The Battle of Neighborhoods

New York City, NY

Table of contents

I) Introduction

Discussion about Business / Human problem:

According to Brett & Kate McKay (The Art of Manliness, 2008) when we think about crafting our lives, we tend to concentrate on two questions: “What will I do for work?” and “Whom will I marry?” These questions, the what and who questions are certainly important. In fact, the answers to those questions constitute the two greatest determinants of our happiness.

But happiness is in fact a three legged stool, with the where question forming the third leg. This makes a great deal of sense: where we live will influence all of the other elements in our life, not just what we do and whom we marry, but the hobbies we pursue, the friends we make, the skills we learn, and how our children grow up.

For many men, where they end up is a matter of happenstance; they never leave their hometown; they take a job in a new city and settle down there; they move to follow a girlfriend and decide to stay after the break-up. However, because the place we live can affect our lives in so many ways, it ought to be a decision we approach deliberately and thoughtfully.

This is the story of one of my friends, Chacha, from his diminutive, who wants to settle with his boyfriend in America to start a new life (She actually lives in France). She loves America and dreams of living there especially in New York, Miami or Los Angeles.

In general, people in the same situation as Chacha tend to analyze and explore different locations before deciding to move to another country, state or city. There are many factors that

people need to consider when deciding to move to a new location like the cost of living, weather conditions, proximity to Family and Friends, Environment for Child Rearing and so on...

Chacha and his boyfriend have already thought a lot about what city they would like to live in and they choose New York for many reasons like convenience, diversity, job opportunity, transportation... ideally in the borough of Brooklyn or Manhattan.

So, it will be New York City! But it remains to know which is the best borough and the best neighborhood in which Chacha and his boyfriend will settle?

There are four criteria that are important to them and that will allow them to choose the place in which they will settle in New York as follows:

1. The cost of living (essentially average apartment rent in Brooklyn & Manhattan)
2. The job opportunity** (essentially presence of many large companies)
3. Borough & Neighborhood diversity (essentially about people)
4. Attractivity of the neighborhood (top venues analysis)

** Chacha is Marketing Manager in a French multinational corporation and his boyfriend work as a Data Scientist in a consulting firm. Ideally, they would like to find the same job in New York.

The main objective of this project is to help Chacha, his boyfriend and people in the same situation as them to choose the best location to live in New York in accordance with the criteria mentioned above.

Informal definition of the problem to be solved:

What is the best place to live in New York based on the four criteria mentioned above?

To find an answer to the problem, we will use data and tools to extract knowledge and make recommendations to Chacha and others in the same situation as her.

II) Analytic Approach

In order to answer to the question asked, we are going to use Machine Learning. First of all, we need to identify the problem of machine learning that we face.

Since we do not know the output (y) and we do not have historical data on what we would like to know, we cannot train a supervised algorithm to make predictions (regression or classification). In conclusion, we are faced with an unsupervised learning problem.

We will therefore use unsupervised learning on data that does not have history for y output. It is up to the machine to find the intrinsic structures in the data.

For this project, we will use K-means Clustering, an unsupervised machine learning algorithm to segment and cluster the different neighborhoods of New York. Thanks to our analysis and exploration, we will have better understanding of the similarities and dissimilarities between neighborhoods in NYC. The results of this project will allow us to guide the choice of Chacha and other people on the best place to live in NYC.

In Summary :

- Machine Learning problem : Unsupervised Learning
- Choice of Unsupervised Learning : K-Means Clustering
- Expected results after this project: The best location to live in NYC.

III) Data Requirements

In this step, we are looking to identify the data needed for this project.

For a supervised learning problem, we need data about what we are trying to predict (y), named the response variable (or variable to explain or dependent variable) and we need data about the explanatory variables or independent variables.

As we face an unsupervised learning problem, the data requirements are different (the response variable is not known).

Nevertheless, we still have to identify the independent variables x1, x2 ..., xn to feed and train the model. In view of the problem we are trying to solve and the four criteria discussed in the first step, the necessary data that we would need for this project are:

- Data about Borough and Neighborhood in New York city
- Data about the location information (Latitude and Longitude of Borough and Neighborhood in NYC)
- Data about average apartment rent by Neighborhood
- Data about Ethnic composition of the population

Finally, we will also use NYC geographic coordinates data as input data for the Foursquare API, which will be used to provide information on the top venues in each NYC neighborhood.

IV) Data Collection

The last step allowed us to identify the data requirements for this project.

In this step, we are going to collect the data needed. For this, we have to find where is the data that we would like to collect.

1) Data about Borough, Neighborhood, location information in NYC:

Neighborhood has a total of 5 boroughs and 306 neighborhoods

Data Source: This dataset is available on the web in the following link:
https://geo.nyu.edu/catalog/nyu_2451_34572

2) Data about average apartment rent by Neighborhood:

NYC Average Neighborhood Rents in June 2018 for two boroughs and their respective neighborhoods: 34 for Manhattan & 20 for Brooklyn. It is the most complete dataset we have found on the price of Manhattan and Brooklyn apartment rentals.

Data Source: This dataset is also available on the web, but in contrary of the Data about NYC Neighborhood, we have to build the dataset. To do this, we have two choices, we can use web scraping to collect data or built an excel file. We chose to build an excel file containing the necessary data because this method is faster than the first.
https://www.cityrealty.com/nyc/market-insight/features/trending-in-ny/nyc-average-neighborhood-rents-november-2018/25161

3) Data about Diversity (Ethnic composition of the population of each Borough in New York) :

Construction of a dataset containing information on the ethnic composition of population for the five boroughs of New York. This Dataset was built on the American Factfinder website by choosing the right filters. Then, we downloaded the data in csv format.
Data Source: https://en.wikipedia.org/wiki/Demographics_of_New_York_City

4) Data about New York Companies:

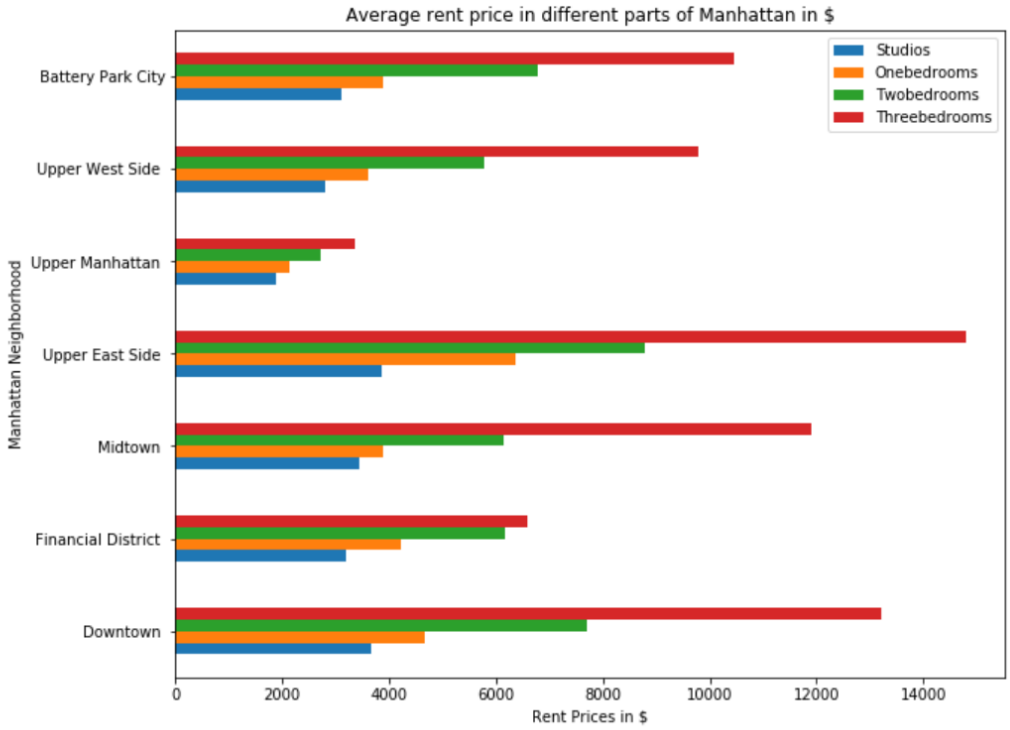
List of companies based in New York City. The table contains two columns (first one for the name of the company and the second one for the industry). We also have to collect addresses or geographic coordinates of these companies to have total information about NYC companies.
Data Source: The Data needed about New York companies is available on Wikipedia on the following link: https://en.wikipedia.org/wiki/List_of_companies_based_in_New_York_City

