New York City—The battle of the Neighbourhoods

IBM Data Science Capstone Project

Analyzing and visualizing the structure of New York City in relation to the requirements of a Clothing Store Investor

Business Problem

The project is based on a hypothetical business case. A Canadian Investor who recently made a fortune with an investment in a Clothing Store in Toronto wants to repeat his idea in New York City.

1. As his brand is exclusive and expensive the location should be one of the most crowded districts with high employment rate and above average income. He prefers not only tourists to buy in the store he would also like to gain many regular customers.
2. Due to the origin of his brand has a touch of Italian design he prefers a location close to Italian restaurants on the basis of window shopping and the chance that people who go for Italian food also have a sympathy for Italian fashion is pretty high.
3. Tourists and business traveler are well known for spending money generously, therefore the criteria to be as close as possible to hotels is highly important, because guests of the city hotels are more likely to buy clothes nearby and guarantee for more walk-in customers.
4. As close to the city Center or other touristic hotspots to benefit from walk-in customers. Approximately 20 Minutes walking distance to the Center of the district. If possible far away from other clothing stores.
5. The Investor wishes to invest in a flat in New York City to be nearby the store. By the reason to live close to the store he has the following criteria to his place of residence: low crime rate, high community trust, close to parks, theatres and art galleries.

The Investor first wants a macro overview of New York City. So we are exploring the community districts.

1. Business Problem Understanding

The Project seems very clear, find the perfect district for an Italian brand clothing store, taking into account the location should be suitable to his imaginations of the perfect place of residence, where you feel safe at the same time.

2. Analytical Approach

The core of the project will be the socio-economic data frame. Complementary we build a venues data frame fetched from foursquare and explore these venues. The final venues frame will contain the most common venues of each district, which we will get through one hot encoding. This data frame is the basis for the k-means algorithm to cluster the districts by their features to compare similarity between these districts.

	Requirement	Weight
0	high population	0.8
1	high income	0.8
2	low unemployment	-0.5
3	low crime rate	-0.7
4	high communtiy trust	0.7
5	away from other Boutique	-0.3
6	away from other Clothing Stores	-0.3
7	close to Italian Restaurants	0.5
8	close to Hotels	0.9
9	close to park	0.5
10	close to Theater	0.3
11	close to Art Gallery	0.3

For the best result the analytical solution to the business problem is to quantify and evaluate the thoughts of the client to full fill his requirements completely. For evaluating his criteria, we will create a **features weighted matrix** to express the investors desires in a scientific way, which we will multiply with the normalized final data frame to add the extra column with the weighted results, which gives us an indication of the best districts.

3. Data requirements and collection

To ensure the best location for the store I decided to add some more complexity to the standard course problem. As you can see from the criteria given by the investor we need some more data.

In the beginning of the Project I found data from many different data sources, but decided to get the data mainly from cccnewyork.org by the reason that the source of their Data is the U.S. Census Bureau and the data was fetched by the American Community Survey https://data.census.gov/. So we can be sure the data is up to date, consistent and reliable.

- the socio-economic data will be obtained from various csv files from **cccnewyork.org**
- the venues will be fetched from **Foursquare** through an API
- the *Geo-coordinates* will be obtained with **nominatim** and **geopy**

4. Data understanding and preparing

First of all, we will build a clean socio-economic data frame with all the necessary information which are related to the business problem. Therefore, we need to load all the files and drop all unnecessary columns and rows.

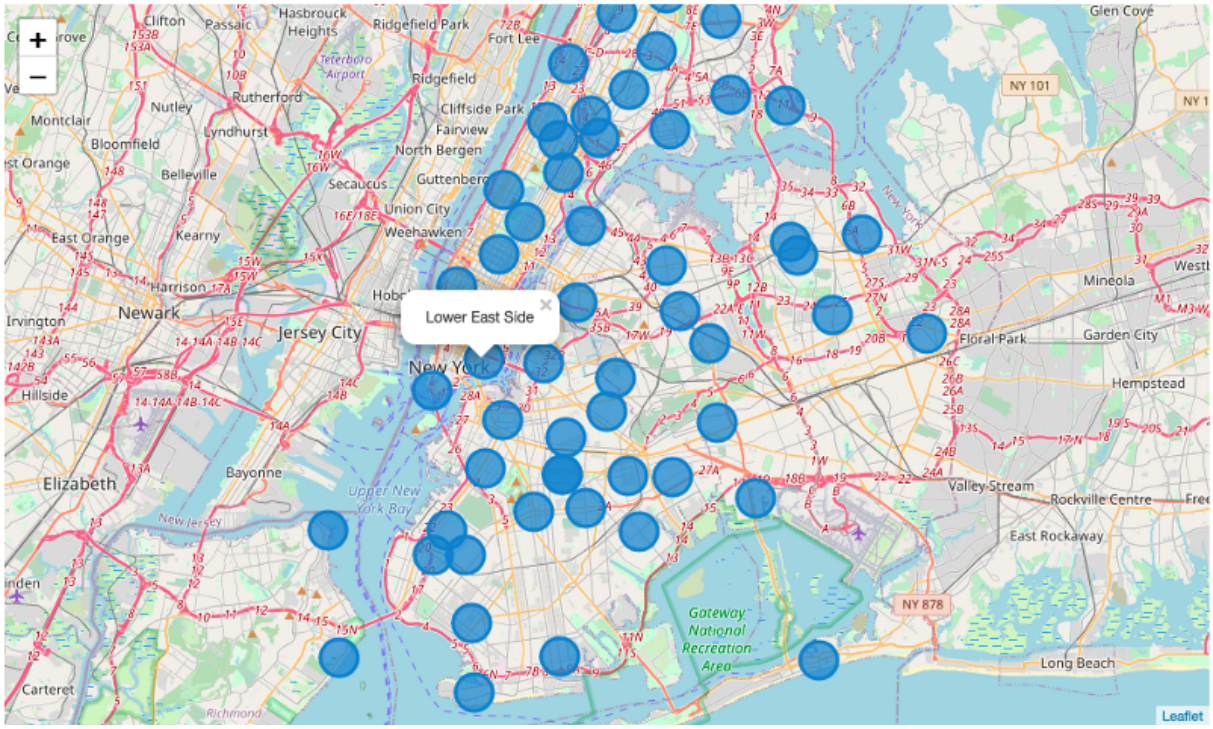
Index	Area	Population	Income	Unemployment in %	Felonies	Trust in %	N_lat	N_long
0	Astoria	160871.0000	67444.0	0.044	769.0	0.73	40.772014	-73.930267
1	Battery Park	61375.7862	148152.0	0.037	143.0	0.70	40.703012	-74.015825
2	Bay Ridge	125200.0000	72402.0	0.036	192.0	0.74	40.633993	-74.014584
3	Bayside	115744.0000	86338.0	0.035	110.0	0.86	40.768435	-73.777077
4	Bedford Park	133784.0000	34349.0	0.098	881.0	0.63	40.870100	-73.885691

