# The Best Neighborhoods to Live in Manhattan

## The Battle of Neighborhoods

## Applied Data Science Capstone Project: IBM/Coursera

Wei Kang

March 2021

# Contents

## Introduction

New York City, the home of more than 8.4 million people, is the most populous city in the US. Nearly half of New York City population is made up of people who were born outside the city or even outside the U.S. Its urban core, Manhattan, is the most densely populated of the five boroughs in New York City. As one of the world's major commercial, finance and cultural centers and a vital hub for many industries, Manhattan attracts people from the rest of the US and the world to live here.

There are different reasons that people move to Manhattan. Different people value certain features more than the others. Some people may want to live close to a grocery store or supermarket, some people may interested in good restaurants nearby, some might prefer to have park nearly, etc. It is not always easy to move to a new city. As people may feel exciting moving to a new place, it could be overwhelming to get to know the new place. Manhattan is an explorer's dream. It is important to know the various neighborboods and figure out the priorities. The better you know about the neighborhoods, the more likely you could pick the best one for you to live in.

In this project, I will explore the neighborhoods in Manhattan. The objective is to provide analysis of average rent and venues in each neighborhood from different aspects to help people who are moving to Manhattan to decide the best neighborhood to live in based on their interests and budget.

## Data

This project utilizes data from the following three sources:

1). New York city neighborhoods coordinates data from https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json

2). Average Rent in Manhattan neighborhoods data from https://www.rentcafe.com/average-rent-market-trends/us/ny/manhattan/

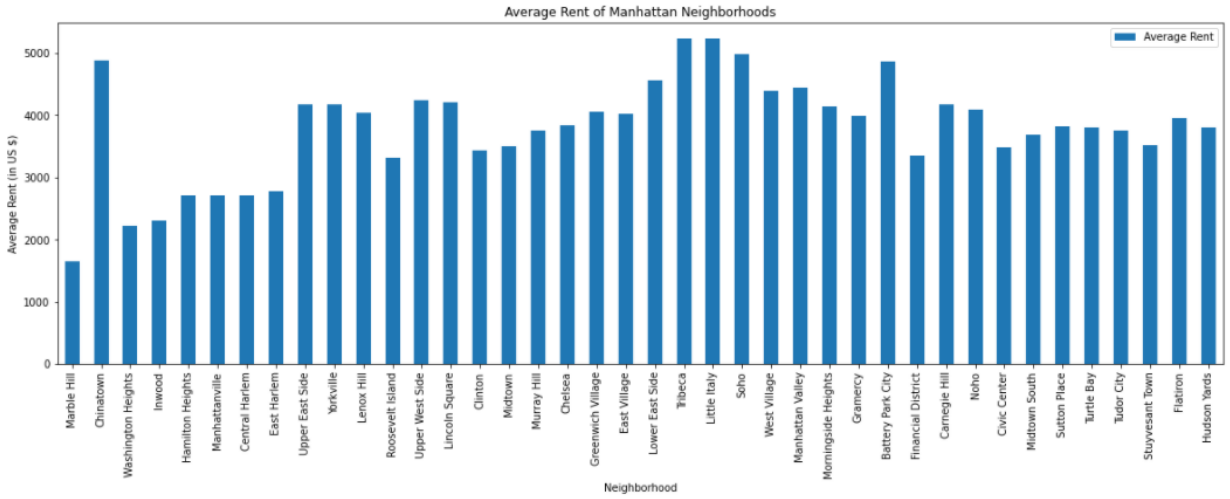
3). Venues for each neighborhood obtained from Foursquare API

The first data sources provide location (latitude and longitude) information for all the neighborboods from five boroughs in New York City. I cleaned the data and keep the coordinates information of the 40 neighborhoods in Manhattan only, which will be used to visualize the location of the neighborhoods in a Manhattan map and to gather venues information from Foursquare. The second data source provides average monthly rent data in U.S. dollars for each of the 40 neighborhoods in Manhattan, which could be used as a reference for budget planning assuming that most people moving to Manhattan will start off renting a place. Using latitudes and longitudes of the neighborhoods from the first data source as input for the Foursquare API, I will explore the neighborhoods by analyzing the venues of people's interests (parks, restaurants, grocery stores, yoga studios, etc).