V) Data Preparation and Exploration

1. Comparison of apartment rent prices between Brooklyn and Manhattan

We will begin our analysis by comparing the price of the rents of the apartments between the two boroughs of New York.

Let us begin our exploration by getting the average rent price for the different parts of Manhattan:



As we can see, Upper Manhattan is the cheapest part of Manhattan and Upper East Side is the most expensive. Let us see the average minimum and maximum rent price for these two parts:

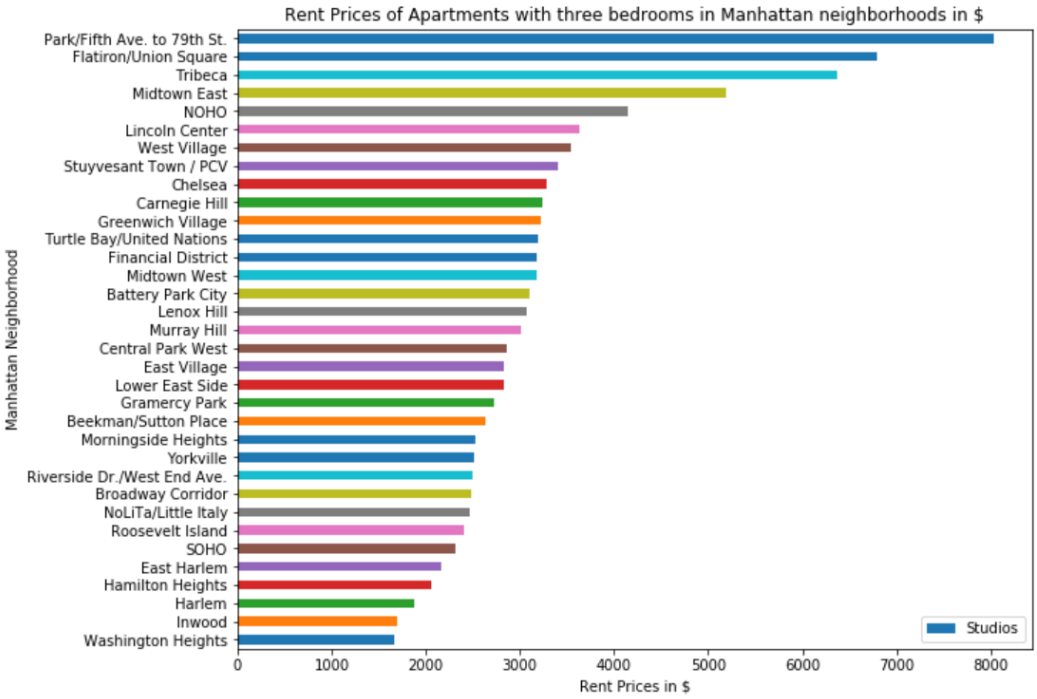
Upper Manhattan is the cheapest part of Manhattan:

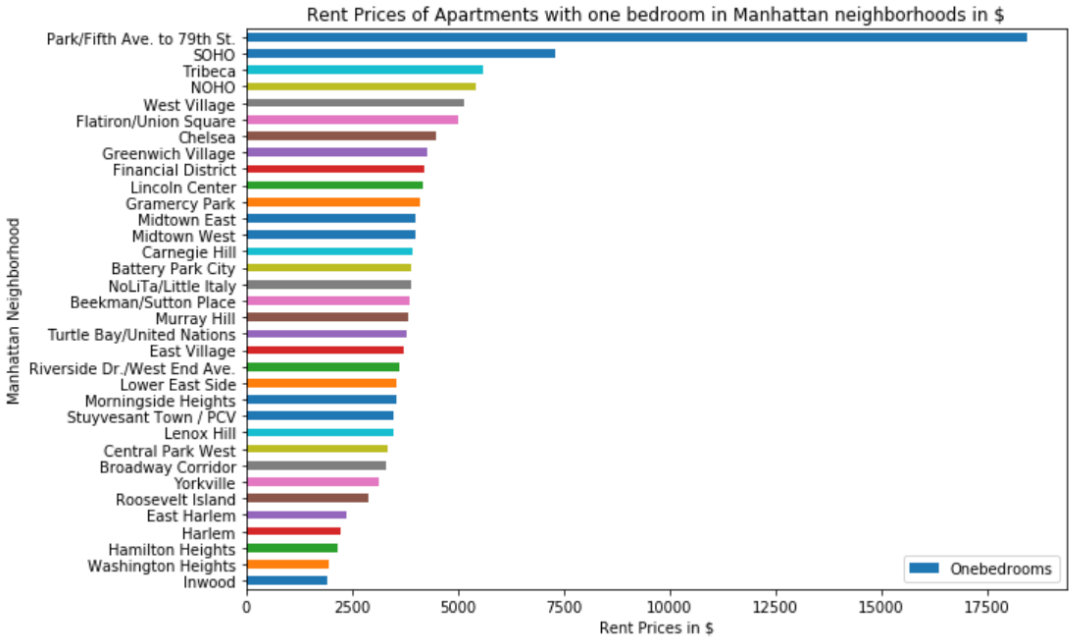
- The average minimum price of rent for a studio is 1892.2 $
- The average minimum price of rent for one bedroom is 2126.0 $
- The average minimum price of rent for two bedrooms is 2717.8 $
- The average minimum price of rent for three bedrooms is 3367.0 $

