

Applied Data Science

week 2

Capstone Project — IBM Data Science

THE BATTLE OF NEIGHBORHOODS

An analysis of neighborhoods in New
York
using data science methodologies

ABSTRACT

Currently, we are sitting on a mountain of data, information that must be classified and ordered so that it can be useful, the correct analysis and reading of the data can determine the success or failure of a business model, department search, etc. In this case we will try to determine the best location for an apartment in the borough of Manhattan, New York City, United States.



1 Introduction

1.1 Background Information

Manhattan is the most densely populated neighborhood of the 5 boroughs in New York City. It consists mainly of the island of Manhattan, surrounded by the Hudson, East and Harlem rivers. It is the heart of “the Big Apple” and is one of the most important cultural, financial and commercial centers in the world. Its iconic sites include skyscrapers like the Empire State Building, neon-lit Times Square, and the theaters of Broadway.

This project aims to estimate the best location to get and move to an apartment in this neighborhood.

This Capstone Project aims to create an analysis of characteristics to search for a better neighborhood as a comparative analysis between neighborhoods. Features include median home prices and top rated restaurants.

Additionally, we will use Foursquare data to explore and compare different neighborhoods. Application of machine learning algorithms to find the most promising district in New York.

This data could be used to determine a subsequent business plan.

1.2 Problem statement:

The problem or uncertainty is where exactly to locate the apartment.

Since the point of this is to create a business plan at the end, we need to make sure the API data is correct. We must also verify that the client may be interested in this specific business.

2 Data

2.1 Description of the Data

This notebook is highly inspired by the template given in the course. I will keep the idea of clustering the city by area and then plot heatmap to find better area.

59 / 5.000

Data link: https://geo.nyu.edu/catalog/nyu_2451_34572

I will change some data:

- ✓ Country/City: New York, United Stated.
- ✓ Goal: Locate an apartment suitable for housing, which has an advantage in terms of location with respect to the rest of the apartments.

I will use the following API:

- ✓ Foursquare API: to find apartament. ([foursquare](#))
- ✓ Google API: reverse geolocalisation.

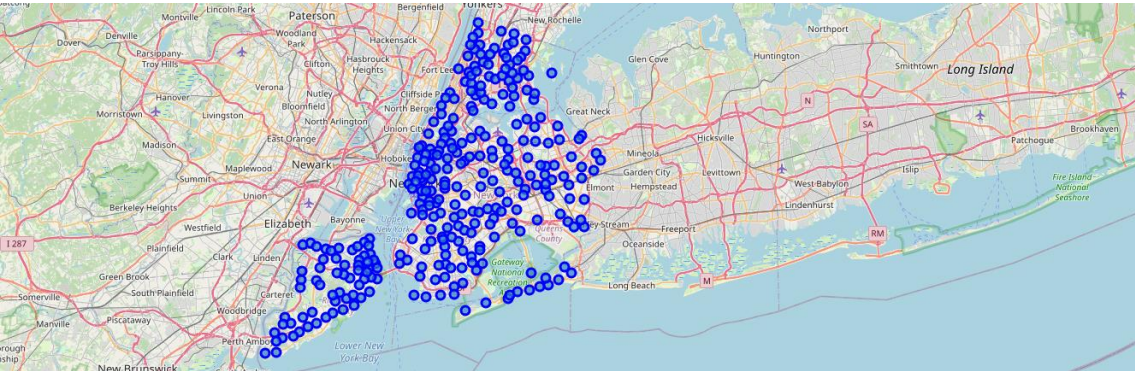
2.2 How the data will be used to solve the problem

The neighborhood has a total of 5 districts and 306 neighborhoods. To segment the neighborhoods and explore them, we will probably need a dataset that contains all 5

neighborhoods and the neighborhoods that exist in each district, as well as the latitude and longitude coordinates of each neighborhood.

Foursquare API Data:

We will need data on different places in different neighborhoods of that specific district. To obtain that information, we will use location information from "Foursquare" (Source: <https://foursquare.com/>). Foursquare is a provider of location data with information about all kinds of places and events within an area of interest. Such information includes place names, dates, menus, tips and more. As such, it is driven as the only data source, as all required information listed can be obtained through the API.



Grouping Approach:

To compare the similarities, I decided to explore neighborhoods, segment them, and group them into groups to find informative data. To be able to do that, we need to cluster data, which is a form of unsupervised machine learning: k-means clustering algorithm.

