

# **Neighborhood Recommender System**

**Based on previous Venues  
Ratings**

**IBM Professional Certificate**

**Jorge Eduardo Hermoso Valle**

**Coursera**

## Introduction

This web app will recommend the user a neighborhood to book a hotel based on previous user ratings to venues (ratings from wherever in the world).

The app will then have two main inputs, the first one will be the previous ratings of the user to different venues in whichever city/town, the second input will be a place or destination (in longitude and latitude) to which the user is looking to find a zone or hotel to stay. These inputs will be processed through a recommendation system, after that, the algorithm will already have insights to recommend a zone that is more likely for the user to find venues near that he or she will be likely to enjoy, and a list of hotels in that zone.

## Data to be used

1.- Dataset generated by the user ratings containing the venues he/she has visited with the rating he/she has given.

|   | venue              | rating |
|---|--------------------|--------|
| 0 | Adult Boutique     | 30     |
| 1 | Gastropub          | 40     |
| 2 | Golf Course        | 40     |
| 3 | River              | 50     |
| 4 | Rock Climbing Spot | 50     |
| 5 | Spa                | 40     |
| 6 | Yoga Studio        | 50     |

2.-Dataset of the venues on a given city obtained via Foursquare.

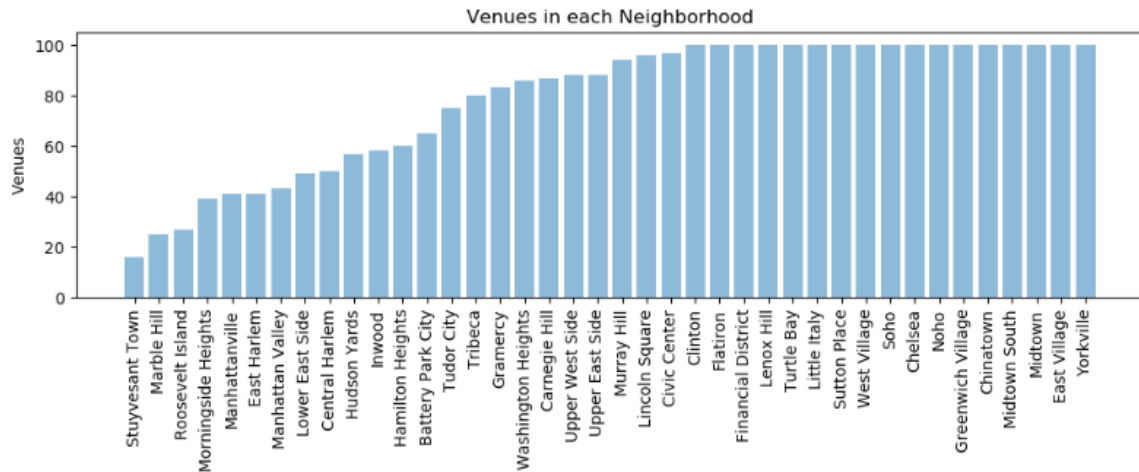
| Neighborhood      |     |
|-------------------|-----|
| Battery Park City | 65  |
| Carnegie Hill     | 87  |
| Central Harlem    | 50  |
| Chelsea           | 100 |
| Chinatown         | 100 |
| Civic Center      | 97  |
| Clinton           | 100 |
| East Harlem       | 41  |
| East Village      | 100 |

3.- Dataset of the different neighborhoods in NYC with its latitude and longitude.

|   | 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 |

## Methodology

First with the NYC coordinates dataset we send a query to Foursquare, one for each neighborhood. To get the top 100 venues in a max. radius of 500 yd. In response to that query we get a json file which we transform into a pandas data frame, and we can make the first exploratory analysis to check how many venues we have in each neighborhood.



Then the data frame is transformed with “one hot” to get the dummies of each kind of venue in each neighborhood.

|   | 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                      |

Once the dummies are created, a new data frame must transform that into a new frame with the mean frequency, also grouping all the venues of each neighborhood.

