# Real Estate Purchase Price correlation to Neighborhoods

COURSERA CAPSTONE PROJECT

Smriti Sharma | IBM Data Science Professional Certificate | 11 June 2020

## Introduction

The cost of living in New York is relatively expensive and in fact, perhaps one of the most expensive in the world. It is very important to consider your geographical location and other factors when deciding to purchase a house there. There is very dense population and along with that, there is very high cost as well because they know that the demand for these houses is very and unfortunately someone with a very low elasticity for these things, would tend to buy houses and therefore it is important to consider where you live.

Furthermore, New York City also has large dining venues, shopping places and a lot of more monumental areas. Due to its very rich diversity in different areas such as food and culture, it is yet again considered something very high in demand.

## Audience

The target audience of this project will be investors or other stakeholders who would like to buy or sell their properties within a certain budget plan in mind. Oftentimes, all stakeholders have certain budget in mind and in cities like Toronto an New York, they always surpass their budget and to avoid such issues the stakeholders face, this project will use complex data science techniques to give relevant information to the stakeholders.

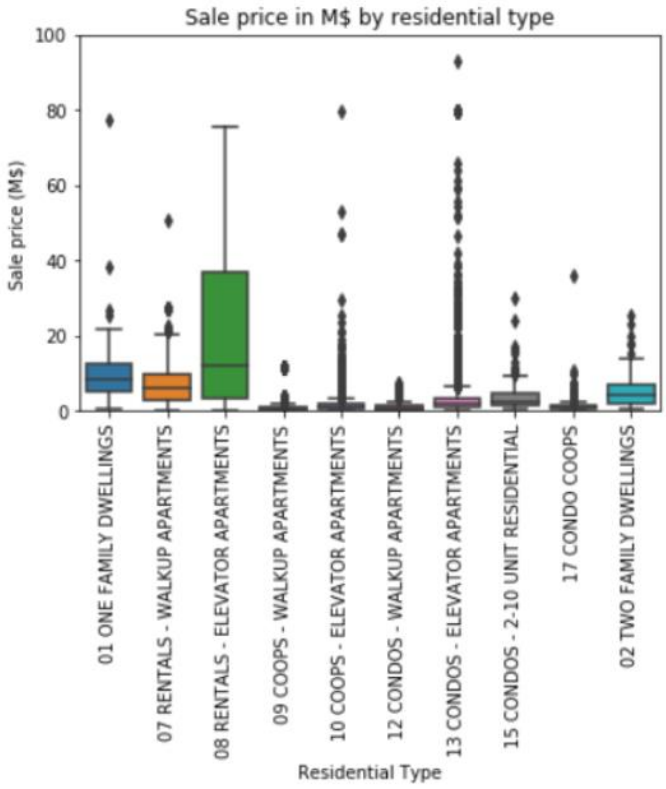
In this work the cost of purchase of a residential unit in different neighborhoods of Manhattan will be analyzed based on residential unit type, neighborhood, and square feet area. The cost of different types of residence I different neighborhood will be discusses. In addition, this work investigates the correlation between number of venues in a every neighborhood in Manhattan to the cost of purchase of residential unit. Last, this work classifies every neighborhood by its venues content and the median cost of purchase of a residential unit.

## Focus question and other questions

The focus/research question of this project will be to find the correlation between the type and number of venues in a specific neighborhood in Manhattan to its residential real estate cost.

## Methodology

The Manhattan sale price by residential category type is shown in the box plot. The price reflects to total cost per purchase including all the units in the property, as shown below rentals elevator apartment has the highest cost probably due to the large number of residential units under each purchase.




In order to avoid data biasing, the purchase price per unit is used as a metrics.

| | BUILDING CLASS CATEGORY | PricePerUnit |
|---|---|---|
| 0 | 01 ONE FAMILY DWELLINGS | 7.825000 |
| 1 | 02 TWO FAMILY DWELLINGS | 2.032500 |
| 2 | 07 RENTALS - WALKUP APARTMENTS | 0.500000 |
| 3 | 08 RENTALS - ELEVATOR APARTMENTS | 0.346698 |
| 4 | 09 COOPS - WALKUP APARTMENTS | 0.625000 |
| 5 | 10 COOPS - ELEVATOR APARTMENTS | 0.880000 |
| 6 | 12 CONDOS - WALKUP APARTMENTS | 0.677500 |
| 7 | 13 CONDOS - ELEVATOR APARTMENTS | 1.700000 |
| 8 | 15 CONDOS - 2-10 UNIT RESIDENTIAL | 2.475000 |
| 9 | 17 CONDO COOPS | 0.810000 |