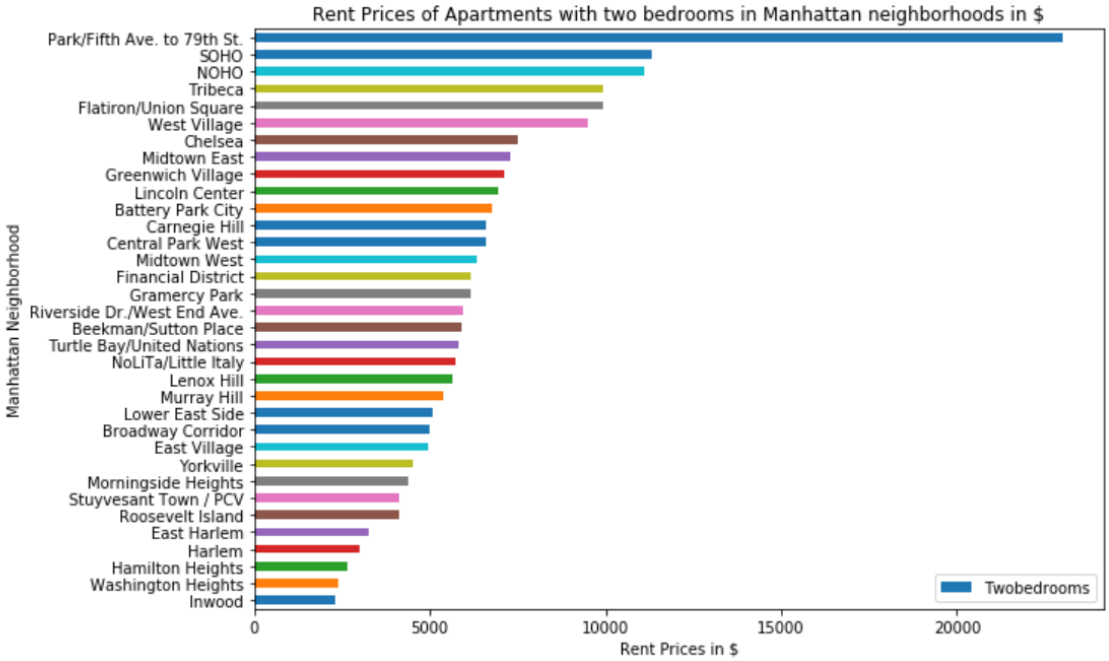
Upper East Side is the most expensive part of Manhattan:

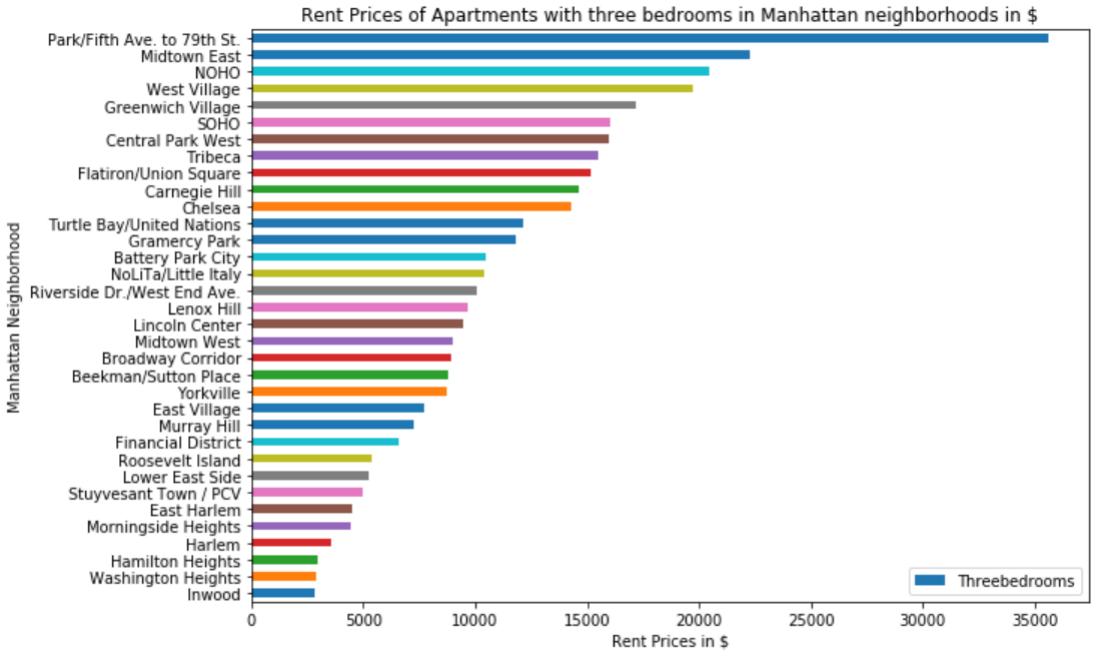
- The average maximum price of rent for a studio is 3852.0 $
- The average maximum price of rent for one bedroom is 6374.4 $
- The average maximum price of rent for two bedrooms is 8776.2 $
- The average maximum price of rent for three bedrooms is 14795.2 $

We are going to plot data and see the rent price for studios and apartments in Manhattan Neighborhoods.




Rent Prices of Apartments with one bedroom in Manhattan neighborhoods in $
Park/Fifth Ave. to 79th St.
SOHO
Tribeca
NOHO
West Village
Flatiron/Union Square
Chelsea
Greenwich Village
Financial District
Lincoln Center
Gramercy Park
Midtown East
Midtown West
Carnegie Hill
Battery Park City
NoLiTa/Little Italy
Beekman/Sutton Place
Murray Hill
Turtle Bay/United Nations
East Village
Riverside Dr./West End Ave.
Lower East Side
Morningside Heights
Stuyvesant Town / PCV
Lenox Hill
Central Park West
Broadway Corridor
Yorkville
Roosevelt Island
East Harlem
Harlem
Hamilton Heights
Washington Heights
Inwood
Manhattan Neighborhood
0
2500
5000
7500
10000
12500
15000
17500
Rent Prices in $
Onebedrooms


Rent Prices of Apartments with two bedrooms in Manhattan neighborhoods in $
Park/Fifth Ave. to 79th St.
SOHO
NOHO
Tribeca
Flatiron/Union Square
West Village
Chelsea
Midtown East
Greenwich Village
Lincoln Center
Battery Park City
Carnegie Hill
Central Park West
Midtown West
Financial District
Gramercy Park
Riverside Dr./West End Ave.
Beekman/Sutton Place
Turtle Bay/United Nations
NoLiTa/Little Italy
Lenox Hill
Murray Hill
Lower East Side
Broadway Corridor
East Village
Yorkville
Morningside Heights
Stuyvesant Town / PCV
Roosevelt Island
East Harlem
Harlem
Hamilton Heights
Washington Heights
Inwood
Manhattan Neighborhood
0
5000
10000
15000
20000
Rent Prices in $
Twobedrooms

