# Setting up a new restaurant in New York City

Tiago Ribeiro

July 2020

# Introduction

- The Business Problem that I wish to help solve is basically in which neighborhood of New York city should a cook/business man build his new restaurant, given the fact that Manhattan is already a place filled with many many restaurants of several cultures.
- "you can't walk a New City block without passing a restaurant". It even states that "80 percent of restaurants fail within five years" (https://vinepair.com/booze-news/new-york-restaurants-eat-at-every-on/)

# The Data

- Foursquare API will be the chosen API to collect the data related to the venues for each geographical point.
- The Source data for the NY City locations should have the following format (after interpreting the JSON file):

| | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|
| 0 | Bronx | Wakefield | 40.894705 | -73.847201 |
| 1 | Bronx | Co-op City | 40.874294 | -73.829939 |
| 2 | Bronx | Eastchester | 40.887556 | -73.827806 |
| 3 | Bronx | Fieldston | 40.895437 | -73.905643 |
| 4 | Bronx | Riverdale | 40.890834 | -73.912585 |

# The Data

- The data for the venues to be analyzed should have the following format:

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | Marble Hill | 40.876551 | -73.91066 | Arturo's | 40.874412 | -73.910271 | Pizza Place |
| 1 | Marble Hill | 40.876551 | -73.91066 | Bikram Yoga | 40.876844 | -73.906204 | Yoga Studio |
| 2 | Marble Hill | 40.876551 | -73.91066 | Tibbett Diner | 40.880404 | -73.908937 | Diner |
| 3 | Marble Hill | 40.876551 | -73.91066 | Starbucks | 40.877531 | -73.905582 | Coffee Shop |
| 4 | Marble Hill | 40.876551 | -73.91066 | Dunkin' | 40.877136 | -73.906666 | Donut Shop |

# Methodology

- First and foremost, the venues of each Manhattan location described in the Data section of this report, were retrieved from Foursquare's API, with the output being a dataframe with following structure:

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | Marble Hill | 40.876551 | -73.91066 | Arturo's | 40.874412 | -73.910271 | Pizza Place |
| 1 | Marble Hill | 40.876551 | -73.91066 | Bikram Yoga | 40.876844 | -73.906204 | Yoga Studio |
| 2 | Marble Hill | 40.876551 | -73.91066 | Tibbett Diner | 40.880404 | -73.908937 | Diner |
| 3 | Marble Hill | 40.876551 | -73.91066 | Starbucks | 40.877531 | -73.905582 | Coffee Shop |
| 4 | Marble Hill | 40.876551 | -73.91066 | Astral Fitness & Wellness Center | 40.876705 | -73.906372 | Gym |

# Methodology

- Then the one-hot coding procedure was done in order to get a matrix with the neighborhoods on the left side and the column names being the type of venue, with 1s marking if the current venue exists for a given location and 0s marking the opposite:

| | Neighborhood | Accessories Store | Adult Boutique | Afghan Restaurant | African Restaurant | American Restaurant | Antique Shop | Arcade | Arepa Restaurant | Argentinian Restaurant |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Marble Hill | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | Marble Hill | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | Marble Hill | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | Marble Hill | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | Marble Hill | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

# Methodology

- Then the data would be structured in the form that will be essential for this case study and will be used for Machine Learning (ML) algorithms:

| | Neighborhood | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Battery Park City | Park | Hotel | Coffee Shop | Gym | Memorial Site | Playground | Plaza | Shopping Mall | Gourmet Shop | BBQ Joint |
| 1 | Carnegie Hill | Coffee Shop | Café | Pizza Place | Yoga Studio | Bookstore | Bakery | French Restaurant | Japanese Restaurant | Italian Restaurant | Bar |
| 2 | Central Harlem | African Restaurant | Seafood Restaurant | Bar | American Restaurant | Chinese Restaurant | French Restaurant | Cosmetics Shop | Market | Caribbean Restaurant | Library |
| 3 | Chelsea | Coffee Shop | Art Gallery | American Restaurant | Italian Restaurant | Bakery | Seafood Restaurant | French Restaurant | Market | Hotel | Pizza Place |
| 4 | Chinatown | Chinese Restaurant | Cocktail Bar | Bakery | Salon / Barbershop | Vietnamese Restaurant | Spa | Bubble Tea Shop | Hotpot Restaurant | Ice Cream Shop | Dessert Shop |

