# Capstone Project - The Battle of Neighborhoods

*Safety in New York City*

*Patrícia Falcão, February 2020*

# Introduction

- Average American moves once every 5 years;

- New York City (NYC) is the most populous city in united states, however, is not the safest one;

- Safety become one of the most important when choosing a place to live;

- NYC is a big city with so many opportunities, it is useful to know which neighbourhoods are the safest to live in when looking for a home

- Goal: find the safest borough in NYC based on the crimes committed between 2014 and 2015. Also, it is set as goal to select the 5 most common venues (of the safest borough) in each neighbourhood and cluster them using k-mean clustering.

# Data

Three diferente data sources:

- A dataset from Kaggle, it was possible to access to the crimes committed in NYC between 2014 and 2015;

- Wikipedia page that had the list of NYC boroughs;

- Wikipedia page referred to the list of neighbourhoods of the safest borough

# Methodology: Data Cleaning

- In the first data source (NYC crime data), it was only considered the crimes from 2015 (most recent). It was calculated the total crimes per borough;

- Using the second data source, it was used the Beautiful Soup library in order to extract the data from the web page. Tables are merged in one dataset. This way it is possible to know the borough with the least and highest crimes recorded in 2015;

- Knowing the safest borough, it used the last source. It is then created a new dataframe using the name of each neighbourhood. Then, the coordinates of each latitude and longitude are obtained using the Google Maps API Geocoding, resulting in Figure 3. However, since it was not possible to use this google tool to locate all the neighbourhoods (it was a big number), it was only considered 27.

# Methodology: Exploratory Data Analysis

- Apply the describe function to the crime data:

| crimes | ABORTION | ADMINISTRATIVE CODE | ADMINISTRATIVE CODES | AGRICULTURE & MRKTS LAW-UNCLASSIFIED | ALCOHOLIC BEVERAGE CONTROL LAW | ANTICIPATORY OFFENSES | ARSON | ASSAULT 3 & RELATED OFFENSES | BURGLAR'S TOOLS | BURGLARY | ... | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 5.000000 | 5.00000 | 5.000000 | 5.000000 | 5.000000 | 5.000000 | 5.000000 | 5.000000 | 5.000000 | 5.000000 | ... | 5.000000 |
| mean | 0.400000 | 210.60000 | 0.400000 | 16.600000 | 15.600000 | 0.400000 | 201.600000 | 10386.400000 | 57.200000 | 3000.000000 | ... | 93480.600000 |
| std | 0.894427 | 145.58434 | 0.547723 | 8.443933 | 14.604794 | 0.547723 | 113.279742 | 5175.539373 | 40.145984 | 1777.280929 | ... | 43854.210776 |
| min | 0.000000 | 34.00000 | 0.000000 | 6.000000 | 0.000000 | 0.000000 | 46.000000 | 2189.000000 | 6.000000 | 560.000000 | ... | 21678.000000 |
| 25% | 0.000000 | 112.00000 | 0.000000 | 13.000000 | 2.000000 | 0.000000 | 146.000000 | 9993.000000 | 33.000000 | 2662.000000 | ... | 92732.000000 |
| 50% | 0.000000 | 212.00000 | 0.000000 | 16.000000 | 20.000000 | 0.000000 | 195.000000 | 10507.000000 | 64.000000 | 2743.000000 | ... | 102712.000000 |
| 75% | 0.000000 | 291.00000 | 1.000000 | 19.000000 | 21.000000 | 1.000000 | 299.000000 | 13224.000000 | 71.000000 | 3544.000000 | ... | 110196.000000 |
| max | 2.000000 | 404.00000 | 1.000000 | 29.000000 | 35.000000 | 1.000000 | 322.000000 | 16019.000000 | 112.000000 | 5491.000000 | ... | 140085.000000 |