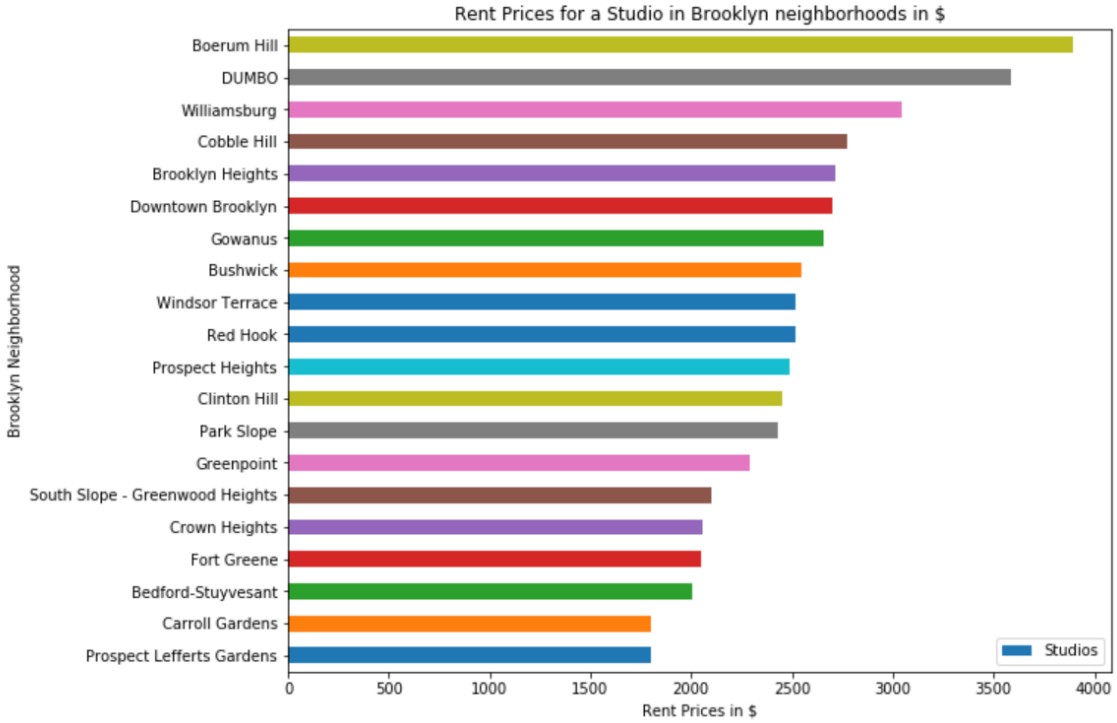


As we can see on the above visualizations, Washington Heights and Inwood are the cheapest neighborhoods in Manhattan to rent a Studio or an Apartment with one, two or three bedrooms. As we could expect, the most expensive Neighborhood in Manhattan is obviously Park/Fifth Ave. to 79th Street.

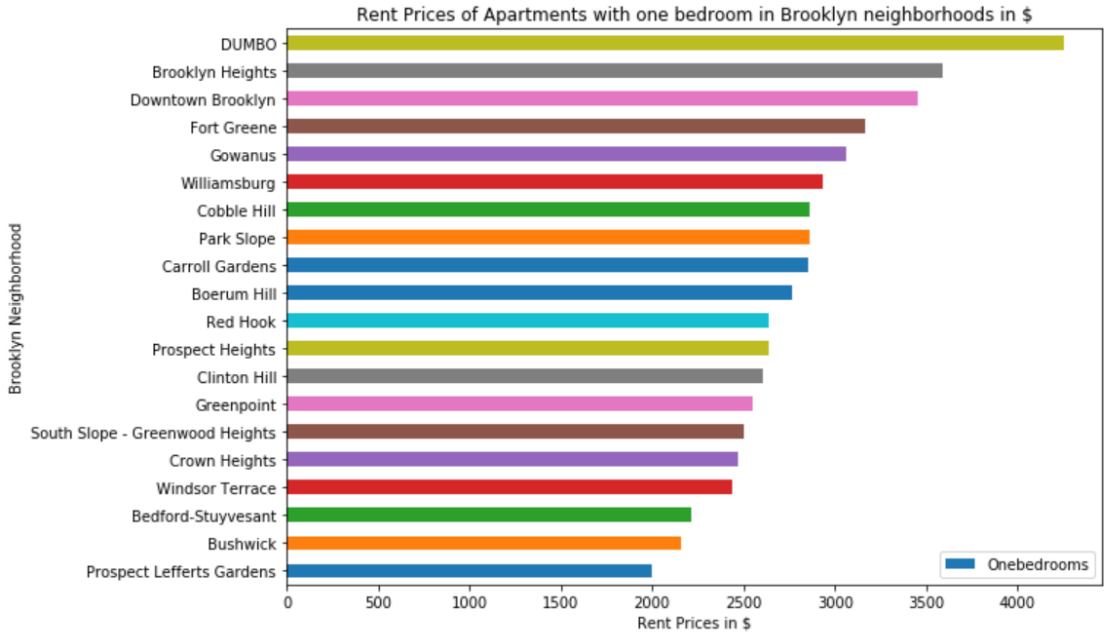
In Summary:

- The cheapest part of Manhattan is Upper Manhattan.
- The most expensive part of Manhattan is Upper East Side.
- The cheapest neighborhoods in Manhattan are Washington Heights and Inwood.
- The most expensive neighborhood in Manhattan is Park/Fifth Ave. to 79th Street.

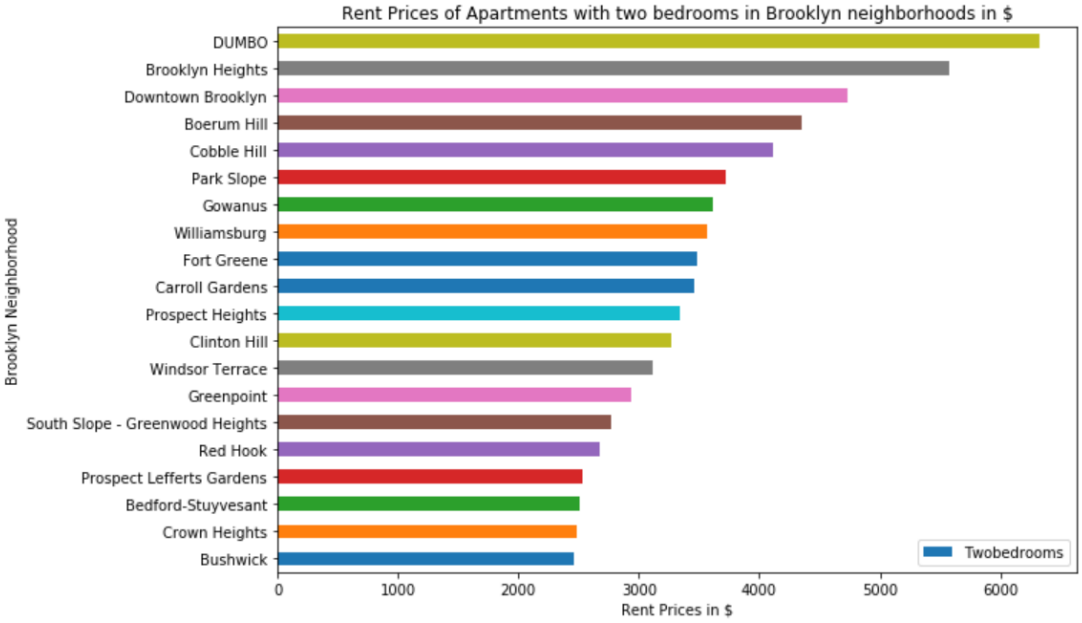
Now, let us compare the real rent prices for each neighborhood in Brooklyn.



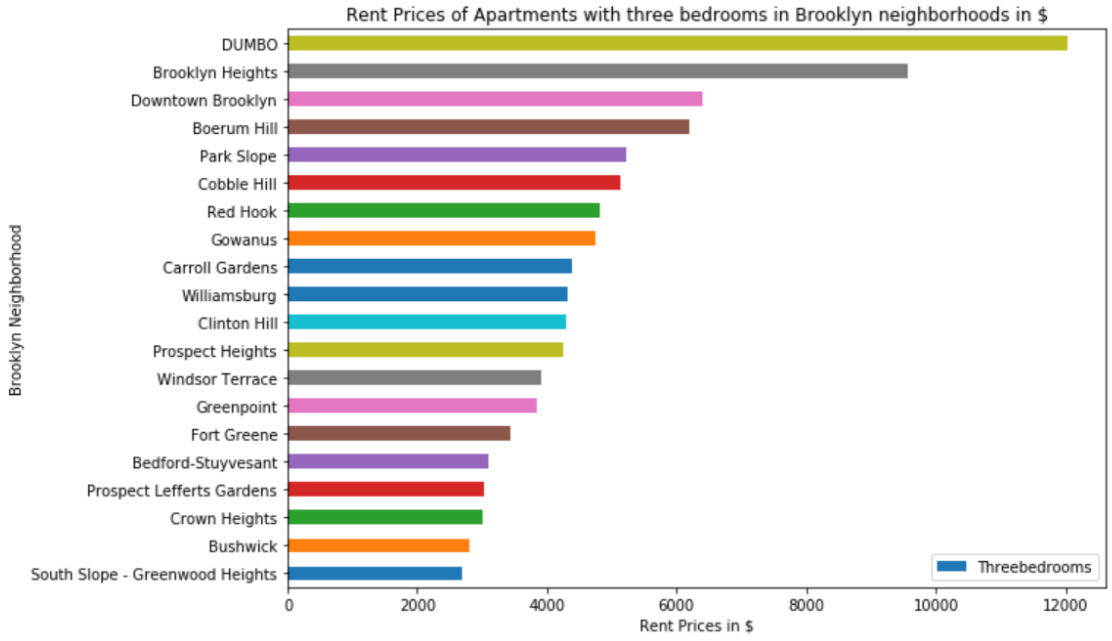
As we can see on the above visualization, Prospect Lefferts and Carrol Gardens are the cheapest neighborhoods in Brooklyn to rent a Studio. In contrast, the most expensive neighborhoods to rent a Studio in Brooklyn are Boerum Hill and Dumbo.