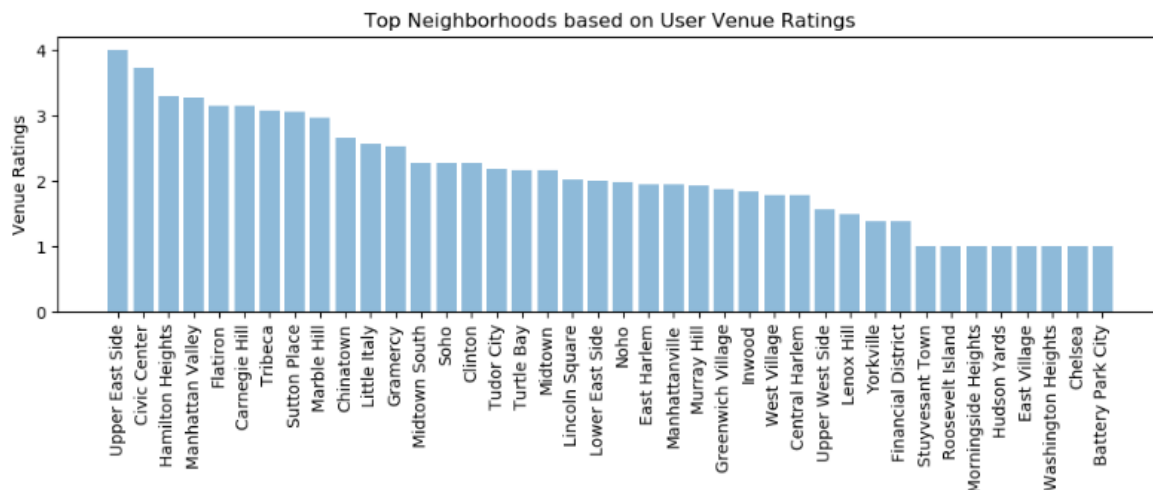
| Neighborhood      | Accessories Store | Adult Boutique | Afghan Restaurant | African Restaurant | American Restaurant | Antique Shop | Arcade | Arepa Restaurant | Argentinian Restaurant |
|-------------------|-------------------|----------------|-------------------|--------------------|---------------------|--------------|--------|------------------|------------------------|
| Battery Park City | 0.0               | 0.0            | 0.0               | 0.00               | 0.015385            | 0.0          | 0.0    | 0.0              | 0.000000               |
| Carnegie Hill     | 0.0               | 0.0            | 0.0               | 0.00               | 0.011494            | 0.0          | 0.0    | 0.0              | 0.011494               |
| Central Harlem    | 0.0               | 0.0            | 0.0               | 0.06               | 0.040000            | 0.0          | 0.0    | 0.0              | 0.000000               |
| Chelsea           | 0.0               | 0.0            | 0.0               | 0.00               | 0.040000            | 0.0          | 0.0    | 0.0              | 0.000000               |
| Chinatown         | 0.0               | 0.0            | 0.0               | 0.00               | 0.030000            | 0.0          | 0.0    | 0.0              | 0.000000               |

## Results

Then a multiplication of the “mean frequency neighborhood” data frame and the “user ratings” data frame to get a new weighted data frame which will be used in the recommender system, this new data frame will be sorted by the total column, to help us determine which neighborhood is a better match for the user.

|    | Neighborhood     | Accessories<br>Store | Adult<br>Boutique | Afghan<br>Restaurant | African<br>Restaurant | American<br>Restaurant | Antique<br>Shop | Arcade | Arepa<br>Restaurant | Argentinian<br>Restaurant |
|----|------------------|----------------------|-------------------|----------------------|-----------------------|------------------------|-----------------|--------|---------------------|---------------------------|
| 35 | Upper East Side  | 0.0                  | 0.0               | 0.0                  | 0.0                   | 0.011364               | 0.000000        | 0.0    | 0.0                 | 0.0                       |
| 5  | Civic Center     | 0.0                  | 0.0               | 0.0                  | 0.0                   | 0.030928               | 0.010309        | 0.0    | 0.0                 | 0.0                       |
| 13 | Hamilton Heights | 0.0                  | 0.0               | 0.0                  | 0.0                   | 0.000000               | 0.000000        | 0.0    | 0.0                 | 0.0                       |
| 20 | Manhattan Valley | 0.0                  | 0.0               | 0.0                  | 0.0                   | 0.000000               | 0.000000        | 0.0    | 0.0                 | 0.0                       |
| 10 | Flatiron         | 0.0                  | 0.0               | 0.0                  | 0.0                   | 0.010000               | 0.000000        | 0.0    | 0.0                 | 0.0                       |

We can create a bar chart to help us visualize the neighborhood-user match.



## Discussion

With this recommender system you can predict which neighborhood in a given city will have good venues matches regarding your previous ratings.

## Conclusion

The best match for the user is Upper East Side Neighborhood with a total rating of 4. And there are also 8 neighborhoods that does not have even one venue matching the user ratings, beneath the radius.