Getting latitudes and longitudes with geocoder

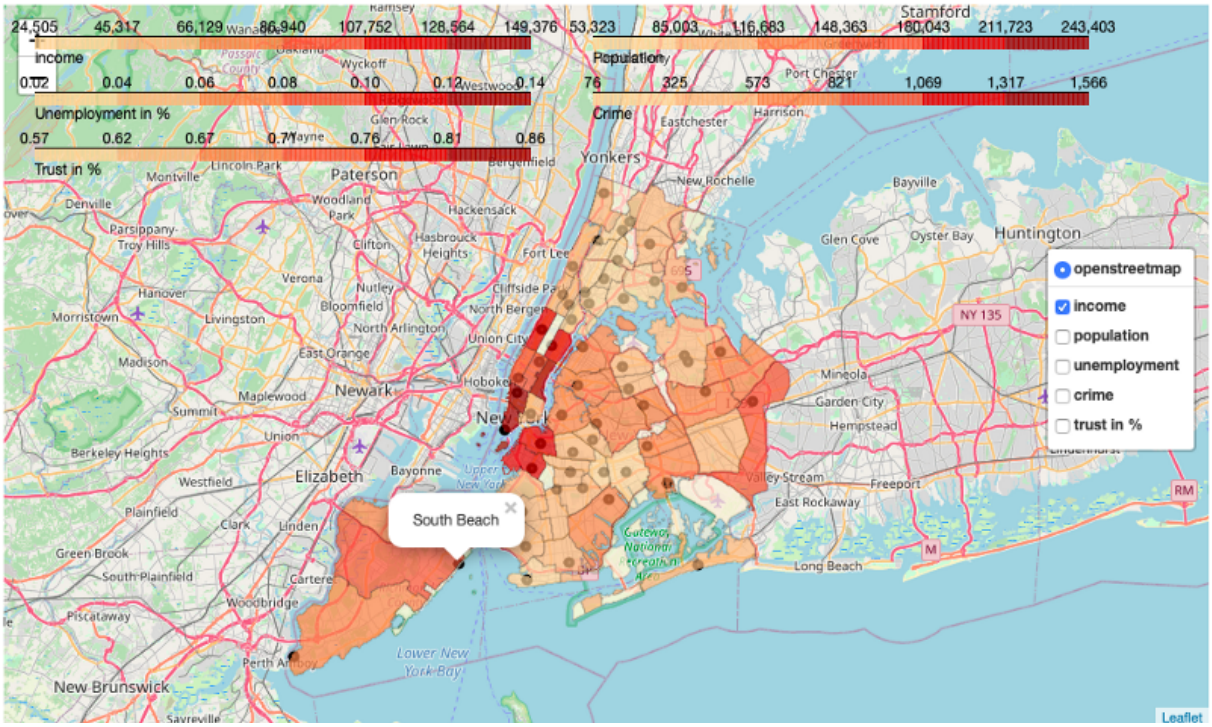
As mentioned before geocoder is a great tool to get the **latitude** and **longitude.** If you query is too large you can use the rate limiter for fetching larger amounts of queries successful.

Mapping with Folium

Folium is a great package to make beautiful maps. We will use it for a general overview of the districts of New York City to get familiar with the structure of the City and for interactive choropleth maps.



By adding **chloropleth** layers for each columnm, the map gets more interactive and informative. Simply add a layer for each column similar to the code below.



Getting the venue data with Foursquare

With Foursquare we can get up to 100 venues for each district, which is great for a free service. We will fetch the data and create a venues data frame, the pandas build in method .get_dummies lets us easily use the one hot encoding process to quantify the venues. After grouping the frame by the districts and calculating the mean value we can compare the different districts perfectly. In the jupyter Notebook you can comprehend the venue exploring detailed. But what is one hot encoding again?

One Hot Encoding is a process in the data processing that is applied to categorical data, to convert it into a binary vector representation for use in machine learning algorithms

One-Hot Encoding simply creates one column for every possible value and put a 1 or 0 in the appropriate column.

	Area	Accessories Store	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Aquarium	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asia Restaurar
0	Astoria	0	0	0	0.02	0	0	0	0	0	0	0.01	0	
1	Battery Park	0	0	0	0.02	0	0	0	0	0	0	0	0	
2	Bay Ridge	0	0	0	0.04	0	0	0	0	0	0.01	0	0	0.0
3	Bayside	0	0	0	0.03	0	0	0	0	0	0	0	0	
4	Bedford Park	0	0	0	0	0	0	0	0	0	0	0	0	
5	Bedford Stuyvesant	0	0	0.01	0.01	0	0	0	0	0	0	0	0	
6	Bensonhurst	0	0	0	0	0	0	0	0	0	0	0	0	
7	Borough Park	0	0	0	0.01	0	0	0	0	0	0	0	0	0.0
8	Brownsville	0	0	0	0	0	0	0	0	0	0	0	0	
9	Bushwick	0	0	0	0.01	0	0	0	0.01	0	0	0	0.01	0.0
10	Canarsie	0	0	0	0	0	0	0	0	0	0	0	0	
11	Central Harlem	0	0	0.04	0.03	0	0	0	0	0	0.01	0.01	0.02	
12	Chelsea	0	0	0	0.02	0	0	0	0	0	0.05	0.02	0.01	
13	Concourse	0	0	0.01	0.02	0	0	0	0	0	0.01	0	0.02	0.0
14	Coney Island	0	0	0	0	0	0.02	0.02	0	0	0	0	0	

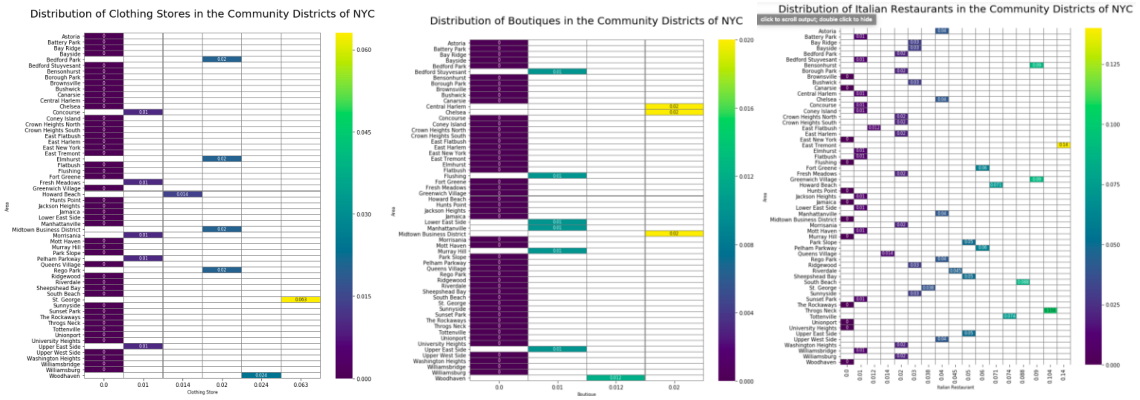
The most common venues

For the comparison of the districts we would like to create a table with a function which gives us the most common venues of each district. We can use this function later to explore the different cluster by their venues.