- The highest ('Petit Larceny') and the lowest ('Fortune Telling', 'Kidnapping', 'Kidnapping and related offenses', 'Loitering', 'Under the influence of drugs') reported crime.
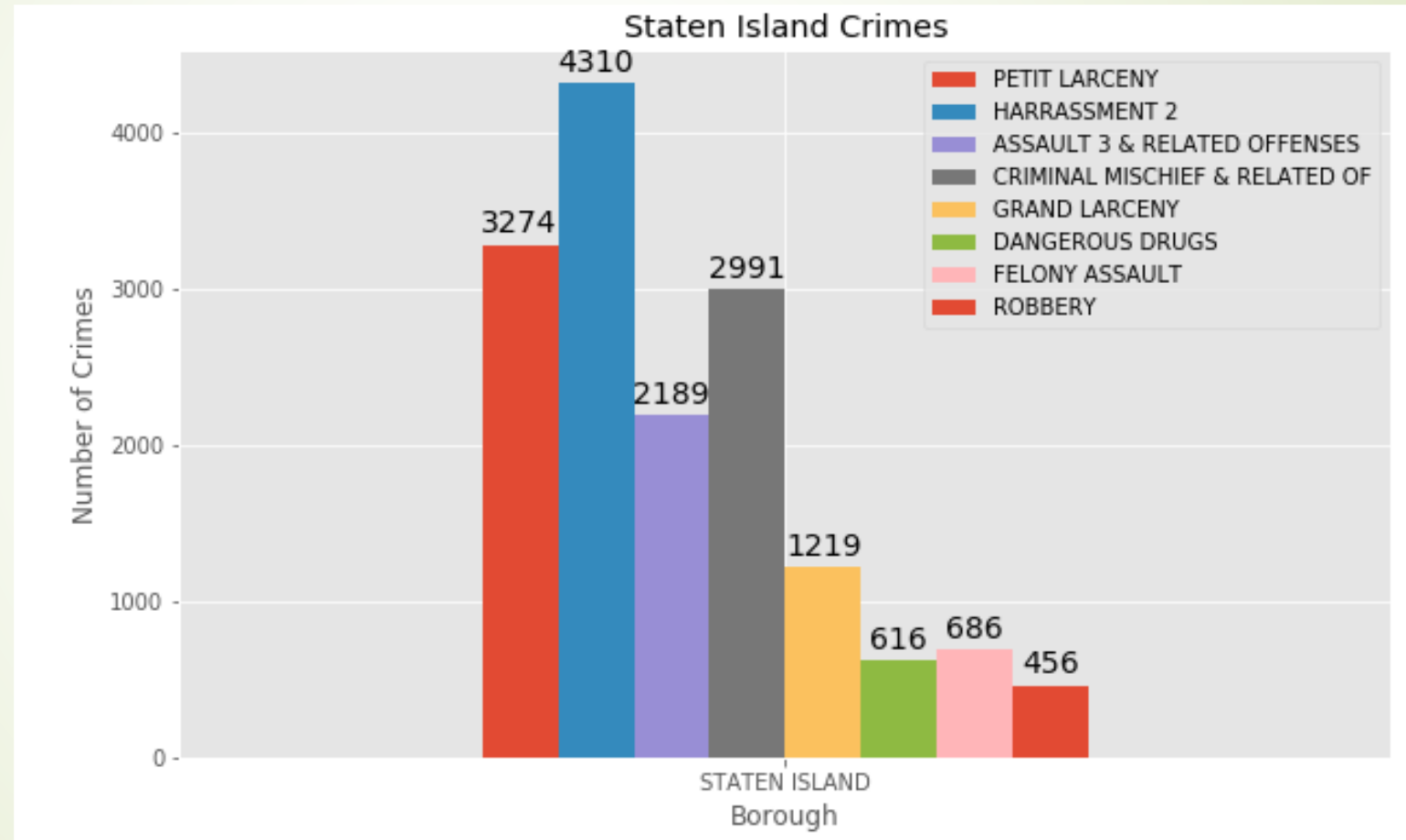
- Crime rate in NYC Boroughs:



Crime rate in NYC Boroughs
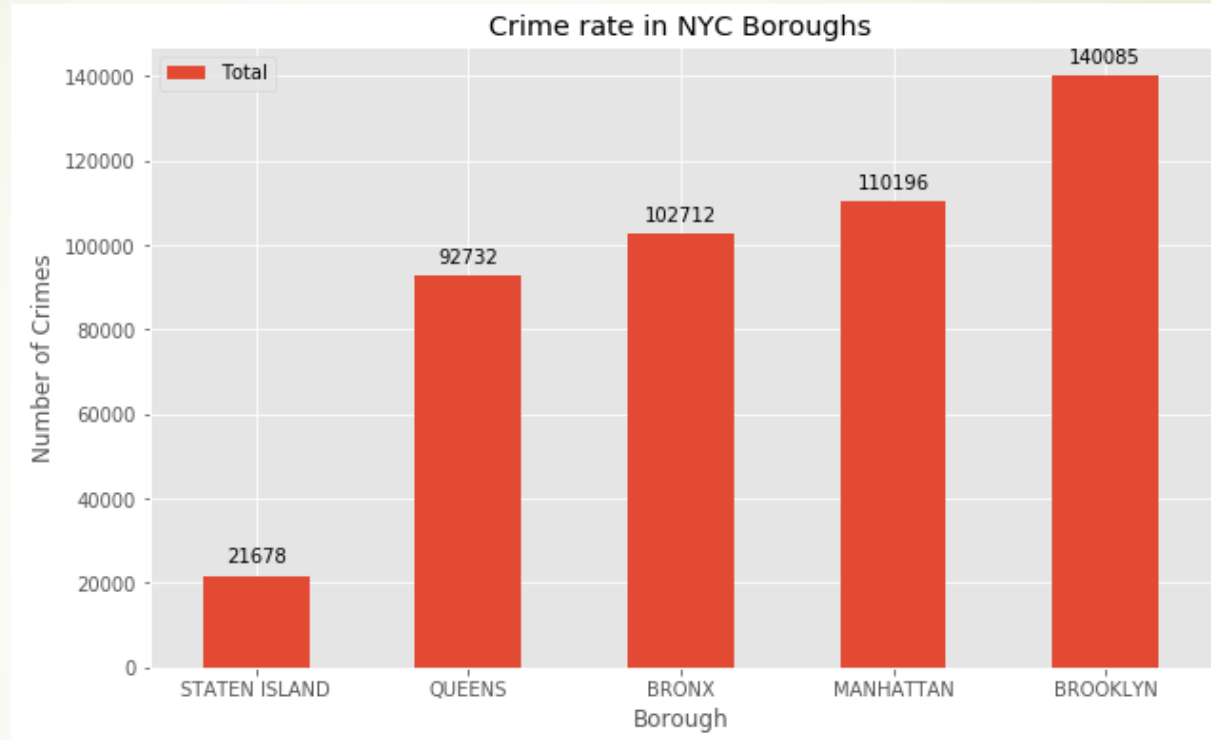
- it is possible to conclude that Staten Island is the borough of NYC with the lowest crime rate in 2015, with 21678 crimes.

- Staten Island Crimes: It is plotted 8 types of crimes committed in Staten Island:

- It is used a Wikipedia page referring to the neighbourhoods of Staten Island. However, since it was not possible to geolocate every neighbourhood with Google Maps API Geocoding, it was considered only 26 of them. It is represented in the next map:

# Modelling

- Using Foursquare API it is possible to find venues 500m away from each neighbourhood of Staten Island considered. Then, the resulting json file is converted to a dataframe, containing all the venues with the coordinates and category;

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | Annadale | 40.54455 | -74.176532 | Annadale Terrace | 40.542555 | -74.177187 | Restaurant |
| 1 | Annadale | 40.54455 | -74.176532 | Annadale Diner | 40.542079 | -74.177325 | Diner |
| 2 | Annadale | 40.54455 | -74.176532 | Play Sports Bar | 40.540418 | -74.177196 | Sports Bar |
| 3 | Annadale | 40.54455 | -74.176532 | Il Sogno | 40.541286 | -74.178489 | Restaurant |
| 4 | Annadale | 40.54455 | -74.176532 | Harbor View Restoration LLC | 40.544934 | -74.174709 | Construction & Landscaping |