As we can see on the above visualization, Prospect Lefferts and Bushwick are the cheapest neighborhoods in Brooklyn to rent an apartment with one bedroom. In contrast, the most expensive neighborhoods to rent an apartment with one bedroom in Brooklyn are Dumbo & Brooklyn Heights.



As we can see on the above visualization, Bushwick and Crown Heights are the cheapest neighborhoods in Brooklyn to rent an apartment with two bedrooms. In contrast, the most expensive neighborhoods to rent an apartment with two bedrooms in Brooklyn are Dumbo & Brooklyn Heights.



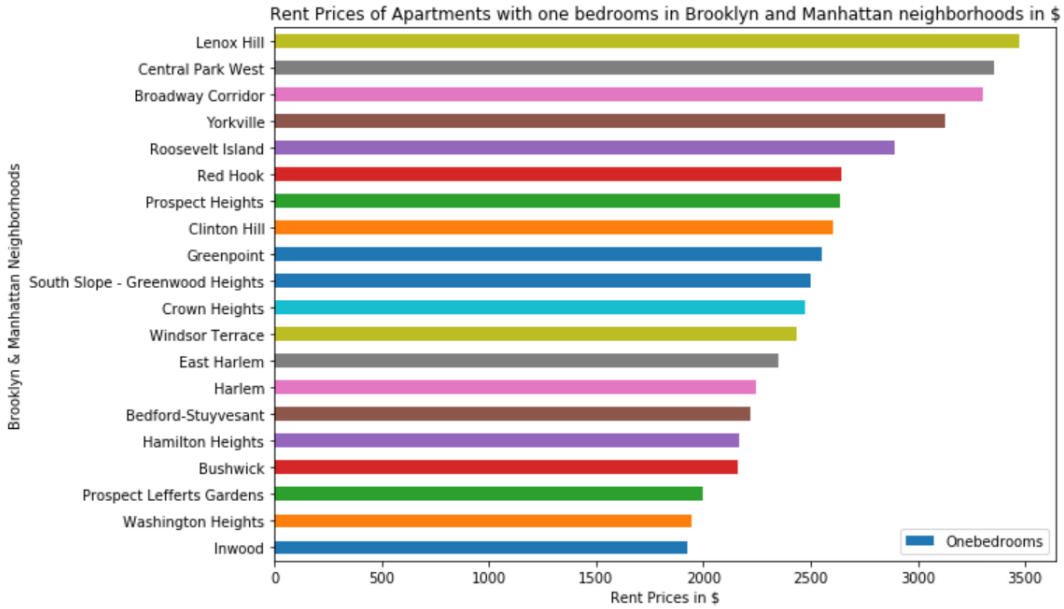
As we can see on the above visualization, South Slope and Bushwick are the cheapest neighborhoods in Brooklyn to rent an apartment with three bedrooms. In contrast, the most expensive neighborhoods to rent an apartment with three bedrooms in Brooklyn are Dumbo & Brooklyn Heights.

In Summary :

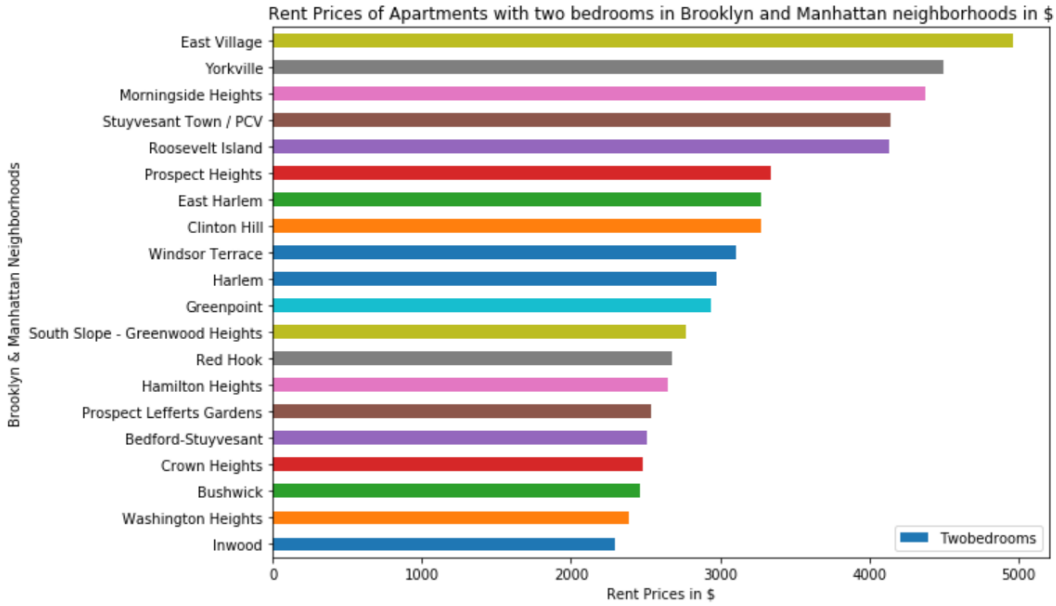
- The cheapiest neighborhoods to rent a Studio in Brooklyn are Prospect Lefferts and Carrol Gardens.
- The most expensive neighborhoods to rent a Studio in Brooklyn are Boerum Hill and Dumbo.
- The cheapiest neighborhoods to rent an apartment with one bedroom in Brooklyn are Prospect Lefferts and Bushwick.
- The most expensive neighborhoods to rent an apartment with one bedroom in Brooklyn are Dumbo & Brooklyn Heights.
- The cheapiest neighborhoods to rent an apartment with two bedrooms in Brooklyn are Bushwick and Crown Heights.
- The most expensive neighborhoods to rent an apartment with two bedrooms in Brooklyn are Dumbo & Brooklyn Heights.
- The cheapiest neighborhoods to rent an apartment with three bedrooms in Brooklyn are South Slope and Bushwick.
- The most expensive neighborhoods to rent an apartment with three bedrooms in Brooklyn are Dumbo & Brooklyn Heights.

Let us continue our analysis by comparing the 10 cheapest neighborhoods in Manhattan and Brooklyn.