	Area	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	11th Most Common Venue	12th Most Common Venue
0	Astoria	Greek Restaurant	Bar	Pizza Place	Café	Italian Restaurant	Park	Bakery	Gym	Grocery Store	Bagel Shop	Deli / Bodega	Gourmet Shop
1	Battery Park	Coffee Shop	Hotel	Park	Pizza Place	Plaza	Gym	Memorial Site	Gym / Fitness Center	Café	Cocktail Bar	Falafel Restaurant	Scenic Lookout
2	Bay Ridge	Pizza Place	Bakery	Mexican Restaurant	Bagel Shop	American Restaurant	Italian Restaurant	Bar	Seafood Restaurant	Chinese Restaurant	Bank	Scenic Lookout	Restaurant
3	Bayside	Korean Restaurant	Greek Restaurant	Bar	Bakery	Italian Restaurant	Coffee Shop	American Restaurant	Cosmetics Shop	Pizza Place	Burger Joint	Indian Restaurant	Sushi Restaurant
4	Bedford Park	Pizza Place	Garden	Diner	Park	Gym	Deli / Bodega	Mobile Phone Shop	Mexican Restaurant	Sandwich Place	Coffee Shop	Latin American Restaurant	Chinese Restaurant
5	Bedford Stuyvesant	Coffee Shop	Bar	Caribbean Restaurant	Pizza Place	Park	Cocktail Bar	French Restaurant	Mexican Restaurant	Bakery	Café	New American Restaurant	Wine Shop
6	Bensonhurst	Italian Restaurant	Pizza Place	Bakery	Pharmacy	Ice Cream Shop	Chinese Restaurant	Sushi Restaurant	Bank	Donut Shop	Bagel Shop	Supermarket	Dessert Shop
7	Borough Park	Pizza Place	Bakery	Chinese Restaurant	Vietnamese Restaurant	Tea Room	Seafood Restaurant	Grocery Store	Bank	Bubble Tea Shop	Asian Restaurant	Dessert Shop	Restaurant

Heat-map of the target venues

Related to the requirements of the customer we will have a closer look at the distribution of the Clothing Stores, Boutiques and Italien Restaurants in the City.



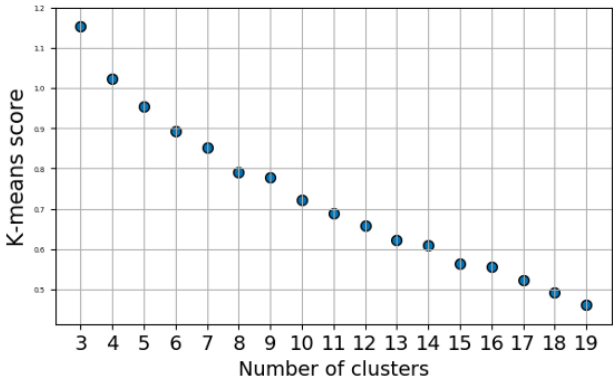
5. Analysing and Modelling

*You can find the detailed code **here***

This project has a need for data analysing through data exploring we will only use a simple classification algorithm but the main part is not about a machine learning model. We will use the **k-means** clustering followed by more data exploring and visualisation to expand our feeling for the data and understanding of the city.

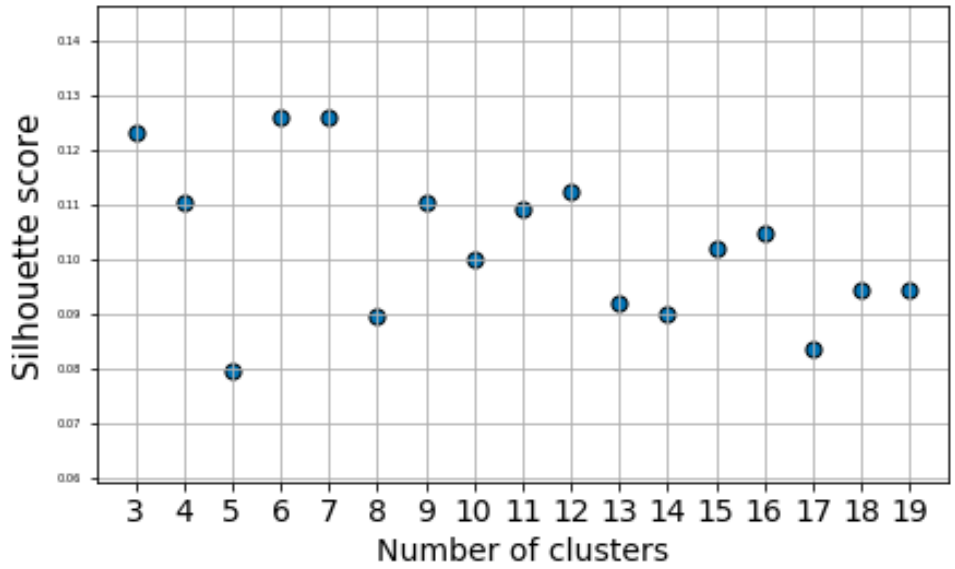
K-means is a method that aims to partition n data points into k clusters where each data point is assigned to the cluster with the nearest mean. The goal is to minimize the sum of all squared distances within a cluster.

To find the perfect number of cluster the most common approach is the **elbow method.** Therefore we run the algorithm multiple times and then plotting the related score.

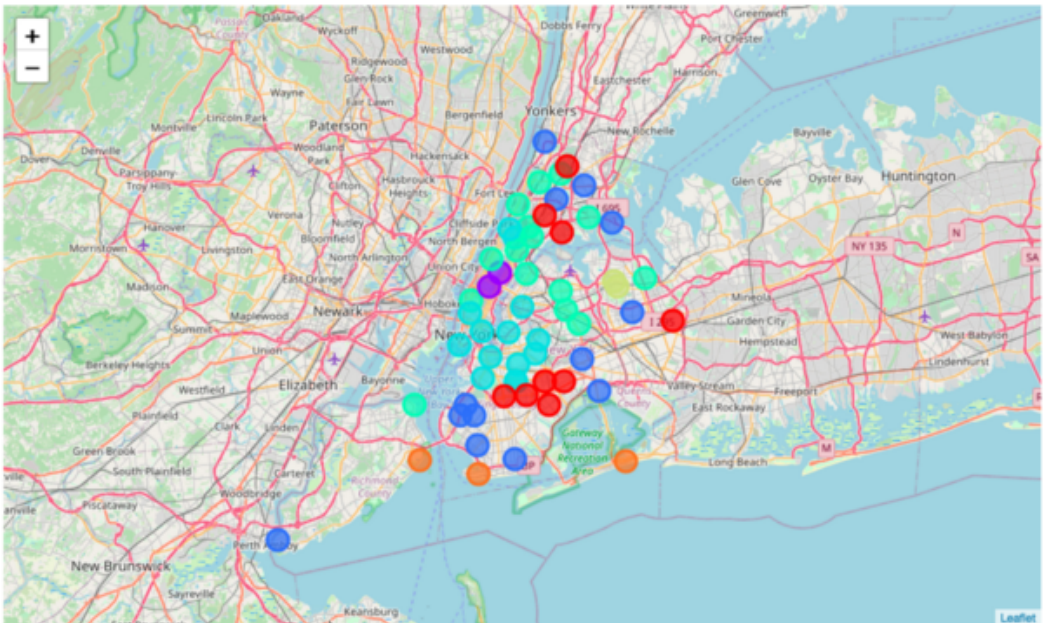


As you can see the elbow method is sometimes not very conclusive. But there are numerous different methods to determine the best number of clusters. The second method I used ist the Silhouette coefficient.