- It is performed one hot encoding technique on the venues data. The venues are grouped by the neighbourhood and the means are calculated;

- The 5 common venues are calculated for each of the neighbourhoods;

- The final goal of this project is to cluster similar neighbourhoods using 5 – means clustering, so people can find similar neighbourhoods, i.e., regions that have similar characteristics and amenities. As mentioned, the 5 – means clustering will then cluster the neighbourhoods into 5 clusters.

# Results

- After applying 5-means clustering, it is possible to access to each one of the 5 clusters and know the neighbourhood that are associated to each of the clusters.

- The second, thirst ant fifth cluster have only neighbourhood each, what means that are no neighbourhood similar to them. In the second one, the venues consist of train station, Italian restaurant and movie theatre, yoga studio and flower shop.

| | Neighborhood | Borough | Latitude | Longitude | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|
| 25 | Huguenot | STATEN ISLAND | 40.537328 | -74.194588 | 1 | Train Station | Italian Restaurant | Movie Theater | Yoga Studio | Flower Shop |

- The first cluster is the biggest one with 18 of the 27 neighbourhoods. It can be deonoted that the most common venue in these neighbourhoods are restaurants (Italian, American, Indian, …), Pizza place and Bagel Shop.

| | Neighborhood | Borough | Latitude | Longitude | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Annadale | STATEN ISLAND | 40.544550 | -74.176532 | 0 | American Restaurant | Pizza Place | Restaurant | Liquor Store | Sports Bar |
| 1 | Arlington | STATEN ISLAND | 41.694069 | -73.886363 | 0 | Automotive Shop | Convenience Store | Event Space | Mobile Phone Shop | Trail |
| 2 | Arrochar | STATEN ISLAND | 40.598438 | -74.072641 | 0 | Pizza Place | Park | Deli / Bodega | Bagel Shop | Bus Stop |
| 3 | Bay Terrace | STATEN ISLAND | 40.555278 | -74.134167 | 0 | Italian Restaurant | American Restaurant | Miscellaneous Shop | Playground | Food Truck |
| 4 | Bloomfield | STATEN ISLAND | 42.895064 | -77.434713 | 0 | Ice Cream Shop | American Restaurant | Pizza Place | Lawyer | Flower Shop |
| 5 | Bulls Head | STATEN ISLAND | 40.607048 | -74.162088 | 0 | Bus Stop | Pharmacy | Diner | Baseball Field | Tattoo Parlor |
| 6 | Castleton Corners | STATEN ISLAND | 40.613160 | -74.122365 | 0 | Pizza Place | Bank | Ice Cream Shop | Convenience Store | Salon / Barbershop |
| 7 | Charleston | STATEN ISLAND | 40.536772 | -74.237367 | 0 | Construction & Landscaping | Bakery | Gym / Fitness Center | Deli / Bodega | Rental Car Location |
| 8 | Chelsea | STATEN ISLAND | 40.746491 | -74.001528 | 0 | Café | French Restaurant | Sushi Restaurant | Coffee Shop | Indian Restaurant |
| 12 | Egbertville | STATEN ISLAND | 40.578622 | -74.131570 | 0 | Construction & Landscaping | Italian Restaurant | Bagel Shop | Trail | Cosmetics Shop |

- The venues in the third cluster comprehends an indie theather, golf course, trail, café and yoga studio (Figure 11).

| | Neighborhood | Borough | Latitude | Longitude | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|
| 22 | Grymes Hill | STATEN ISLAND | 40.618715 | -74.093475 | 2 | Indie Theater | Golf Course | Trail | Café | Yoga Studio |

- In turn, some of the venues in the fifth cluster are pizza place, Asian restaurant and flower shop.