Most common places near the neighborhood:

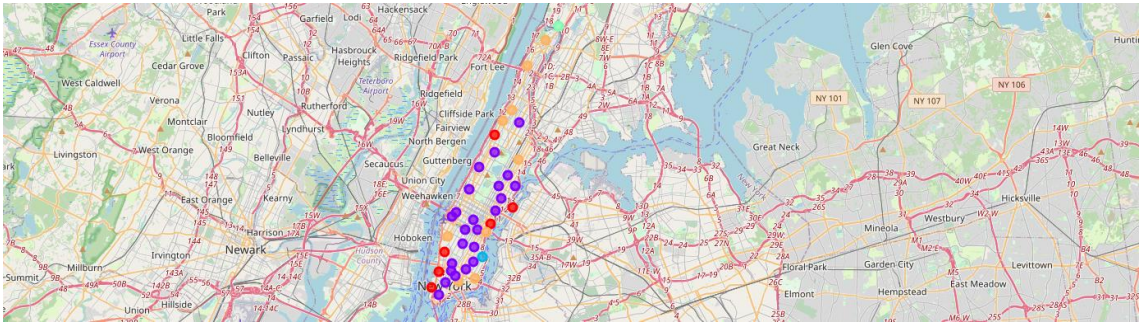
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Battery Park City	Park	Hotel	Memorial Site	Gym	Burger Joint	Italian Restaurant	Gourmet Shop	Boat or Ferry	Beer Garden	
1	Carnegie Hill	Coffee Shop	Yoga Studio	Grocery Store	Bookstore	Café	Wine Shop	Gym / Fitness Center	French Restaurant	Bakery	
2	Central Harlem	African Restaurant	Seafood Restaurant	Pizza Place	Bar	Chinese Restaurant	French Restaurant	Coffee Shop	Public Art	American Restaurant	
3	Chelsea	Art Gallery	Coffee Shop	Café	Bakery	Seafood Restaurant	Market	Park	Cocktail Bar	Electronics Store	
4	Chinatown	Chinese Restaurant	Cocktail Bar	Bakery	American Restaurant	Spa	Dim Sum Restaurant	Hotpot Restaurant	Ice Cream Shop	Optical Shop	

Using the K-Means clustering approach:

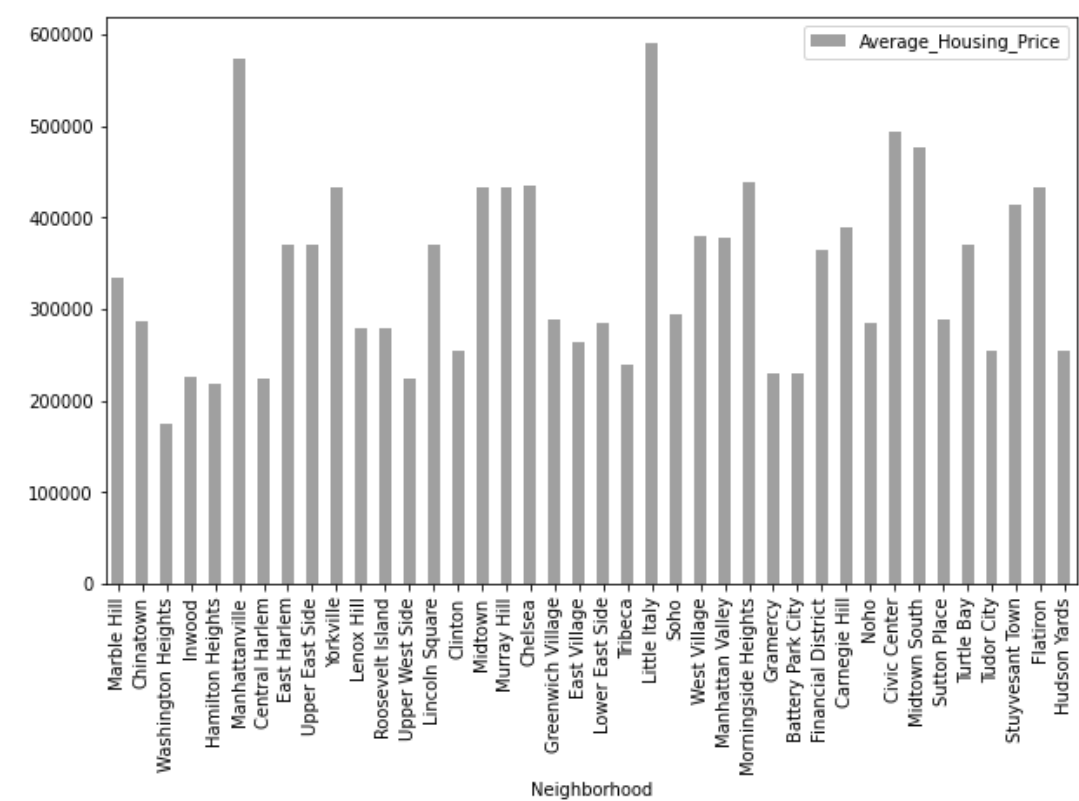
	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Manhattan	Marble Hill	40.876551	-73.910660	3	Gym	Coffee Shop	Sandwich Place	Yoga Studio	Kids Store	Storage Facility	Steakhouse	Seafood Restaurant	Deli / Bodega	
1	Manhattan	Chinatown	40.715618	-73.994279	4	Chinese Restaurant	Cocktail Bar	Bakery	American Restaurant	Spa	Dim Sum Restaurant	Hotpot Restaurant	Ice Cream Shop	Optical Shop	
2	Manhattan	Washington Heights	40.851903	-73.936900	4	Café	Latin American Restaurant	Bakery	Mobile Phone Shop	Chinese Restaurant	Deli / Bodega	Supplement Shop	Supermarket	Bank	
3	Manhattan	Inwood	40.867684	-73.921210	4	Mexican Restaurant	Lounge	Restaurant	Café	Chinese Restaurant	Park	Deli / Bodega	Pizza Place	Spanish Restaurant	
4	Manhattan	Hamilton Heights	40.823604	-73.949688	4	Pizza Place	Café	Mexican Restaurant	Deli / Bodega	Coffee Shop	Bakery	Sandwich Place	School	Cocktail Bar	