## Methodology

In this sector, I start with a statistical data analysis of the neighborhoods rent by looking at the distribution of average rent in all Manhattan neighborhoods. Then, I create a map of the neighborhoods to provide a better idea of the locations of different neighborhoods. In order to dig deeper into the neighborhoods, I create a function that takes inputs of neighborhoods with latitudes and longitudes and returns the venues with in a certain defined radius of the neighborhood locations by utilizing the Foursquare API. Finally, I prepare a dataframe containing all the venues in Manhattan neighborhoods for further exploration in the Results sector.

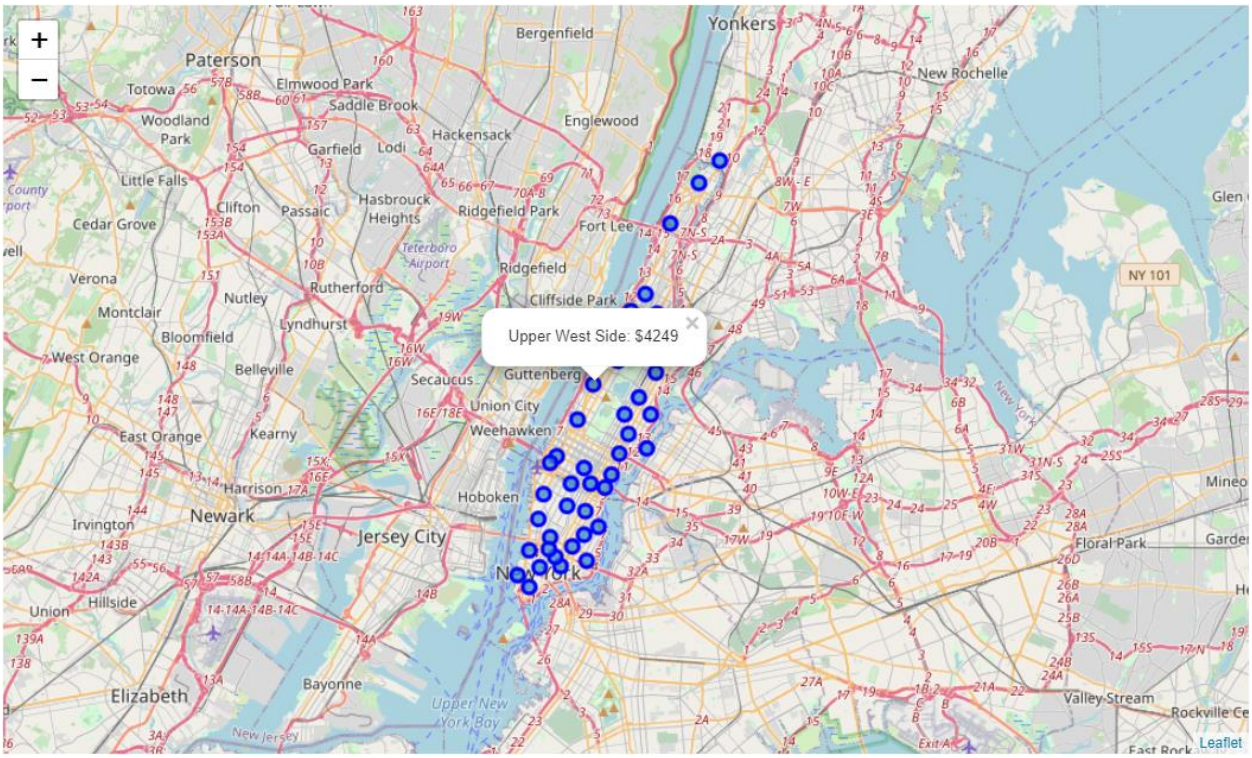
To find the best neighborhood to live in, the first important thing to do is to identify affordable neighborhoods with reasonable rent level. I create a bar chart using matplotlib to visualize the average rent data in all Manhattan neighborhoods, which is shown below.




Budget might be one of the most important factors that affects people's decision of where to live. By viewing some basic statistical details of the data, people know better about the rent level in Manhattan neighborhoods. Little Italy and Tribeca have the highest average rent at 5228 dollars among the 40 neighborhoods, while Marble Hill has the lowest rent at 1658 dollars. There are 16 neighborhoods with rent lower than the mean value of 3800.675 dollars, and 24 neighborhoods withe the rent above the mean.

| | Average Rent |
|---|---|
| **count** | 40.000000 |
| **mean** | 3800.675000 |
| **std** | 802.371242 |
| **min** | 1658.000000 |
| **25%** | 3476.750000 |
| **50%** | 3901.000000 |
| **75%** | 4185.500000 |
| **max** | 5228.000000 |

Another important factor that might affect people's decision is the location of neighborhoods. Many people prefer to live not too far from where they work or study. I create a map of Manhattan for people to easily visualize the location of neighborhoods with corresponding rent.




