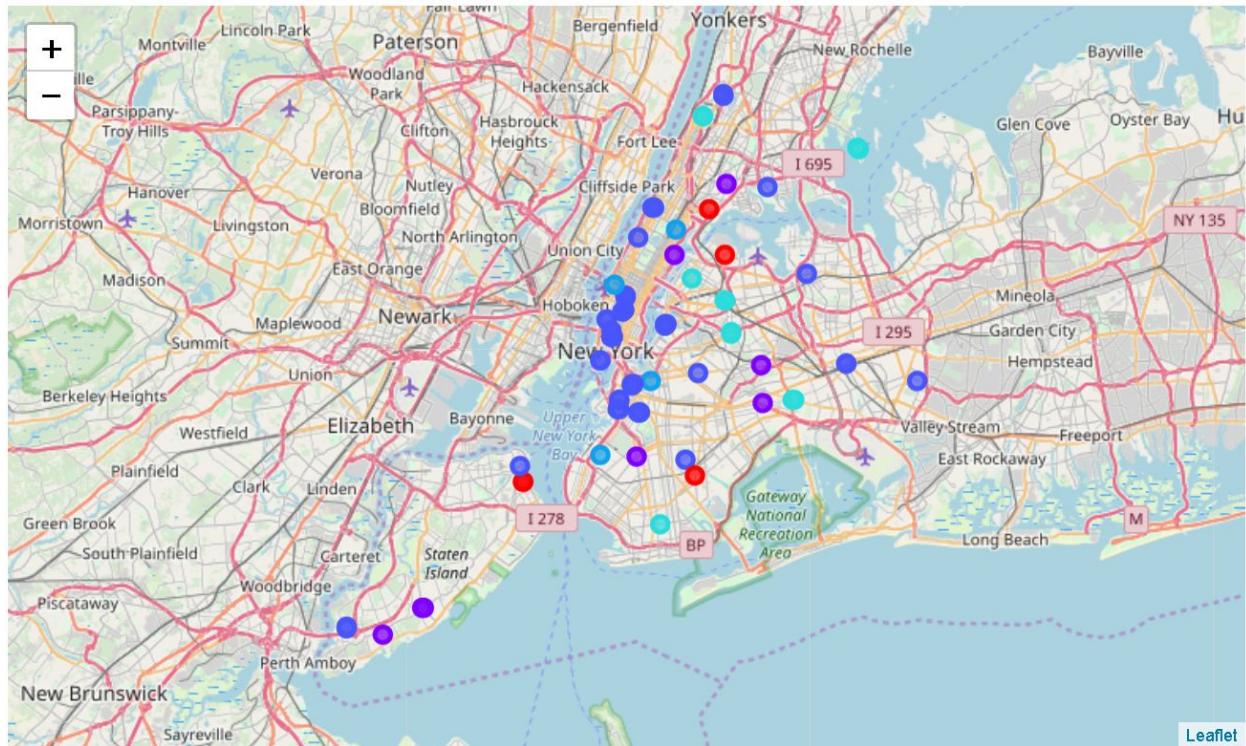
All the venues within a radius of 100 m from each neighbourhood are requested from Foursquare. There are intotal 873 venues of interest. The data look as follow:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Fieldston	40.895437	-73.905643	8106	40.895668	-73.904750	Wine Shop
1	Fieldston	40.895437	-73.905643	nicksemlerSPA	40.894942	-73.905475	Spa
2	Kingsbridge	40.881687	-73.902818	Garden Gourmet Market	40.881350	-73.903389	Gourmet Shop
3	Kingsbridge	40.881687	-73.902818	MyUnique	40.881966	-73.903584	Thrift / Vintage Store
4	Kingsbridge	40.881687	-73.902818	Mattress Firm	40.881580	-73.903277	Mattress Store

There are 218 unique venue categories, next step is fine out only the venues which are store type.

```
array(['Thrift / Vintage Store', 'Mattress Store', 'Discount Store',
      'Grocery Store', 'Jewelry Store', 'Liquor Store',
      'Convenience Store', 'Pet Store', 'Arts & Crafts Store',
      'Fruit & Vegetable Store', 'Hardware Store', 'Shoe Store',
      'Furniture / Home Store', 'Electronics Store', 'Accessories Store',
      'Toy / Game Store', 'Department Store', 'Big Box Store',
      'Lingerie Store', 'Camera Store', "Women's Store", "Men's Store",
      'Shipping Store', 'Paper / Office Supplies Store', 'Video Store',
      'Kids Store', 'Health Food Store', 'Clothing Store'], dtype=object)
```

Using K Means with 5 clusters we can have the location of neighbourhood with store as follow