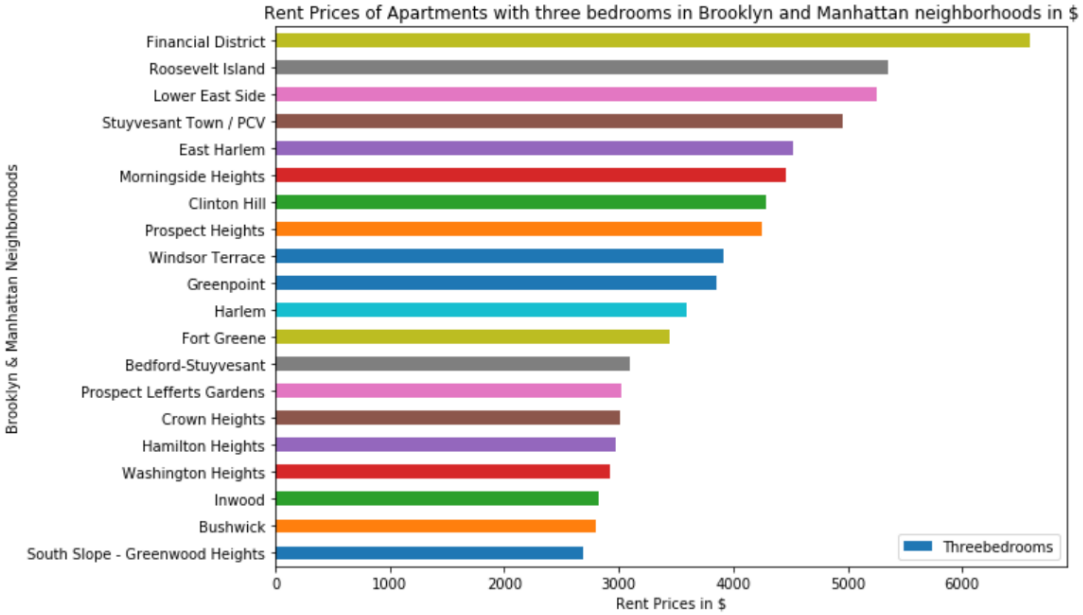
In the following of our exploration, we will focus our analysis on the rents of apartments and we will exclude analysis on studios. The reason is simple, Chacha and his boyfriend want to rent an apartment with one or two bedrooms but after a few months in NYC, they would like more bedrooms to accommodate their families and friends. Therefore, in the penultimate section, we will choose the best neighborhood for an apartment with one or two bedrooms.



That is interesting! Indeed, at first glance, we might think that the rent are more expensive in Manhattan than in Brooklyn. This is true for most neighborhoods in Manhattan but not for those located in Upper Manhattan! Inwood, Washington Heights and Prospect Lefferts Gardens are the cheapest neighborhoods to rent an apartment with one bedroom! Actually two of them (Inwood, Washington Heights) are Manhattan Neighborhoods and are the cheapest neighborhoods in to rent an apartment with one bedroom.



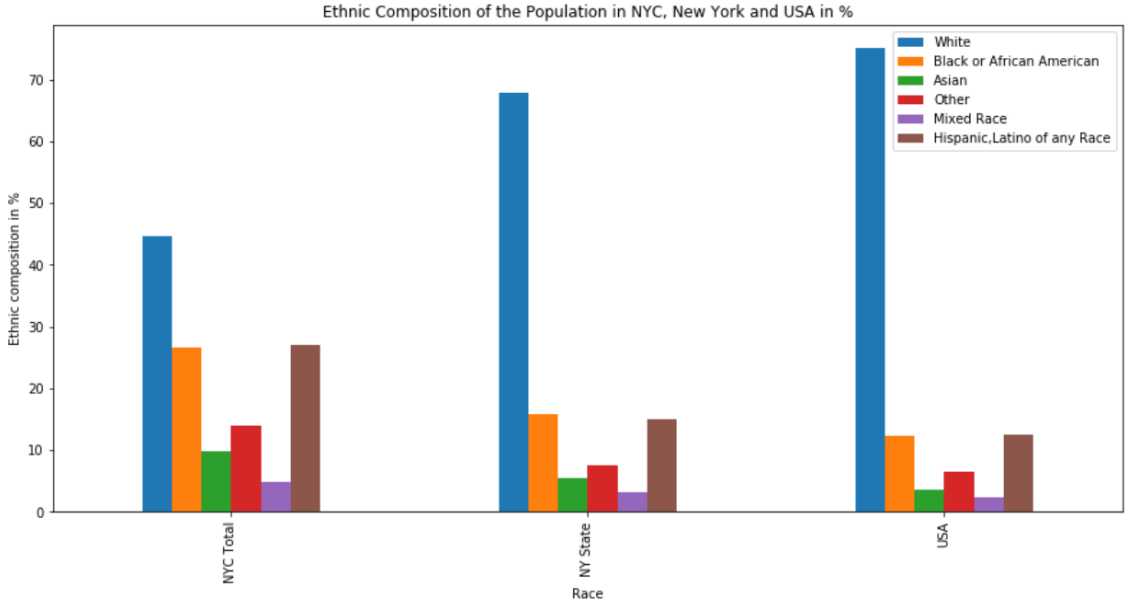
Inwood (Manhattan), Washington Heights (Manhattan) and Bushwick (Brooklyn) are the cheapest neighborhoods to rent an apartment with two bedrooms!



Sout Slope (Brooklyn), Bushwick (Brooklyn) and Inwood (Manhattan) are the cheapest neighborhoods to rent an apartment with three bedrooms!

Our analysis on comparing the price of the rents of the apartments between Manhattan and Brooklyn is now finish. Now, we are going to analyze ethnic composition of the New York's population.

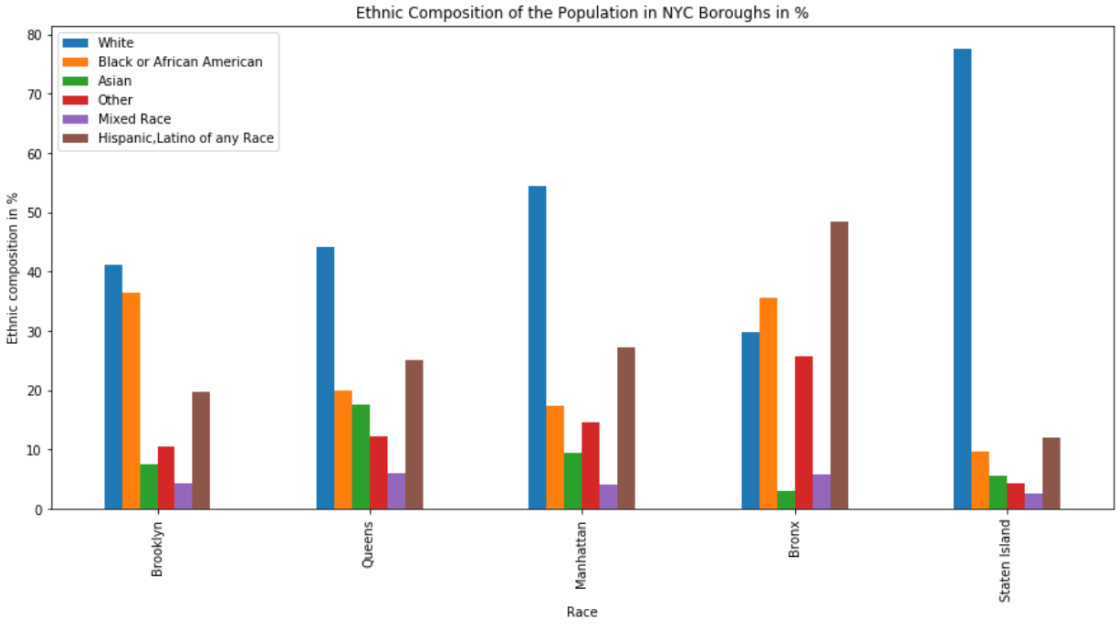
2. Analysis of the ethnic composition of the populations of the various boroughs of New York.