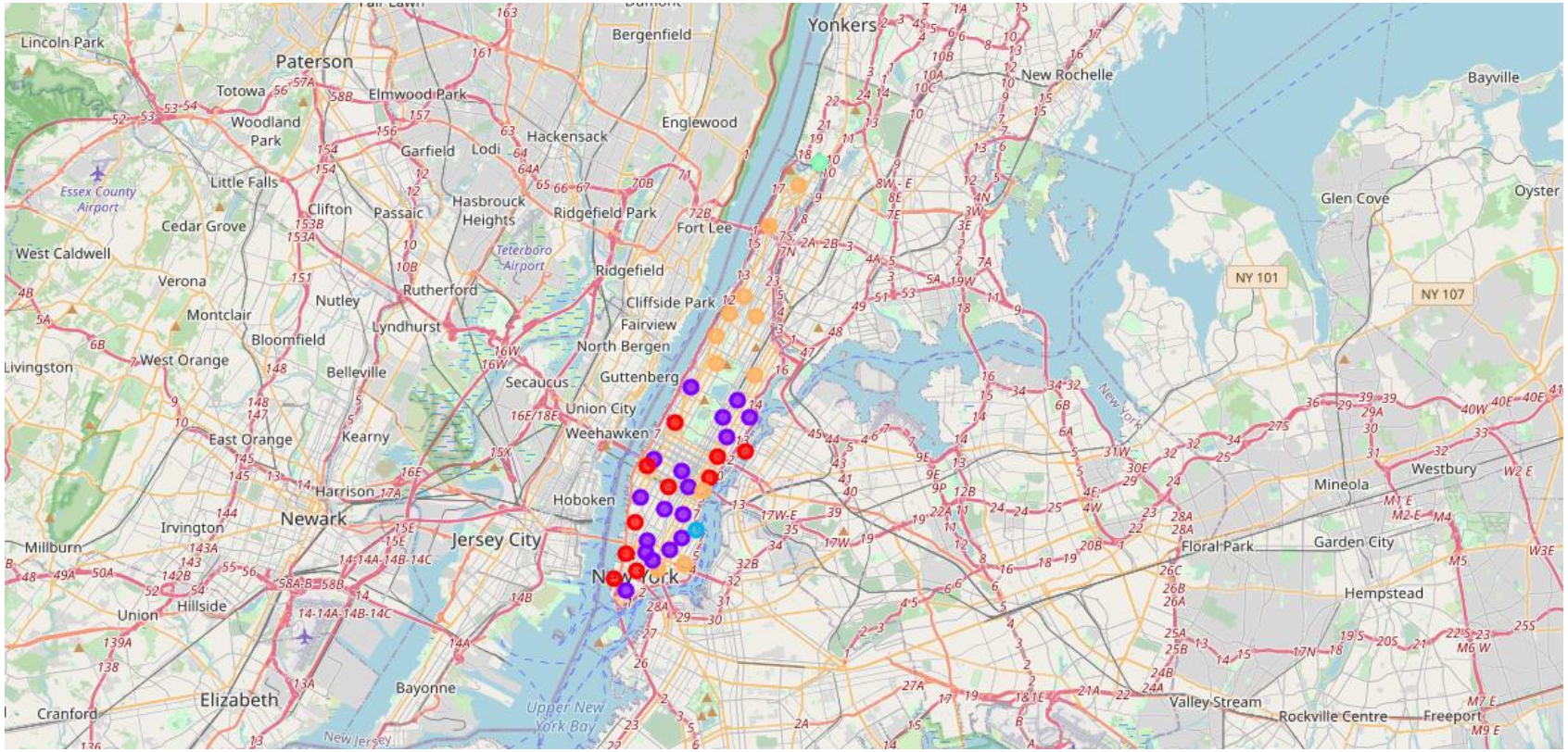
# Methodology

- After using clustering to allow us to label our data:

| | Borough | Neighborhood | Latitude | Longitude | Cluster Labels | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Manhattan | Marble Hill | 40.876551 | -73.910660 | 3 | Gym | Coffee Shop | Yoga Studio | Diner | Seafood Restaurant | Sandwich Place | Supplement Shop | Tennis Stadium | Donut Shop | Shopping Mall |
| 1 | Manhattan | Chinatown | 40.715618 | -73.994279 | 4 | Chinese Restaurant | Cocktail Bar | Bakery | Salon / Barbershop | Vietnamese Restaurant | Spa | Bubble Tea Shop | Hotpot Restaurant | Ice Cream Shop | Dessert Shop |
| 2 | Manhattan | Washington Heights | 40.851903 | -73.936900 | 4 | Café | Bakery | Mobile Phone Shop | Chinese Restaurant | Seafood Restaurant | Bank | Tapas Restaurant | Mexican Restaurant | Coffee Shop | Italian Restaurant |
| 3 | Manhattan | Inwood | 40.867684 | -73.921210 | 4 | Mexican Restaurant | Lounge | Restaurant | Café | Frozen Yogurt Shop | Bakery | Spanish Restaurant | Chinese Restaurant | Caribbean Restaurant | American Restaurant |
| 4 | Manhattan | Hamilton Heights | 40.823604 | -73.949688 | 4 | Pizza Place | Coffee Shop | Café | Deli / Bodega | Mexican Restaurant | Bakery | Park | Cocktail Bar | Sandwich Place | Chinese Restaurant |

# Methodology

K = 5 Clusters



# Methodology – Using a new test entry to obtain the candidate cluster

- 8 – Park
- 11 – Clothing Store
- 11 – Plaza
- 11 – Coffee Shop
- 12 – American Restaurant
- 1 – Spa
- 20 – Dog Run
- 7 – Liquor Store
- 0 – Donut Shop
- 20 - Thai Restaurant

# Results

| ML Techniques Accuracy Scores | | |
|---|---|---|
| Algorithm | Jaccard | F1-score |
| KNN | 0.625 | 0.643 |
| Decision Tree | 0.375 | 0.383 |
| SVM | 0.625 | 0.625 |
| Logistic Regression | 0.625 | 0.611 |

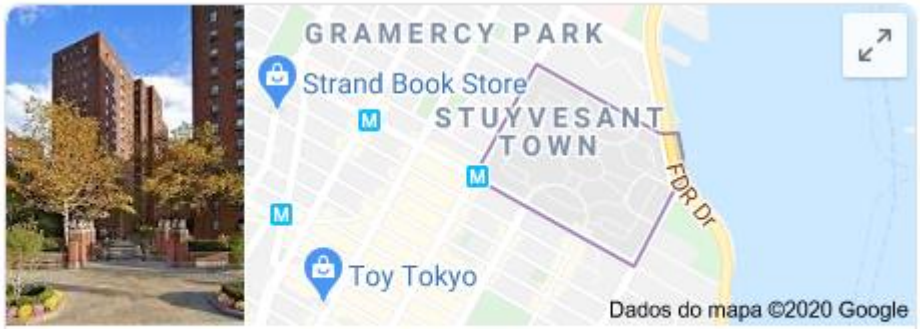
KNN was selected to check the label for our candidate cluster, thus cluster 1 was obtained!

# Results

| | Neighborhood | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 11 | Roosevelt Island | Park | Hotel | School | Gym | Coffee Shop | Greek Restaurant | Dry Cleaner | Sandwich Place | Liquor Store | Noodle House |
| 13 | Lincoln Square | Café | Gym / Fitness Center | Plaza | Theater | Concert Hall | Performing Arts Venue | Wine Shop | American Restaurant | Italian Restaurant | Coffee Shop |
| 21 | Tribeca | Park | Italian Restaurant | Café | American Restaurant | Wine Bar | Spa | Coffee Shop | Skate Park | Hotel | Greek Restaurant |
| 24 | West Village | Italian Restaurant | New American Restaurant | American Restaurant | Park | Cocktail Bar | French Restaurant | Jazz Club | Coffee Shop | Wine Bar | Theater |
| 28 | Battery Park City | Park | Hotel | Coffee Shop | Gym | Memorial Site | Playground | Plaza | Shopping Mall | Gourmet Shop | BBQ Joint |
| 32 | Civic Center | Coffee Shop | Cocktail Bar | Hotel | Gym / Fitness Center | Spa | Yoga Studio | Café | French Restaurant | Italian Restaurant | Bakery |
| 33 | Midtown South | Korean Restaurant | Hotel | Japanese Restaurant | Burger Joint | Cosmetics Shop | Gym / Fitness Center | Clothing Store | Coffee Shop | Bakery | Pizza Place |
| 34 | Sutton Place | Italian Restaurant | Gym / Fitness Center | Furniture / Home Store | Park | Coffee Shop | Gym | Bakery | Thai Restaurant | Beer Bar | Spa |
| 35 | Turtle Bay | Coffee Shop | Sushi Restaurant | Italian Restaurant | Wine Bar | Park | Seafood Restaurant | Café | Japanese Restaurant | Deli / Bodega | French Restaurant |
| 39 | Hudson Yards | Gym / Fitness Center | Hotel | American Restaurant | Café | Italian Restaurant | Burger Joint | Dog Run | Gym | Park | Coffee Shop |