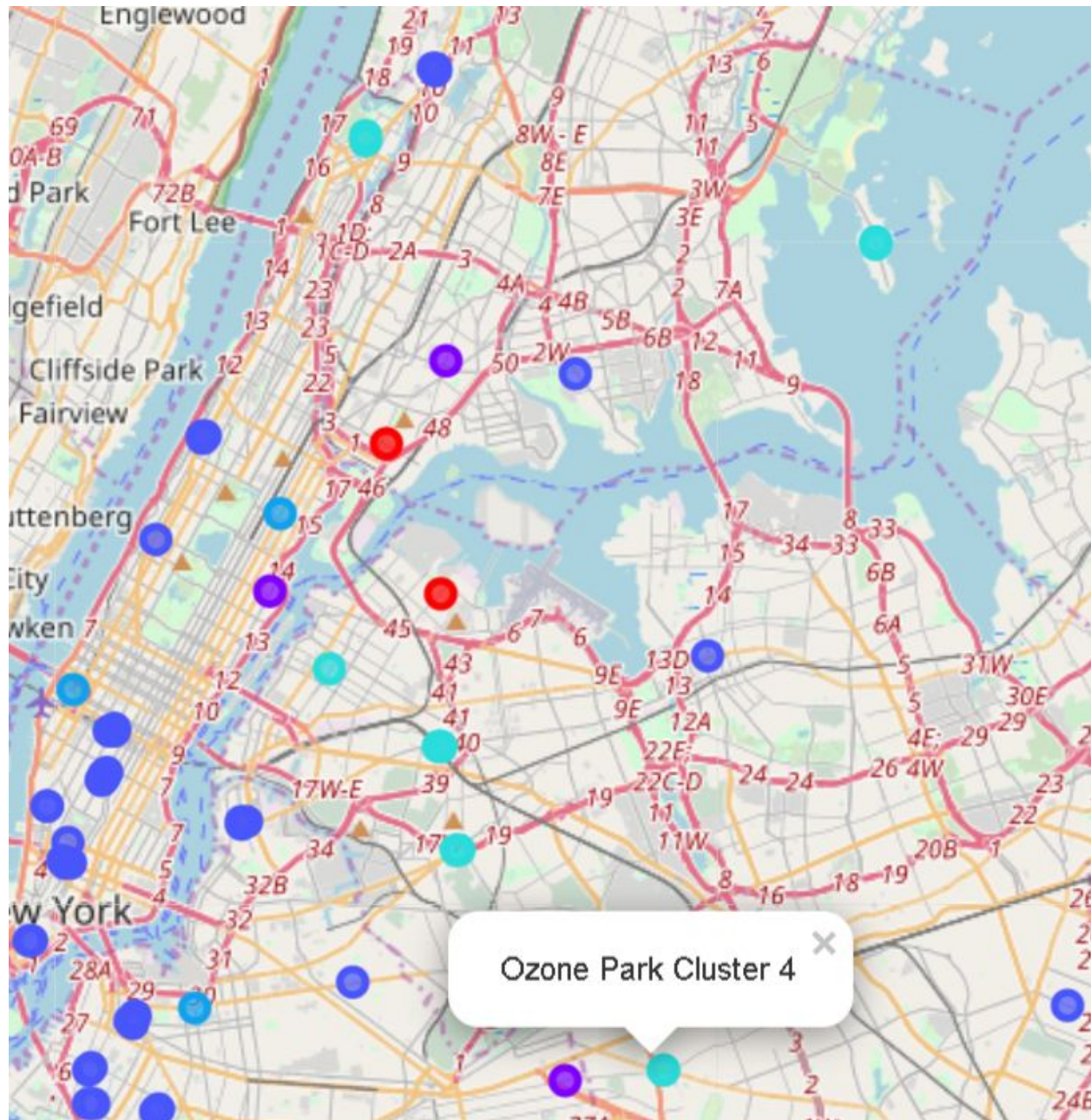
From the result we have 5 type of neighbourhood according to their 3 most interest store type

- Cluster 1: Discount store, Women's store, Video Store
- Cluster 2: Liquor Store, Women's store, Hardware Store
- Cluster 3: Thrift / Vintage Store, Mattress Store, Discount Store
- Cluster 4: Pet Store, Women's store, Hardware Store
- Cluster 5: Grocery Store, Jewelry Store, Women's Store

## Discussion

Clearly I should choose neighbourhoods in cluster 5 to open my new store





## Conclusion

By using data provided by Course creator and Foursquare, I was able to find the most suitable location to open my new store. Further analysis could be made to compare and find the best location from those neighborhoods we found but the initial results are relatively encouraging.