By utilizing the Foursquare API, I created a function that takes inputs of neighborhoods with latitudes and longitudes and returns the venues within a defined radius of the neighborhoods locations. The function returns a total number of 3201 venues of 332 categories. In the rest of the project, I will use the venues data to explore the neighborhoods in Manhattan.

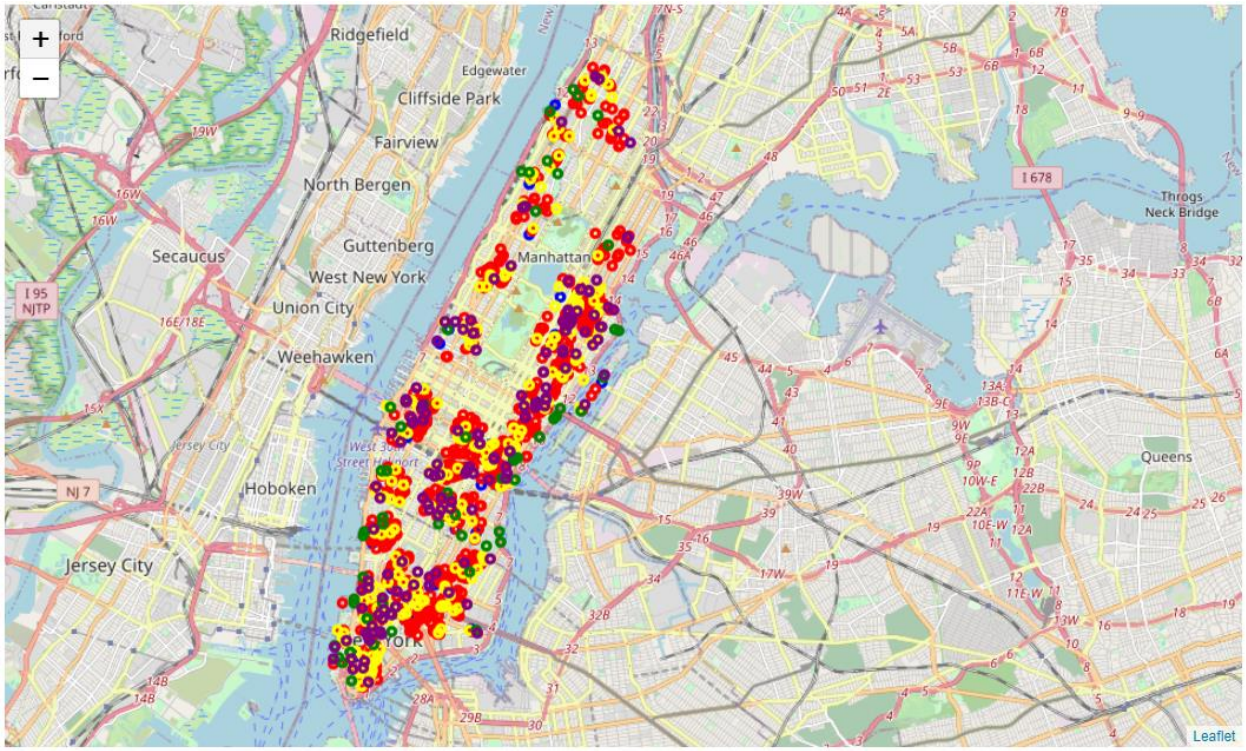
## Results

In this sector, I will first do a general check of the types of venues, filter out the venues of interest, and visualize these venues on a Manhattan map. Then I will perform one hot encoding to narrow the list of neighborhoods to the ones including the venues of interest only and identify the best neighborhood to live in. Lastly, a k-means clustering analysis is used to provide suggestions of similar neighborhoods that fits one's budget.

