

House Sales Prices & Venues Data Analysis of New York City

Xia Zhang

June 2020

1. Introduction

New York City (NYC) is the most populous city in the United States. The estimated population in NYC in 2019 was 8,336,817 distributed over about 302.6 square miles. NYC is also the most densely populated major city in the United States. With almost 20 million people in its metropolitan statistical area and approximately 23 million in its combined statistical area, it is one of the world's most populous megacities. New York City has been described as the cultural, financial, and media capital of the world, significantly influencing commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports [1].

Situated on one of the world's largest natural harbors, New York City is composed of five boroughs, each of which is a county of the State of New York. The five boroughs are: Brooklyn, Queens, Manhattan, the Bronx, and Staten Island. At its core is Manhattan, a densely populated borough that's among the world's major commercial, financial and cultural centers. Its iconic sites include skyscrapers such as the Empire State Building and sprawling Central Park. Broadway theater is staged in neon-lit Times Square.

With the above statistics and information, one kind easily imagine that living in New York City can be very exciting especially if you are a city person. However, it can also be very expensive, especially the housing. While on the other hand, NYC has always been very attractive for investors all over the world, whether to invest in the housing market, opening restaurants, bars, entertainment center, etc. Whatever the perspective objectives are, a comprehensive understanding of different areas in NYC with the corresponding statistics including housing prices, & venues information will be very critical for either a personal buyer or an investor. Although one can easily search for information on the internet for an area of interest, combined information of average housing price and venues information is not readily available. Thus, the primary objective for this study is to perform a comprehensive investigation and analysis of the mean unit housing prices in different neighborhoods of New York City as well as to explore the surrounding venues for each neighborhood and combine this two information to show on the same map.

In this study, eventually, two maps will be created:

- 1.1 A map of housing located at different neighborhood of New York City for the five boroughs, each borough shown with a different color. A popup label was created for each building labeled with building class category name, borough, neighborhood, and the mean unit housing price.
- 1.2 A map of clustered neighborhoods of NYC with combined information of mean unit housing price and venues information: popup labels showing the borough, neighborhood, top surrounding venues and the mean unit housing prices. The neighborhoods will also be clustered into 4 clusters by DBSCAN method and a name will be given to each cluster based on the appearance frequency of different venues for each cluster.

2. Data Description

The following data were used for the present study:

- The Department of Finance's Rolling Sales files of New York City, which lists properties that sold in the last twelve-month period in New York City for all tax classes [2]. These files include information on:
 - The neighborhood
 - Building type
 - Square footage
 - Sale price
 - Other data

Data source : [link1](#) and [link2](#)

- Geocoder.arcgis method was used to get the latitudes and longitudes pairs of each neighborhood in the housing sales file.
- Foursquare API was used to explore nearby venues for each neighborhood of NYC [3].

References:

[1] https://en.wikipedia.org/wiki/New_York_City

[2] <https://www1.nyc.gov/site/finance/taxes/property-rolling-sales-data.page>

[3] Foursquare API