| | Neighborhood | Borough | Latitude | Longitude | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|
| 11 | Dongan Hills | STATEN ISLAND | 40.597927 | -74.098027 | 4 | Pizza Place | Asian Restaurant | Flower Shop | Yoga Studio | Food |

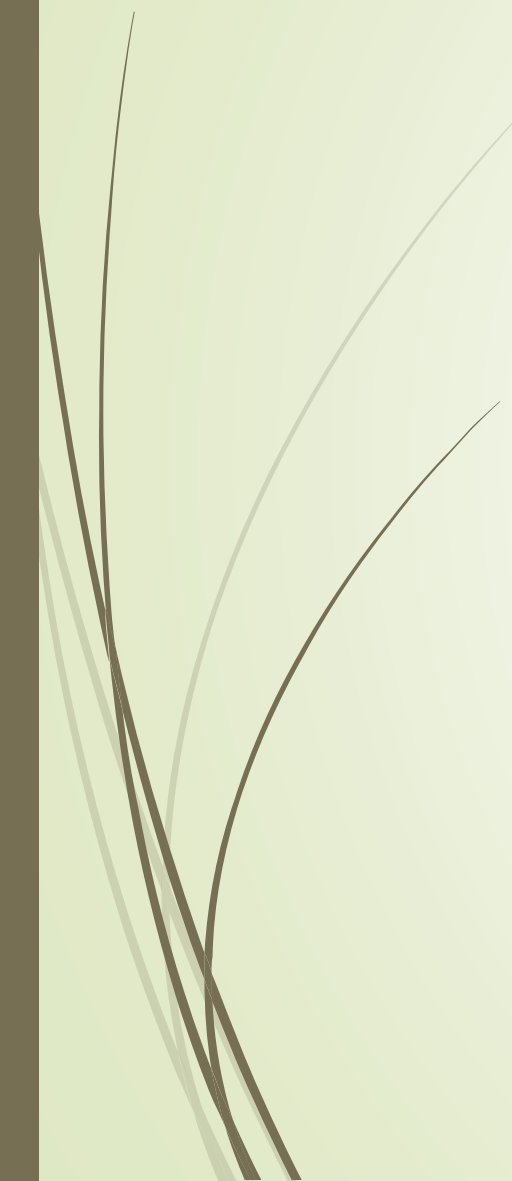- The last one, the fourth cluste, has 5 neighbourhoods that have common venues such as bus stop and deli/bodega.

| | Neighborhood | Borough | Latitude | Longitude | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|
| 9 | Clifton | STATEN ISLAND | 40.620104 | -74.077086 | 3 | Bus Stop | Grocery Store | Intersection | Deli / Bodega | Train Station |
| 13 | Elm Park | STATEN ISLAND | 40.631493 | -74.148754 | 3 | Deli / Bodega | Italian Restaurant | Bus Stop | Athletics & Sports | Furniture / Home Store |
| 15 | Emerson Hill | STATEN ISLAND | 40.608585 | -74.094564 | 3 | Bus Stop | Acupuncturist | Deli / Bodega | Intersection | Automotive Shop |
| 23 | Hamilton Park | STATEN ISLAND | 40.641726 | -74.090072 | 3 | Bus Stop | Park | Bowling Alley | Mexican Restaurant | Deli / Bodega |
| 24 | Heartland Village | STATEN ISLAND | 40.588333 | -74.157778 | 3 | Gym / Fitness Center | Sandwich Place | Bus Stop | Bus Station | Yoga Studio |