The Silhouette coefficient is calculated using the mean intra-cluster distance and the mean nearest-cluster distance for each sample. For each point p, first find the average distance between p and all other points in the same cluster this is a measure of cohesion (A). Then find the average distance between p and all points in the nearest cluster, this is a measure of separation from the closest other cluster (B). The silhouette coefficient for p is defined as the difference between B and A (B-A) divided by the greater of the two (max(A,B))



There are numerous quantitative methods of evaluating clustering results, you will see by using them as tools with the full understanding of the limitations the combination of contrasting methods rises the quality of your choice, if you be aware of actually examine the results, kind of a human inspection and making a determination based on an understanding of what the data represents, what a cluster represents, and what the clustering is intended to achieve, you will find the perfect number of clusters.



This is the clustered map of each Community District by the venue structure and similarity. #

Analyzing the Investor requirements

where solving the Business Problem begins

The clustered map above includes all venues we have fetched from Foursquare including the irrelevant venues, except the socio economic data. For the quality of the result it is important to deal only with relevant features, which have an impact on the decision of the Investor. Beginning from this part we will deploy the recently mentioned features weight matrix.

Initially we prepare and merge the data frames to include only the necessary columns.

	Area	Population	Income	Unemployment in %	Felonies	Trust in %	boro_cd	Boutique	Clothing Store	Italian Restaurant	Hotel	Park	Theater	Art Gallery
0	Astoria	160871.0000	67444.0	0.044	769.0	0.73	401	0.0	0.00	0.03	0.00	0.04	0.00	0.00
1	Battery Park	61375.7862	148152.0	0.037	143.0	0.70	101	0.0	0.00	0.01	0.05	0.05	0.00	0.00
2	Bay Ridge	125200.0000	72402.0	0.036	192.0	0.74	310	0.0	0.00	0.03	0.00	0.01	0.00	0.01
3	Bayside	115744.0000	86338.0	0.035	110.0	0.86	411	0.0	0.00	0.03	0.00	0.01	0.00	0.00
4	Bedford Park	133784.0000	34349.0	0.098	881.0	0.63	207	0.0	0.02	0.02	0.00	0.06	0.01	0.00

For the next part **Feature Scaling** is very important.

Feature scaling is a technique to change the values of columns in the dataset to use a common scale, without losing information or distorting the differences in the ranges of the values. This can be achieved through Normalization and Standardization

Normalization is a scaling technique which rescales the features so that the data will fall in the range of [0,1] to bring them to a comparable grade.

Standardization is a scaling technique which rescales the features the way they range between [-1,1] by the properties of a standard normal distribution with the **mean μ=0** and the **standard deviation, σ=1,** where μ is the average and σ is the standard deviation from the average.

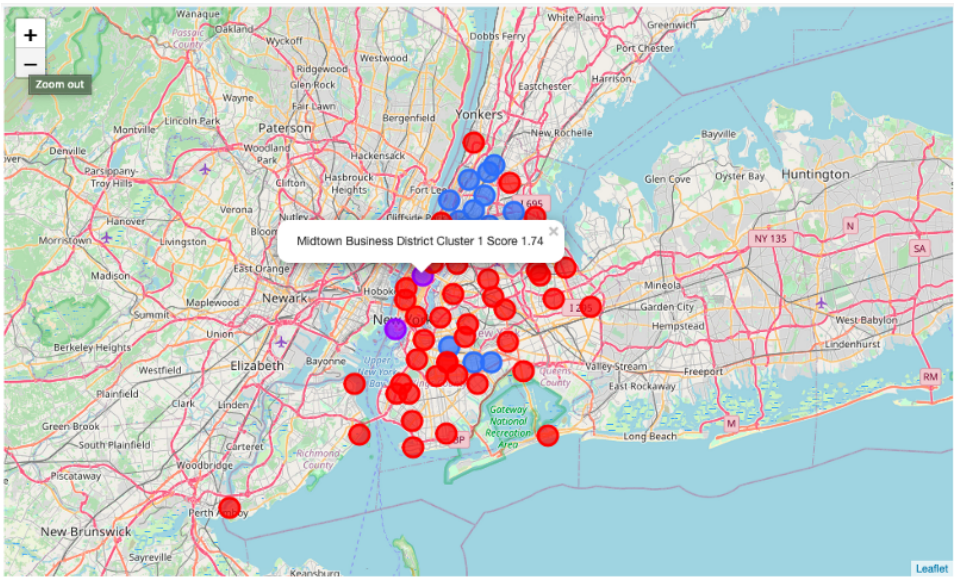
So after the normalization and setting the index on the Area our data frame looks like the following:

Area	Population	Income	Unemployment in %	Felonies	Trust in %	Boutique	Clothing Store	Italian Restaurant	Hotel	Park	Theater	Art Gallery
Astoria	0.666025	0.455235	0.314286	0.495809	0.848837	0.0	0.00000	0.214286	0.000000	0.4	0.0000	0.0
Battery Park	0.254103	1.000000	0.264286	0.092199	0.813953	0.0	0.00000	0.071429	0.714286	0.5	0.0000	0.0
Bay Ridge	0.518343	0.488701	0.257143	0.123791	0.860465	0.0	0.00000	0.214286	0.000000	0.1	0.0000	0.2
Bayside	0.479194	0.582766	0.250000	0.070922	1.000000	0.0	0.00000	0.214286	0.000000	0.1	0.0000	0.0
Bedford Park	0.553882	0.231850	0.700000	0.568021	0.732558	0.0	0.27027	0.142857	0.000000	0.6	0.0625	0.0

Now we can multiply the features weight matrix and calculate the total score column, with some simple visualisation the data frame looks pretty informative.

Area	Population	Income	Unemployment in %	Felonies	Trust in %	Boutique	Clothing Store	Italian Restaurant	Hotel	Park	Theater	Art Gallery	Total Score
South Beach	0.46	0.49	-0.136	-0.074	0.667	-0	-0	0.35	0.36	0.14	0	0	2.26
Battery Park	0.21	0.8	-0.132	-0.065	0.57	-0	-0	0.04	0.56	0.23	0	0	2.2
Upper West Side	0.64	0.66	-0.182	-0.181	0.578	-0	-0	0.14	0	0.45	0.04	0	2.15
Tottenville	0.55	0.47	-0.1	-0.046	0.7	-0	-0	0.26	0	0.17	0	0	2
Upper East Side	0.68	0.66	-0.096	-0.129	0.505	-0.15	-0.048	0.14	0.11	0.23	0.02	0	1.92
Sunnyside	0.46	0.38	-0.154	-0.09	0.619	-0	-0	0.11	0.45	0.05	0	0.06	1.88
Greenwich Village	0.3	0.8	-0.132	-0.162	0.57	-0	-0.048	0.32	0	0.18	0	0	1.83
Midtown Business District	0.18	0.55	-0.189	-0.232	0.537	-0.3	-0.097	0	0.9	0.09	0.3	0	1.74
Chelsea	0.37	0.55	-0.189	-0.234	0.537	-0.3	-0	0.14	0.22	0.27	0.02	0.3	1.69
Park Slope	0.39	0.66	-0.196	-0.113	0.667	-0	-0	0.18	0	0.09	0	0	1.68

After applying the k-means method featuring this data frame (dropping the total score column) won't get a visualization of the best districts numerically, but it shows us which districts are similar in accordance to the investor requirements. We will repeat the same process as mentioned before, finding the perfect number of clusters with the two method previously explained.