Result:

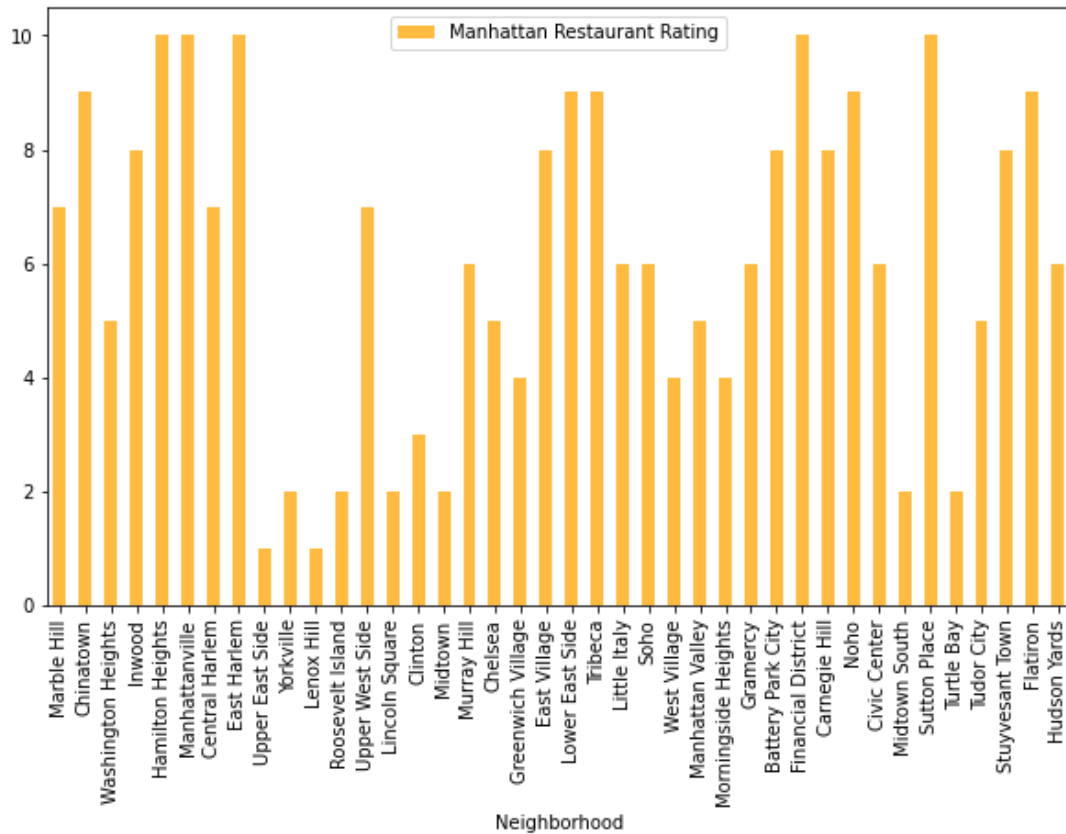
Map of Clusters in Manhattan



Median home price:



Restaurant Ratings:



Solution to the problem

The main objective of this project is to suggest a better neighborhood in a new city for someone who is interested in moving to the site.

In the present work, the following study was carried out

1. Ranked list of houses in terms of house prices in ascending order.
2. Ranked restaurant ratings in terms of location, rating, and reviews.

The project may extend to the following or may include more scenarios, to name a few below:

1. Making it more precise in terms of finding the best house/restaurant/school in Manhattan.
2. Connectivity with the airport, the train station, the city center, the markets and other daily necessities nearby.

Conclusion section:

You have used the k-means clustering algorithm. I separated the neighborhood into 5 (five) different groups and for 40 different latitudes and longitudes of the data set, which have very similar neighborhoods around them. Using the charts above, the results presented for a particular neighborhood have been made based on median home prices and restaurant rating.

This course provides a huge set of tools that allow you to be actively involved in data analysis, most of the decisions that life requires you to make day by day can be simplified or guided by a simple study of data, day by day we compile an impressive amount of information that we must learn to classify and use to minimize the possible error.