- In the next figure is represented the five clusters in the map, in order to understand its geographical localisation. Each cluster is associated to a colour (red, blue, green, orange and purple).
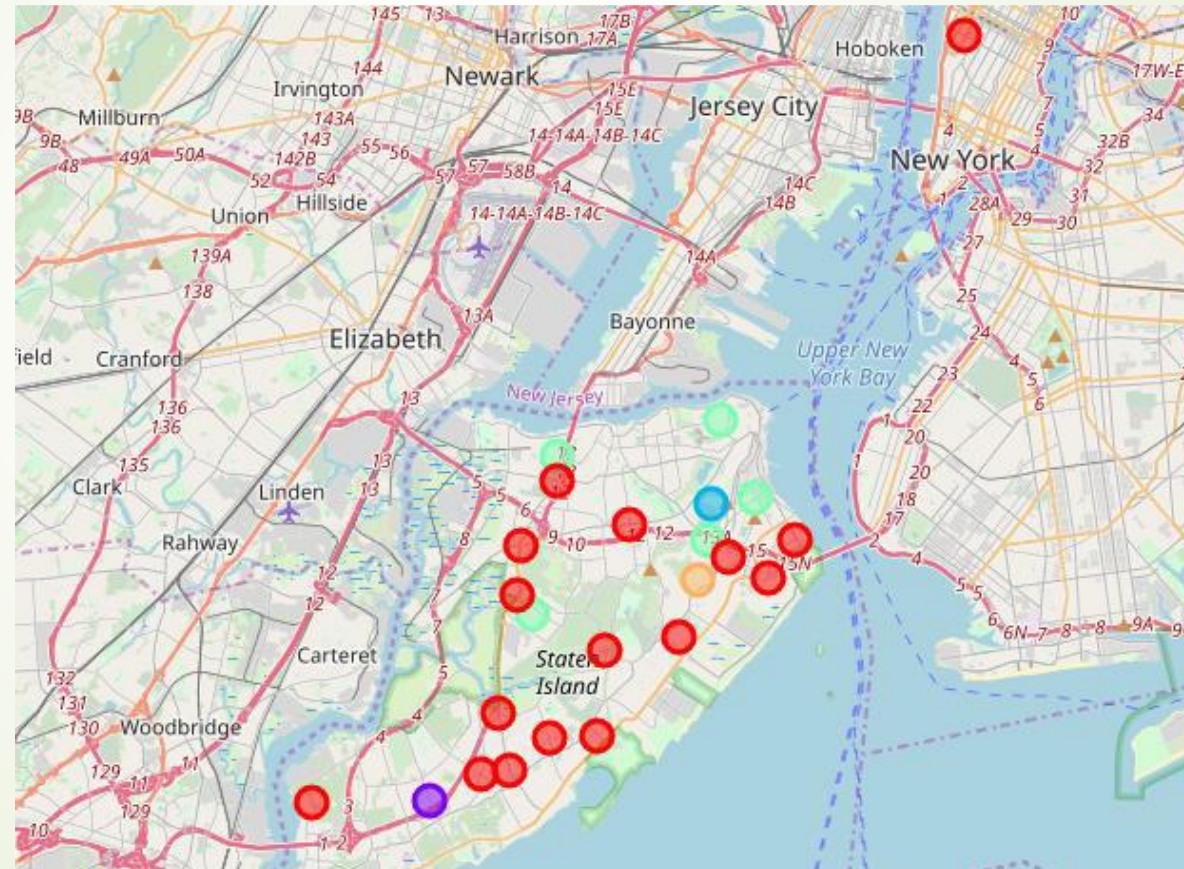
# Discussion

- This project has as aim to help people to find the safest borough in New York City based on the crimes committed between 2014 and 2015;

- Based in a cluster technique, also allow people to find neighbourhoods with similarities. This way, as soon as someone want to move to NYC or to relocate inside the city, if one of the wishes is to find a safer borough, the answer is staten island;

- However, inside this borough, there are a lot of neighbourhoods. So, if someone really want a place that has indie theather, golf course, trail, café and yoga studio, maybe the best neighbourhood is Grymes Hill (third cluster);

- Conversely, if it desires to have a train station, some Italian restaurant and movie theatre, yoga studio and flower shop, then the best neighbourhood is Huguenot (second cluster). However, the final choice may vary depending on the interests of each person.

# Conclusion

- This works is a big help for people that has as intention to relocate or to move to the safest borough of NYC (This was my criteria, but it could be other one), as well as, have the notion of the similar neighbourhoods of that borough;

- So, this type of technology can be a decision support, make it easy and more reliable.

# Thank you!