Red cluster 0 ist the medium level cluster the total mean of the features is mediocre. The mean total score is 1.39, but it's notable that it includes 5 of the top scored districts, especially South Beach and Tottenville, which are located in Staten Island. There are also 3 high ranked districts from Manhatten included. The rest of the cluster is moderate.

The **purple** Cluster 1 is the high ranked Cluster it consists of only 2 districts with an median total score of 1.97. The districts of this cluster Battery Park and Midtown Business District scoring with a high occurrence of hotels but low population.

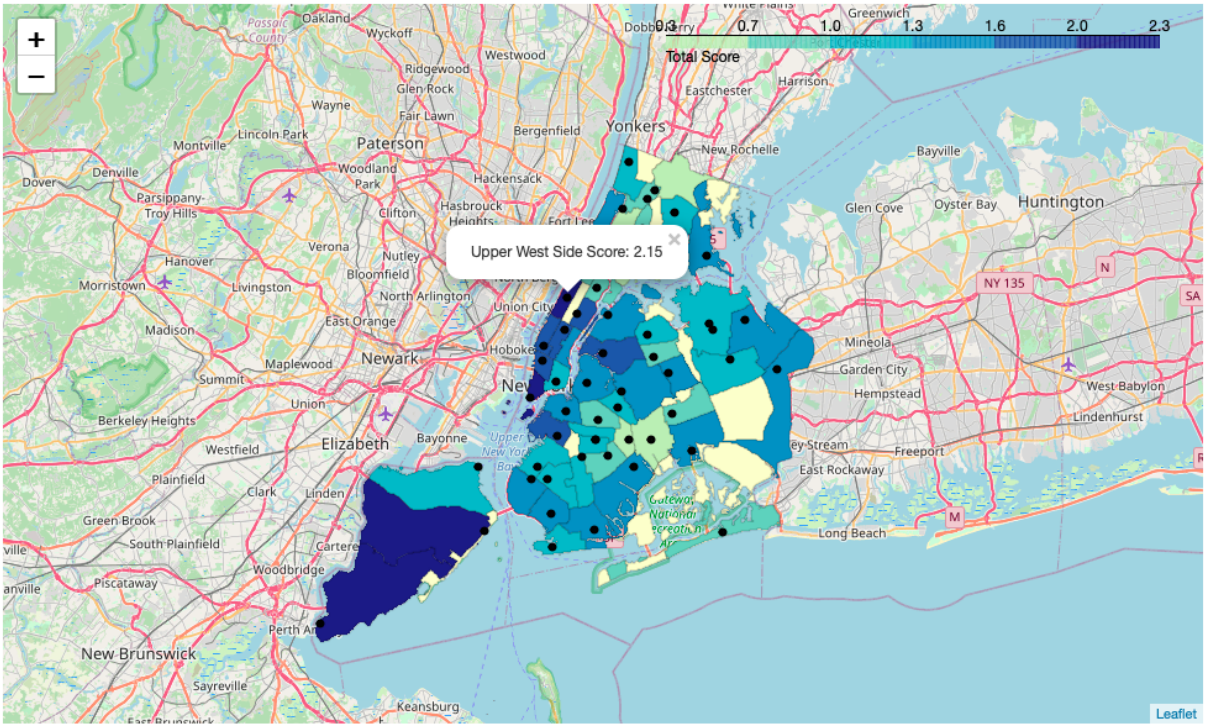
The **blue** Cluster 2 is the substandard faction, with a median total score of 0.67 and except of the population, trust and parks the mean values are very low.

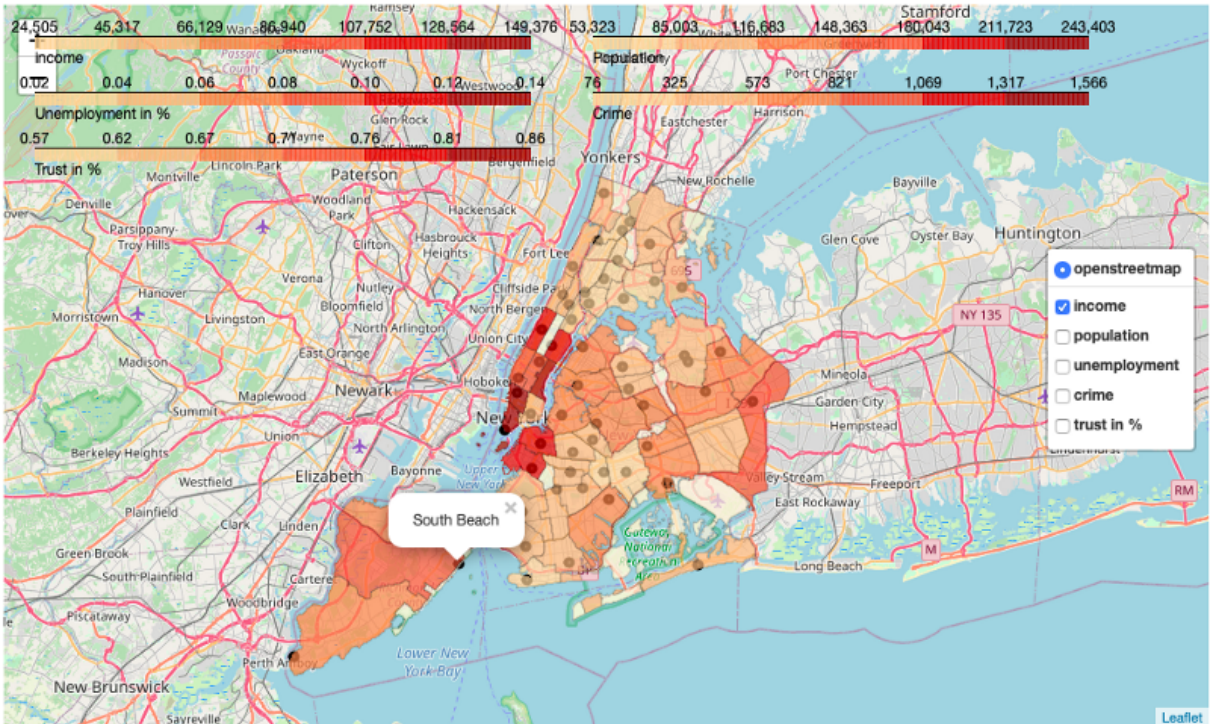
6. Evaluation

As you may see presenting the customer a clustered map is not a result, which is a good foundation for finding the perfect location of the Store. But the weighted heatmap is great to work with. We are going to explore this data frame further. Presenting a map with the total scores has much more information for the decision of the customer, combining this map with the choropleth map of the socio economic data is superb to visualize the data frames interactive.