Assuming that people would prefer to live in neighborhoods that have five major types of venues (grocery stores, restaurants, coffee places, parks and gyms) nearby, I have summarized the detailed venue categories into proper types of venues and create a map to show these five types of venues in Manhattan

neighborhoods. The distribution of these five types of venues are shown in different colors of markers in the following map.

| venue | color |
|---|---|
| grocery | blue |
| restaurants | red |
| coffee | yellow |
| parks | green |
| gyms | purple |



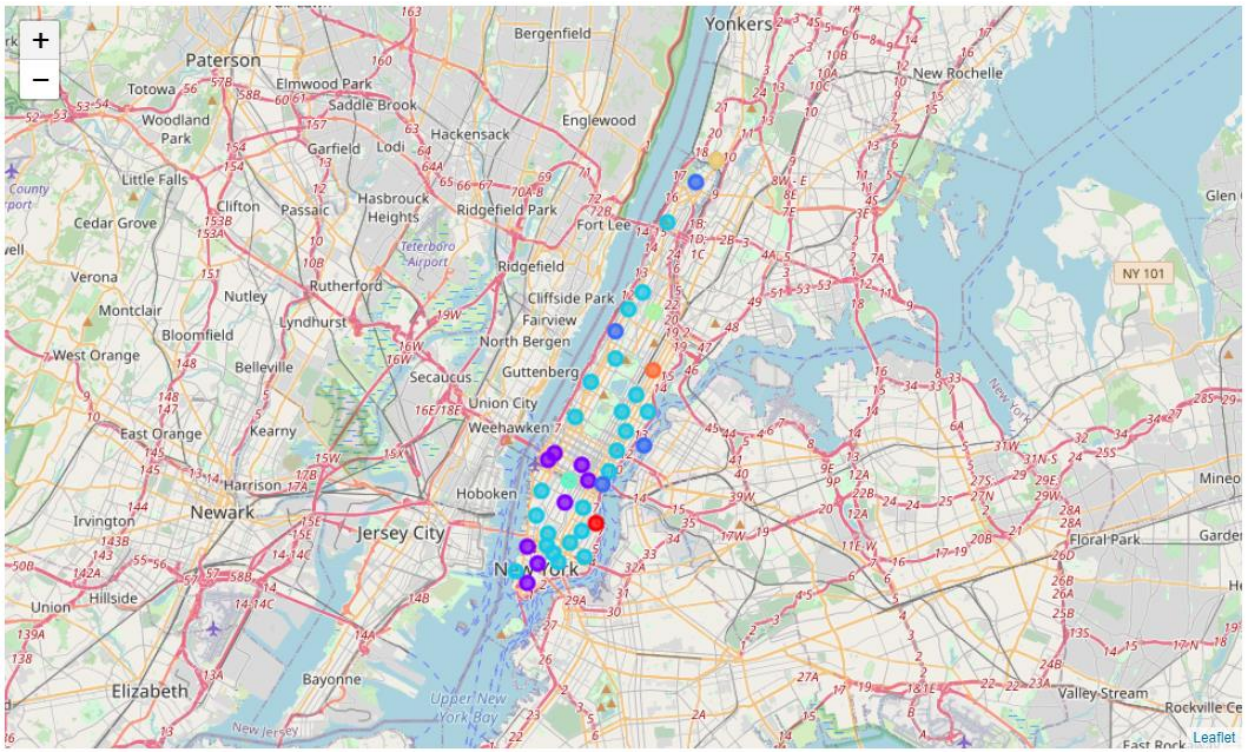

To better explore the venues of interest in each neighborhood, I use one hot encoding to count venues of interest (grocery stores, restaurants, coffee places, parks, and gyms) located in each neighborhood, filter out the neighborhoods which do not have all the of five desired types of venues, and reduced the list of neighborhoods from 40 to 17. For these 17 neighborhoods, I calculated a score using the normalized sum of venues of interest and sort the 17 neighborhoods by the “score” in the following table.