Cluster 1

# Discussion




Stuyvesant Town

- Cluster 3, which only has the neighborhood Stuyvesant Town, can be discarded since by just googling location we notice that this is a large private residential development, which are usually not a great ideal place to build your restaurant.
- Cluster 4 can also be discarded, but for a different reason. We only have one entry for this particular cluster which, if the model's chosen label were to be this one, would not be a result giving us great confidence in it.
- In cluster 5 restaurants are predominant, thus it isn't a good choice!

# Discussion

- Cluster 2 has the following statistics:

| | Neighborhood | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 17 |
| unique | 17 | 6 | 12 | 14 | 17 | 14 | 13 | 12 | 15 | 15 | 14 |
| top | Little Italy | Italian Restaurant | Coffee Shop | Pizza Place | Mexican Restaurant | Mediterranean Restaurant | Bakery | French Restaurant | Mediterranean Restaurant | Italian Restaurant | Japanese Restaurant |
| freq | 1 | 7 | 3 | 3 | 1 | 2 | 3 | 4 | 2 | 2 | 2 |

- Which features a great predominance of restaurants with a total number of 20 most common restaurants (not counting the Pizza Places). In contrast with cluster 1, which has the following statistics:

| | Neighborhood | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| unique | 10 | 6 | 6 | 9 | 8 | 9 | 10 | 9 | 9 | 9 | 9 |
| top | West Village | Park | Hotel | American Restaurant | Park | Coffee Shop | French Restaurant | Café | Coffee Shop | Italian Restaurant | Coffee Shop |
| freq | 1 | 3 | 4 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 2 |

# Discussion

- Looking back at Cluster 1's dataframe:

| | Neighborhood | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 11 | Roosevelt Island | Park | Hotel | School | Gym | Coffee Shop | Greek Restaurant | Dry Cleaner | Sandwich Place | Liquor Store | Noodle House |
| 13 | Lincoln Square | Café | Gym / Fitness Center | Plaza | Theater | Concert Hall | Performing Arts Venue | Wine Shop | American Restaurant | Italian Restaurant | Coffee Shop |
| 21 | Tribeca | Park | Italian Restaurant | Café | American Restaurant | Wine Bar | Spa | Coffee Shop | Skate Park | Hotel | Greek Restaurant |
| 24 | West Village | Italian Restaurant | New American Restaurant | American Restaurant | Park | Cocktail Bar | French Restaurant | Jazz Club | Coffee Shop | Wine Bar | Theater |
| 28 | Battery Park City | Park | Hotel | Coffee Shop | Gym | Memorial Site | Playground | Plaza | Shopping Mall | Gourmet Shop | BBQ Joint |
| 32 | Civic Center | Coffee Shop | Cocktail Bar | Hotel | Gym / Fitness Center | Spa | Yoga Studio | Café | French Restaurant | Italian Restaurant | Bakery |
| 33 | Midtown South | Korean Restaurant | Hotel | Japanese Restaurant | Burger Joint | Cosmetics Shop | Gym / Fitness Center | Clothing Store | Coffee Shop | Bakery | Pizza Place |
| 34 | Sutton Place | Italian Restaurant | Gym / Fitness Center | Furniture / Home Store | Park | Coffee Shop | Gym | Bakery | Thai Restaurant | Beer Bar | Spa |
| 35 | Turtle Bay | Coffee Shop | Sushi Restaurant | Italian Restaurant | Wine Bar | Park | Seafood Restaurant | Café | Japanese Restaurant | Deli / Bodega | French Restaurant |
| 39 | Hudson Yards | Gym / Fitness Center | Hotel | American Restaurant | Café | Italian Restaurant | Burger Joint | Dog Run | Gym | Park | Coffee Shop |