Area	Population	Income	Unemployment in %	Felonies	Trust in %	Boutique	Clothing Store	Italian Restaurant	Hotel	Park	Theater	Art Gallery	Total Score
South Beach	0.46	0.49	-0.136	-0.074	0.667	-0	-0	0.35	0.36	0.14	0	0	2.26
Battery Park	0.21	0.8	-0.132	-0.065	0.57	-0	-0	0.04	0.56	0.23	0	0	2.2
Upper West Side	0.64	0.66	-0.182	-0.181	0.578	-0	-0	0.14	0	0.45	0.04	0	2.15
Tottenville	0.55	0.47	-0.1	-0.046	0.7	-0	-0	0.26	0	0.17	0	0	2
Upper East Side	0.68	0.66	-0.096	-0.129	0.505	-0.15	-0.048	0.14	0.11	0.23	0.02	0	1.92
Sunnyside	0.46	0.38	-0.154	-0.09	0.619	-0	-0	0.11	0.45	0.05	0	0.06	1.88
Greenwich Village	0.3	0.8	-0.132	-0.162	0.57	-0	-0.048	0.32	0	0.18	0	0	1.83
Midtown Business District	0.18	0.55	-0.189	-0.232	0.537	-0.3	-0.097	0	0.9	0.09	0.3	0	1.74
Chelsea	0.37	0.55	-0.189	-0.234	0.537	-0.3	-0	0.14	0.22	0.27	0.02	0.3	1.69
Park Slope	0.39	0.66	-0.196	-0.113	0.667	-0	-0	0.18	0	0.09	0	0	1.68
Bayside	0.39	0.47	-0.125	-0.05	0.7	-0	-0	0.11	0	0.05	0	0	1.53
Howard Beach	0.43	0.41	-0.2	-0.195	0.627	-0	-0.068	0.24	0.16	0.12	0	0	1.53
Ridgewood	0.62	0.41	-0.146	-0.142	0.627	-0	-0	0.11	0	0	0.04	0	1.52
Murray Hill	0.48	0.73	-0.096	-0.142	0.61	-0.15	-0	0	0	0.09	0	0	1.52
Fort Greene	0.49	0.59	-0.182	-0.292	0.627	-0	-0	0.21	0	0.05	0.02	0	1.51

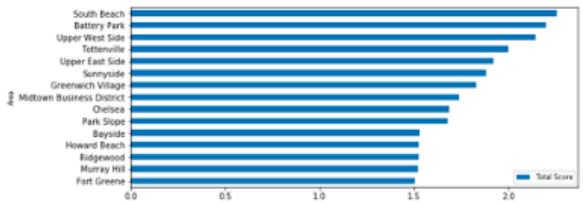
Visualize the total score

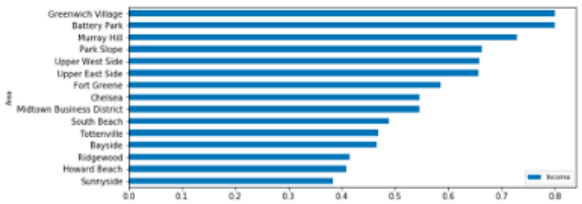


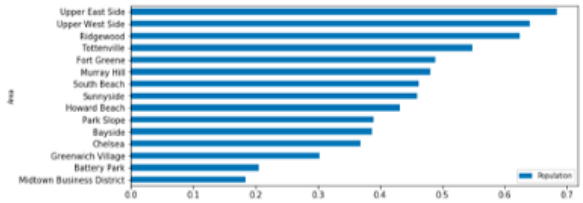


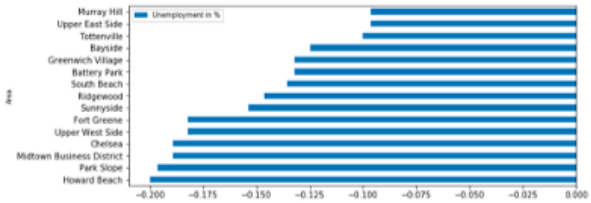
Bar plots of the top 15 districts

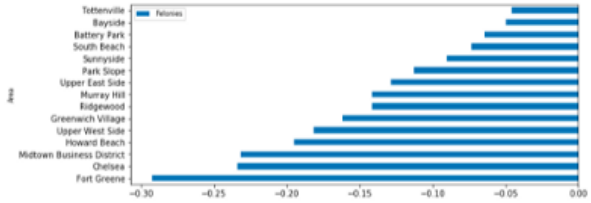
Supportive to the maps are the following bar plots of the sorted top 15 results, to get a contrasting view of the data.

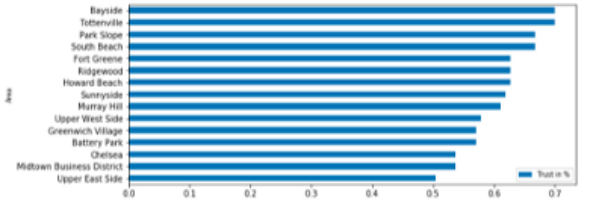

South Beach
Battery Park
Upper West Side
Tottenville
Upper East Side
Sunnyside
Greenwich Village
Midtown Business District
Chelsea
Park Slope
Bayside
Howard Beach
Ridgewood
Murray Hill
Fort Greene
Area
Total Score
0.0
0.5
1.0
1.5
2.0

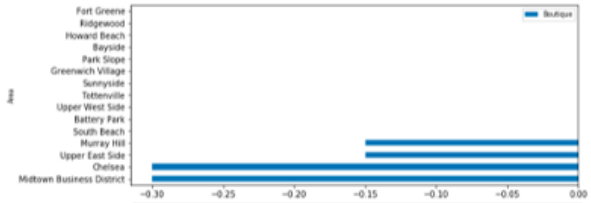

Greenwich Village
Battery Park
Murray Hill
Park Slope
Upper West Side
Upper East Side
Fort Greene
Chelsea
Midtown Business District
South Beach
Tottenville
Bayside
Ridgewood
Howard Beach
Sunnyside
Area
Income
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8

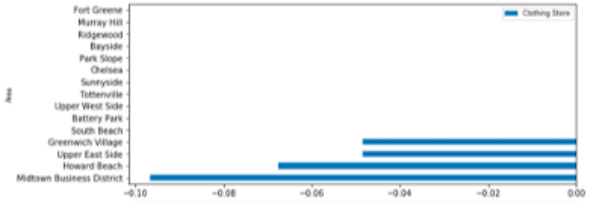

Upper East Side
Upper West Side
Ridgewood
Tottenville
Fort Greene
Murray Hill
South Beach
Sunnyside
Howard Beach
Park Slope
Bayside
Chelsea
Greenwich Village
Battery Park
Midtown Business District
Area
Population
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7

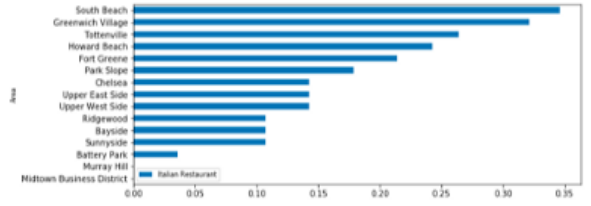

Murray Hill
Upper East Side
Tottenville
Bayside
Greenwich Village
Battery Park
South Beach
Ridgewood
Sunnyside
Fort Greene
Upper West Side
Chelsea
Midtown Business District
Park Slope
Howard Beach
Area
Unemployment in %
−0.200
−0.175
−0.150
−0.125
−0.100
−0.075
−0.050
−0.025
0.000