| | Neighborhood | Coffee | Grocery | Gym | Park | Restaurant | Score |
|---|---|---|---|---|---|---|---|
| 0 | Flatiron | 5 | 1 | 8 | 2 | 34 | 0.090090 |
| 1 | Turtle Bay | 8 | 1 | 1 | 3 | 36 | 0.088288 |
| 2 | Sutton Place | 4 | 2 | 11 | 3 | 27 | 0.084685 |
| 3 | West Village | 4 | 1 | 1 | 4 | 35 | 0.081081 |
| 4 | East Village | 4 | 1 | 1 | 1 | 34 | 0.073874 |
| 5 | Washington Heights | 8 | 2 | 3 | 2 | 22 | 0.066667 |
| 6 | Chelsea | 8 | 1 | 2 | 2 | 23 | 0.064865 |
| 7 | Hudson Yards | 5 | 2 | 7 | 2 | 18 | 0.061261 |
| 8 | Midtown | 5 | 1 | 4 | 1 | 22 | 0.059459 |
| 9 | Lincoln Square | 7 | 2 | 7 | 1 | 13 | 0.054054 |
| 10 | Gramercy | 5 | 2 | 1 | 2 | 18 | 0.050450 |
| 11 | Manhattanville | 4 | 2 | 1 | 1 | 19 | 0.048649 |
| 12 | Inwood | 4 | 2 | 1 | 2 | 18 | 0.048649 |
| 13 | Manhattan Valley | 3 | 1 | 2 | 1 | 15 | 0.039640 |
| 14 | Battery Park City | 5 | 1 | 4 | 4 | 7 | 0.037838 |
| 15 | East Harlem | 1 | 1 | 2 | 1 | 14 | 0.034234 |
| 16 | Roosevelt Island | 2 | 1 | 2 | 2 | 2 | 0.016216 |

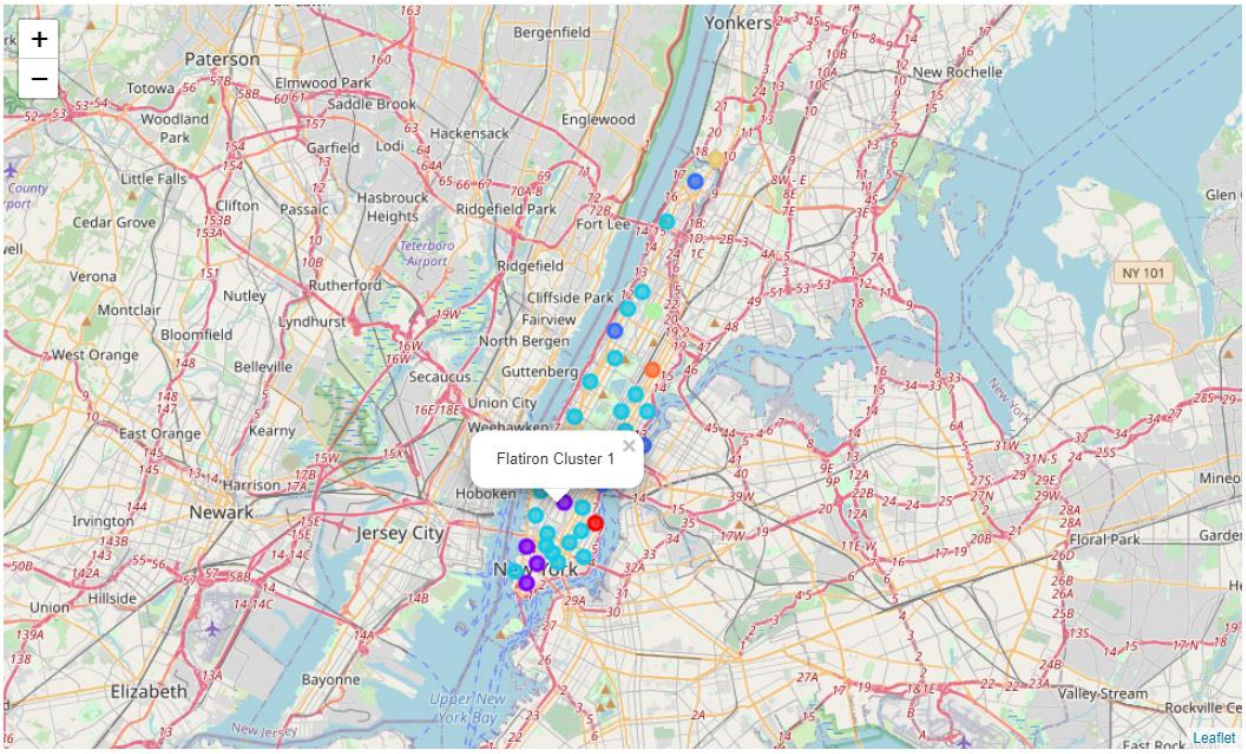
As is shown in the table, Flatiron ranks the first among the 17 neighborhoods with all of the five types of venues of interest, so it is identified as the best neighborhood to live in Manhattan. It is followed by Turtle Bay, Sutton Place, West Village, and Clinton.