# Discussion

- We can already discard neighborhoods Tribeca, West Village, Midtown South, Sutton Place, Turtle Bay and Hudson Yards because they have restaurants at least as their 3rd most common venue. Leaving us with just Roosevelt Island, Lincoln Square, Battery Park City and Civic Center neighborhoods.

| | Neighborhood | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 11 | Roosevelt Island | Park | Hotel | School | Gym | Coffee Shop | Greek Restaurant | Dry Cleaner | Sandwich Place | Liquor Store | Noodle House |
| 13 | Lincoln Square | Café | Gym / Fitness Center | Plaza | Theater | Concert Hall | Performing Arts Venue | Wine Shop | American Restaurant | Italian Restaurant | Coffee Shop |
| 21 | | | | | | | | | | | |
| 24 | | | | | | | | | | | |
| 28 | Battery Park City | Park | Hotel | Coffee Shop | Gym | Memorial Site | Playground | Plaza | Shopping Mall | Gourmet Shop | BBQ Joint |
| 32 | Civic Center | Coffee Shop | Cocktail Bar | Hotel | Gym / Fitness Center | Spa | Yoga Studio | Café | French Restaurant | Italian Restaurant | Bakery |
| 33 | | | | | | | | | | | |
| 34 | | | | | | | | | | | |
| 35 | | | | | | | | | | | |
| 39 | | | | | | | | | | | |

# Discussion

- Then in order to choose just one neighborhood, we started to analyze from left to right until we found the first neighborhood with a restaurant as its most common venue, which is Roosevelt Island with a Greek Restaurant as its 6th most common venue.
- Again, reading from left to right, we get to the "8th Most Common Venue" column, which has, for both Lincoln Square and Civic Center a restaurant. We eliminate these entries and reach the conclusion that our best choice of a neighborhood is **Battery Park City**, featuring virtually no restaurant as its most common venue, featuring entries such as Park and Plaza which are in the single-entry dataframe that we used as test data.

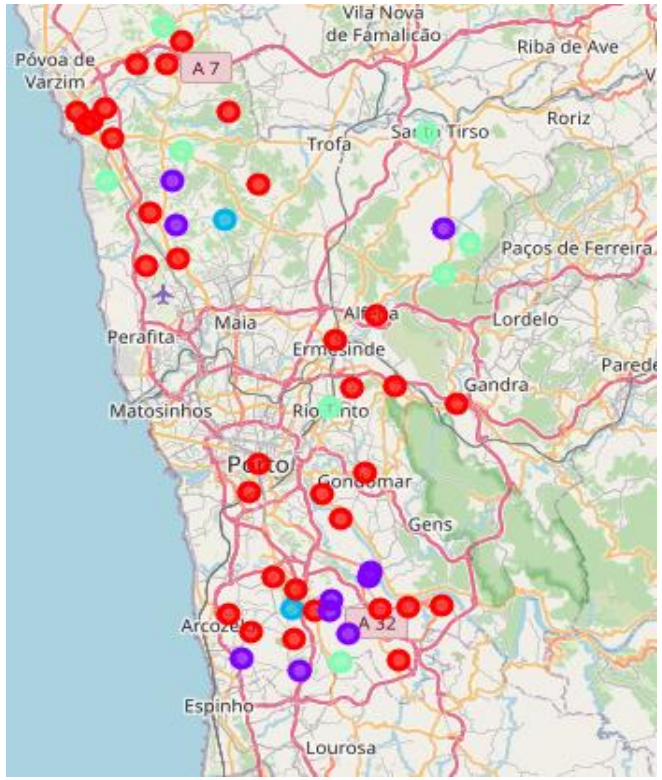
# Discussion


Battery Park City

Our Winner!

# Discussion

- Now which type of restaurant do we build? -> Porto as the most different city overall




**Portuguese Cuisine is the differentiator!**

# Conclusion

- Lack of geographical points to some cities (like Porto for example, in which the entire district of Porto had to be considered and not just the city itself)
- An American restaurant is in that same dataframe, but I opted for a more realistic approach, since it is difficult to assume that in New York City, a place already filled up to the top with so many restaurants (as discussed in the Introduction section), would have a neighborhood with no restaurants whatsoever.
- The test data set could be, of course, enhanced in order to feature more entries to check the validity of the model according to the results obtained in this work.
- And, also, many more cities could have been researched upon in order to place the new restaurant, however New York City, as it was shown, proved to be a great challenge.

# Thank you