The legend of the visualization for "Other" corresponds to:

- American Indian, Native Alaskan, Native Hawaiian, and Pacific Islander

As we can see on the visualization above and as we could expect, New York City is a great diverse city and probably the most diverse city in USA. Indeed, if we compare Ethnic composition of NYC with The State of New York or USA, we can see that there is a big gap between the ethnic composition of the state of New York and in general the USA in comparison to that of the city of New York. So now, we confirmed that New York City is a great diverse city, let us see which neighborhoods in New York contribute the most to this diversity!



Thanks to the visualization above, we can say that three New York neighborhoods contribute the most to the diversity of the city:

- Bronx
- Queens
- Brooklyn

Now, if we compare Brooklyn and Manhattan on the diversity, we can affirm that Brooklyn is the most diverse borough. If we wanted to go into more details, we should analyze the ethnic composition of each neighborhood in Brooklyn and Manhattan.

However, this is not necessary in this project, we just wanted to confirm with this analysis that New York is a much-diversified city and we wanted to identify which boroughs contribute the most to this diversity. We also know from this analysis that Brooklyn is a more ethnically diverse borough compared to Manhattan.

In the remainder of our study, we will focus on Brooklyn because we now know that Brooklyn apartment rentals are generally cheaper than rented apartments in Manhattan and that the Brooklyn is more diversified than Manhattan. The following analysis will allow us to identify the best locations in each neighborhood of Brooklyn. Finally, taking into account all the criteria used in our study, we will be able to determine the best neighborhood (s) to settle in New York City.

3. Segmenting and Clustering Neighborhoods in New York City.

Create a map of New York with neighborhoods superimposed on top.

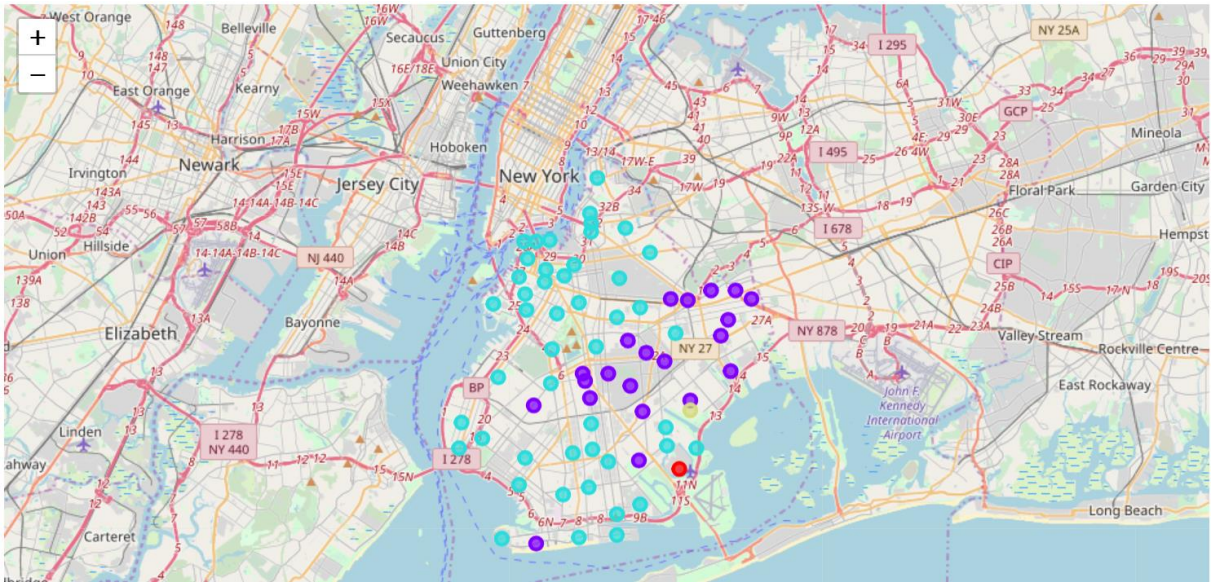
The visualization above represent all the neighborhoods in New York City. However, our analysis will be only on Brooklyn Neighborhoods.

As we did with all of New York City, let us visualize the neighborhoods in Brooklyn! We will use the geographical coordinates of Brooklyn to create a Map of Brooklyn's Neighborhoods.

Now, we are going to start utilizing the Foursquare API to explore the neighborhoods and segment them.

VI) Modeling the problem to be solved: Cluster Neighborhoods.

We will run *k*-means to cluster the neighborhood into 4 clusters.



Examine Clusters

In this step, we are going to examine each cluster and determine the discriminating venue categories that distinguish each cluster.

K-means will partition Brooklyn Neighborhoods into mutually exclusive groups, in our example, into four clusters. The neighborhoods in each cluster are similar to each other. Now we can create a profile for each group, considering the common characteristics of each cluster. For example, the four clusters can be:

- Cluster one (Red): Neighborhood rich in aquatic activities and fish trade.
- Cluster two (Purple): Neighborhoods rich in sporting activities and with many restaurants.
- Cluster three (Yellow): Neighborhoods very rich in culinary diversity and with many local stores.
- Cluster four (Blue): Neighborhood with many places to eat.

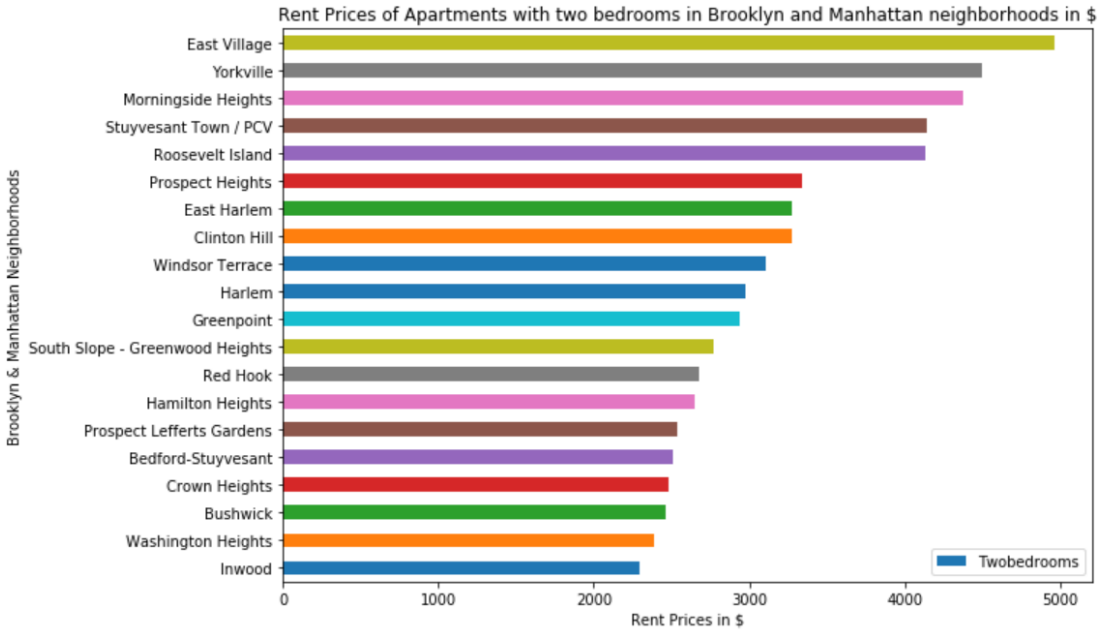
VII) Analysis of Results

To choose which Neighborhood(s) is/are the best to live in NYC, we will recall the criteria that were defined at the beginning of this study.