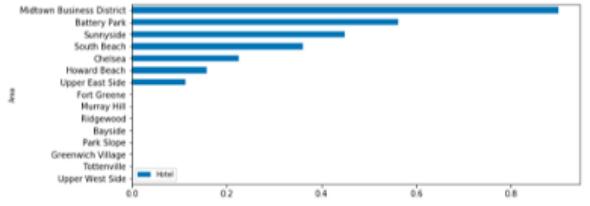

Tottenville
Bayside
Battery Park
South Beach
Sunnyside
Park Slope
Upper East Side
Murray Hill
Ridgewood
Greenwich Village
Upper West Side
Howard Beach
Midtown Business District
Chelsea
Fort Greene
Area
Felonies
−0.30
−0.25
−0.20
−0.15
−0.10
−0.05
0.00

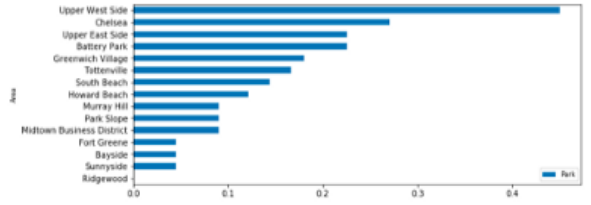

Bayside
Tottenville
Park Slope
South Beach
Fort Greene
Ridgewood
Howard Beach
Sunnyside
Murray Hill
Upper West Side
Greenwich Village
Battery Park
Chelsea
Midtown Business District
Upper East Side
Area
Trust in %
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7

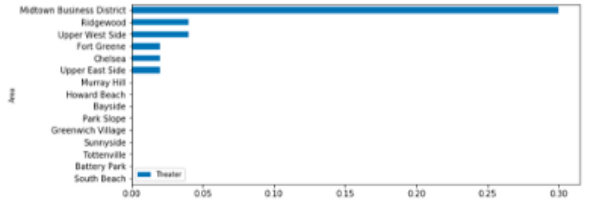

Fort Greene
Ridgewood
Howard Beach
Bayside
Park Slope
Greenwich Village
Sunnyside
Tottenville
Upper West Side
Battery Park
South Beach
Murray Hill
Upper East Side
Chelsea
Midtown Business District
Area
Boutique
−0.30
−0.25
−0.20
−0.15
−0.10
−0.05
0.00

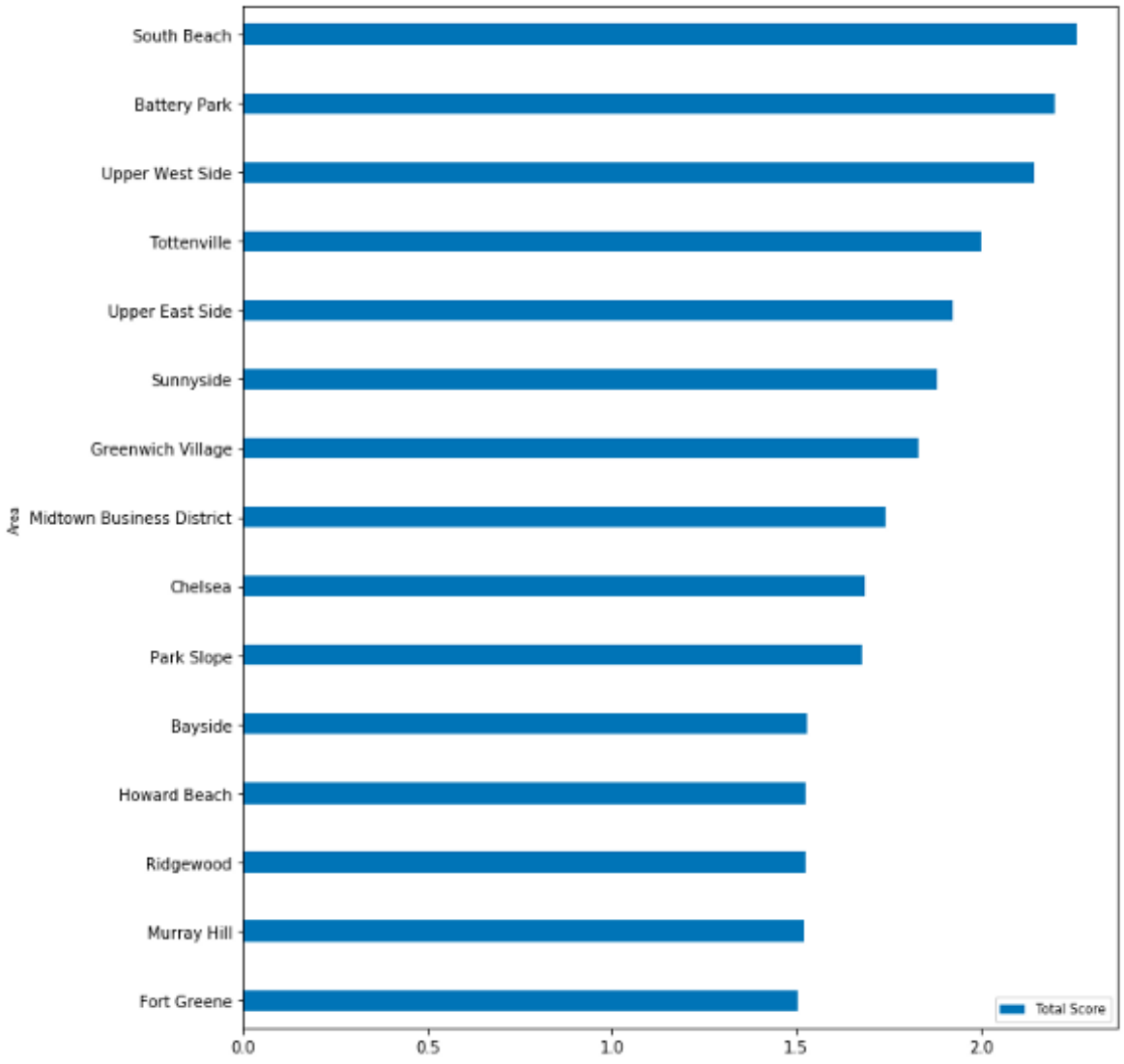

Fort Greene
Murray Hill
Ridgewood
Bayside
Park Slope
Chelsea
Sunnyside
Tottenville
Upper West Side
Battery Park
South Beach
Greenwich Village
Upper East Side
Howard Beach
Midtown Business District
Area
Clothing Store
−0.10
−0.08
−0.06
−0.04
−0.02
0.00