I also understand that people sometimes would like to have more than one choices, and that Flatiron might not the best choice for various reasons. People might want to look around more neighborhoods before they make the decision where to settle down. It would be a good idea to identify more neighborhoods that are similar to Flatiron based on the venues and average rent level and provide more suggestions to people.

To do so, I apply k-means clustering methodology to cluster the 40 Manhattan neighborhoods into 8 clusters based on a normalized mean value of all the venues and the normalized average rent of neighborhoods. The resulting 8 clusters of neighborhoods are shown in the following map by different colors of markers.




The previous identified best neighborhood Flatiron is shown by a purple marker in the next map. All of the purple markers stand for neighborhoods in cluster 1. There are 8 neighborhoods in total in cluster 1 and they are shown in the following table. As the neighborhoods Battery Park City, Civic Center, Clinton, Financial District, Hudson Yards, Midtown, Murray Hill, and Upper East Side are in the same cluster as Flatiron, they are considered the most similar neighborhoods to the best neighborhoods. People could also take these neighborhoods into consideration when selecting a neighborhood to live in Manhattan.




| | Neighborhood |
|---|---|
| 0 | Civic Center |
| 1 | Clinton |
| 2 | Financial District |
| 3 | Flatiron |
| 4 | Hudson Yards |
| 5 | Midtown |
| 6 | Murray Hill |
| 7 | Tribeca |

## Discussion

In the previous sectors, various analysis have been performed to identify the best neighborhood(s) to live in Manhattan. We have looked into the neighborhoods from different aspects including average rent level, location, venues of interested and by k-means clustering.

There are 40 neighborhoods in Manhattan. The average rent of all neighborhoods is 3800.68 dollars per month. People could choose in live in neighborhoods with average rent as low as 1658 dollars/month, or in neighborhoods with average rent as high as 5228 dollars/month based on their budget and their expectations of the neighborhoods.

Assuming that people would prefer to live in a neighborhood that has all of the 5 major types of venues (grocery stores, restaurants, coffee places, parks, and gyms) nearby, I narrowed down the list of neighborhoods from 40 to 17 in the following table. I calculated a "Score" for each of these neighborhoods based on the normalized sum of venues to identify the best neighborhood, which is Flatiron.

| | Neighborhood | Average Rent |
|---|---|---|
| 0 | Civic Center | 3493 |
| 1 | Clinton | 3428 |
| 2 | Financial District | 3354 |
| 3 | Flatiron | 3962 |
| 4 | Hudson Yards | 3809 |
| 5 | Midtown | 3509 |
| 6 | Murray Hill | 3754 |
| 7 | Tribeca | 5228 |

Flatiron has an average rent that is slightly above the mean of average rent of all the neighborhoods. In order to provide more neighborhoods of suggestion, a k-means clustering is performed to identify 7 neighborhoods that are most similar to Flatiron. These neighborhoods in the same cluster with Flatiron are shown in the following table. People could consider these neighborhoods based on their budget.

## Conclusion

The purpose of this project is to provide suggestions to people on which is the best neighborhood to live in Manhattan. In order to provide proper suggestions, I have explored the 40 Manhattan neighborhoods through analysis of three important factors: rent, location, and venues. This project has utilized three datasets as described in the data sector.

This project has created a bar chart to visualize and compare average monthly rent across neighborhoods, and generated folium maps of Manhattan to visualize the locations of neighborhoods and venues of interest. The Foursquare API is utilized in this project to extract data of a total number of 3201 venues of 332 categories in Manhattan neighborhoods.

I have assume that people would prefer to live in a neighborhood that has all of the 5 major types of venues (grocery stores, restaurants, coffee places, parks, and gyms) nearby, and identified 17 neighborhoods that have all of the 5 types of venues. I calculated a "score" of neighborhoods based on the normalized sum of the 5 types of venues to identify the best neighborhood to live in Manhattan is Flatiron.

Lastly, a k-means clustering is applied to cluster 40 Manhattan neighborhoods into 8 clusters. The 7 neighborhoods in the same cluster with Flatiron, which are Battery Park City, Civic Center, Clinton, Financial District, Hudson Yards, Midtown, Murray Hill, and Upper East Side, could also be considered since they are most similar to the best neighborhoods.

This project could provide a starting point in searching for the best neighborhood to live in Manhattan. People should combine the findings of average rent, location and venues of neighborhoods in this project with their personal interest and budget to find the neighborhood that fits them the best.