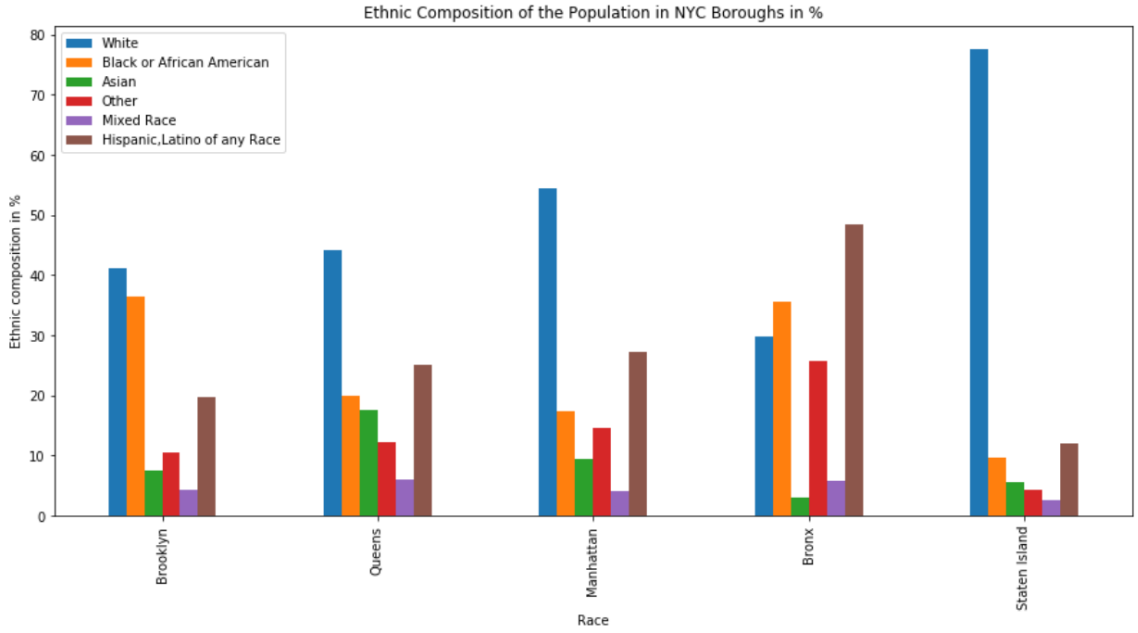
As we said in the beginning, there are four criteria that are important for Chacha and his boyfriend and that will allow them to choose the place in which they will settle in New York as follows:

1. The cost of living (essentially average apartment rent in Brooklyn & Manhattan)
2. The job opportunity** (essentially presence of many large companies)
3. Borough & Neighborhood diversity (essentially about people)
4. Attractivity of the neighborhood (top venues analysis)

So let us begin analysis of Results by plotting the cheapest neighborhoods in Brooklyn and Manhattan.



We can see that in general, Brooklyn's Neighborhoods are cheaper than Manhattan Neighborhood (except for Upper Manhattan).



We can see on the visualization above that if we compare Brooklyn and Manhattan on the diversity, we can affirm that Brooklyn is the most diverse borough.

In the end of our study, we had focus on Brooklyn because we knew that Brooklyn apartment rentals are generally cheaper than rented apartments in Manhattan and that Brooklyn is more diversified than Manhattan. The end of our analysis allowed us to identify the best locations in each neighborhood of Brooklyn.

Finally, taking into account all the criteria used in our study, we now be able to determine the best neighborhood (s) to settle in New York City.
Now in the final step of our analysis, let us visualize the DataFrame containing the cheapest neighborhoods in Brooklyn to rent an apartment with one or two bedrooms.

	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 I Com Venu
3	Brooklyn	Greenpoint	40.730201	-73.954241	2	Bar	Pizza Place	Coffee Shop	Cocktail Bar	Café	Bake
9	Brooklyn	Crown Heights	40.670829	-73.943291	2	Pizza Place	Museum	Café	Bagel Shop	Bakery	Play
12	Brooklyn	Windsor Terrace	40.656946	-73.980073	2	Diner	Food Truck	Plaza	Park	Bagel Shop	Café
13	Brooklyn	Prospect Heights	40.676822	-73.964859	2	Bar	Mexican Restaurant	Cocktail Bar	Thai Restaurant	Wine Shop	Ame Rest
16	Brooklyn	Bushwick	40.698116	-73.925258	2	Bar	Mexican Restaurant	Coffee Shop	Pizza Place	Discount Store	Bake
17	Brooklyn	Bedford Stuyvesant	40.687232	-73.941785	2	Coffee Shop	Café	Pizza Place	Bar	Pharmacy	BBQ
21	Brooklyn	Red Hook	40.676253	-74.012759	2	Art Gallery	Seafood Restaurant	Café	Park	Bagel Shop	Wine
38	Brooklyn	Clinton Hill	40.693229	-73.967843	2	Italian Restaurant	Pizza Place	Yoga Studio	Grocery Store	Japanese Restaurant	Rest
42	Brooklyn	Prospect Lefferts Gardens	40.658420	-73.954899	2	Bakery	Café	Caribbean Restaurant	Pizza Place	Ice Cream Shop	Supe

As we can see in the DataFrame, nine of the ten cheapest neighborhoods in Brooklyn belong to Cluster 3.

As the cluster index starts at zero, our cluster 2 is the third cluster:

- Cluster 3: Neighborhoods very rich in culinary diversity and with many local stores.

So nine of ten cheapest neighborhoods in Brooklyn are Neighborhoods with a very rich culinary diversity and with many local stores!

Our best neighborhoods to live are definitely in this top ten!

VIII) Conclusion

The best neighborhoods with the respect of the criteria defined at the beginning of the project are:

- Clinton Hill
- Green Point
- Bushwick

These three neighborhoods are in the top 10 of the cheapest neighborhoods in Brooklyn. In addition, these neighborhoods are close to Manhattan, which is a good thing for job opportunities. Indeed, the majority of large companies (consulting firms included) are located in the heart of Manhattan, so a good proximity to these neighborhoods with Manhattan is a crucial point. Not to mention that these neighborhoods are part of cluster 4. Therefore, these neighborhoods are very rich in culinary diversity and have many local shops.

Note however that we chose three neighborhoods without being able to establish a ranking between them. This is simple, arrived at this point of the study, we would have to consider other variables in order to differentiate them completely and know which the best is. These factors to be taken into account could be :

- Affordability (The price of gasoline, consumable goods...)
- Taxes
- Crime Rates and Statistics
- Healthcare Facilities...

The great thing about Data Science is that we can always go further. Future prospects for this study would be to continue this project taking into account other factors such as those mentioned above!

In terms of conclusion and to answer the problem of Chacha and his boyfriend, we can say that the best borough for living in New York City is Brooklyn and that the three best neighborhoods to settle are Clinton Hill, Green Point, and Bushwick.

Oh and I forgot to tell you but I am the data scientist looking to settle to New York with his girlfriend! ;)