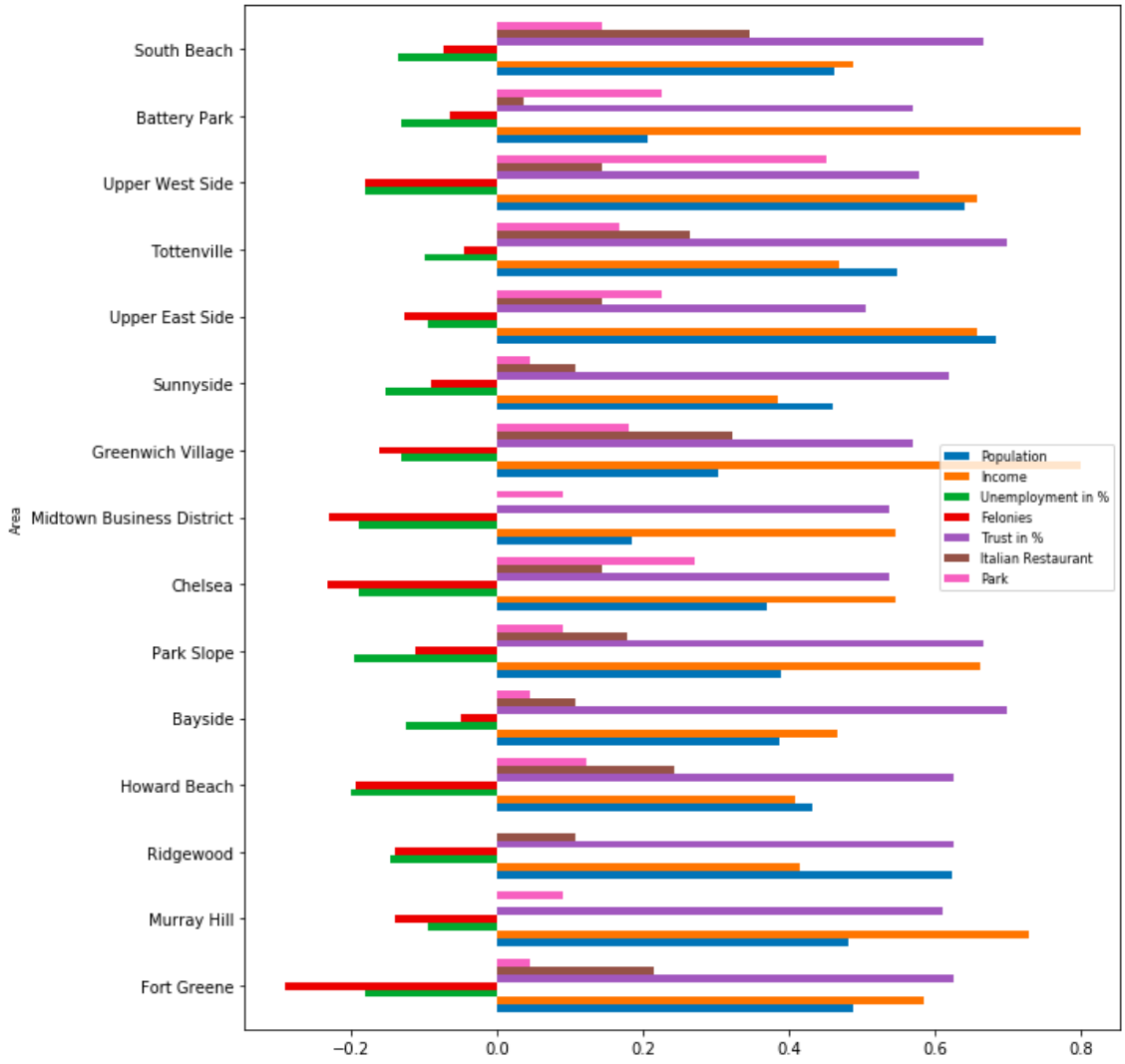

South Beach
Greenwich Village
Tottenville
Howard Beach
Fort Greene
Park Slope
Chelsea
Upper East Side
Upper West Side
Ridgewood
Bayside
Sunnyside
Battery Park
Murray Hill
Midtown Business District
Area
Italian Restaurant
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35


Midtown Business District
Battery Park
Sunnyside
South Beach
Chelsea
Howard Beach
Upper East Side
Fort Greene
Murray Hill
Ridgewood
Bayside
Park Slope
Greenwich Village
Tottenville
Upper West Side
Area
Hotel
0.0
0.2
0.4
0.6
0.8


Upper West Side
Chelsea
Upper East Side
Battery Park
Greenwich Village
Tottenville
South Beach
Howard Beach
Murray Hill
Park Slope
Midtown Business District
Fort Greene
Bayside
Sunnyside
Ridgewood
Area
Park
0.0
0.1
0.2
0.3
0.4


Midtown Business District
Ridgewood
Upper West Side
Fort Greene
Chelsea
Upper East Side
Murray Hill
Howard Beach
Bayside
Park Slope
Greenwich Village
Sunnyside
Tottenville
Battery Park
South Beach
Area
Theater
0.00
0.05
0.10
0.15
0.20
0.25
0.30


South Beach
Battery Park
Upper West Side
Tottenville
Upper East Side
Sunnyside
Greenwich Village
Midtown Business District
Chelsea
Park Slope
Bayside
Howard Beach
Ridgewood
Murray Hill
Fort Greene
Area
Total Score
0.0
0.5
1.0
1.5
2.0



7. Discussion of the Result

Our analysis shows that there are several promising districts for the store. Especially ***South Beach***, ***Upper West Side*** and ***Battery Park*** are high rated. The Distribution of Clothing Stores was the highest in St.George and of Boutiques in Midtown Business District.

As you can see from the map the Cluster 0 (***red***) is the medium cluster for the requirements of the Customer, it is a pretty big cluster and includes some of the best scored districts.
The ***purple*** Cluster is mostly located in Manhatten and consist of only 2 high ranked districts.
The ***blue*** cluster should be ignored.

South Beach located in Staten Island gained the highest score. There is a high frequency of Italian restaurants and the factor that it is a good place to live with a low Crime Rate compensates the medium socio economic data. Choosing this location could mean that the Store will profit from regular customers but there won't be as many tourists and walk in customers as in Manhatten.

Upper Westside scores with high income, population and parks but there are no hotels directly in the district, which could lead to less touristic customers. On the oter hand the Central Park is close by, which is a touristic hotspot. But probably more touristic than South Beach. On the other hand this is a place where a lot of wealthy people live and the store could benefit from

regular customers. It could be a great place to live if the customer prefers to live right in the city Center. The proximity to the Central Park a touristic hot spot could maybe compensate the lack of hotels in the relation of touristic customers.

The ***Battery Park*** is a touristic hotspot in New York even though the low population it is in the top 3 districts and got the highest income score. The few People who can afford to live in the top of Manhatten have a high income furthermore there are lots of hotels located in and around the area which guarantees for a great mix of tourists and regular customers.

Tottenville the 4. place is in Staten Island too and has the lowest crime and the lowest unemployment rate of the top 15. The trust score is also one of the highest. Furthermore it has a high overall score and is pretty similar to South Beach.

There is one main decision to make:
Manhatten or Staten Island

8. Conclusion

Purpose of this project was to identify districts which fits best to the diverse requirements of the customer. By evaluating and quantifying his imaginations with the weighted matrix it was possible to identify several districts which combines his requirements for the location of the store and personal living wishes.

For finding the perfect location we now have to go deeper and analyse the top 10 to 15 districts more detailed. We could compare specific neighbourhoods and add more detailed data like tourism frequency to finally find the perfect neighbourhood or even the best street for the store.

If you don´t use the free Foursquare version the quality oft he venue data will rise too.