*Figure 7- Median price per unit by residential type*

Figure 7 shows a summary table of the price per unit median by residential type. The data shows the following:

- One family house is more expensive by approximately 4 times than 2 family houses
- Elevator co-ops and elevator apartments are significantly more expensive than non-elevator apartment and co-ops. One can claim that elevator improves the property value. But it is also possible the building with elevators are never, modern, and larger and therefore more expensive.
- In case of rental classified properties – walkup apartments are more expensive than elevator apartments.
- Generally speaking – condos are more expensive than co-ops. In condos, the buyer purchases the apartment while co-op the buyer purchases a chare in a unit that reflects the size of the apartment in a building – similar to buying a share in the stock market.

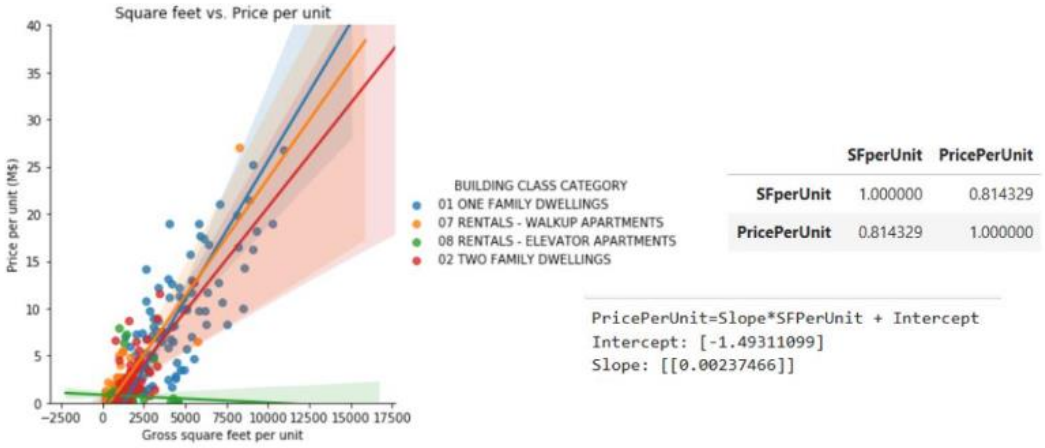



*Figure 8. Gross square feet per unit vs. price per unit*

An exploratory correlation between the size of the residential unit measured by square feet per unit vs. the price per unit was made. Some building categories missed the gross square feet data therefore only building categories that had this type of data ere considered. Fig. 8 shows the correlation between housing cost and the size of the property. A strong linear correlation was found it price per unit is auto correlated with square feet per unit with $r^2$ of 0.814. the slope of the correlation is 0.002, which means that roughly every square feet in Manhattan costs about $2000. that indicates the very high cost properties in Manhattan as of 2019.

## Discussion

A summary of price per residential unit by neighborhood is shown in Fig. 9. In the analysis all the residential types were consolidated. One can see that there is a large variation in price based on the neighborhood. For example, a residential unit in Hudson Yards; a new developed area with designer sky-scrapers, top-notch shopping centers and new highline gardens was sold at a median value of $4.7 million while Inwood; a low income working-class neighborhood, the median cost was $0.4 million.

| | Hood | PricePerUnit |
|---|---|---|
| 0 | Central Harlem | 0.710000 |
| 1 | Chelsea | 1.195000 |
| 2 | Chinatown | 1.059500 |
| 3 | Civic Center | 2.466200 |
| 4 | Clinton | 0.871500 |
| 5 | East Harlem | 0.701500 |
| 6 | East Village | 0.820000 |
| 7 | Financial District | 1.022250 |
| 8 | Flatiron | 1.675000 |
| 9 | Gramercy | 1.225000 |
| 10 | Greenwich Village | 1.400000 |
| 11 | Hudson Yards | 4.666812 |
| 12 | Inwood | 0.407043 |
| 13 | Little Italy | 2.825000 |
| 14 | Lower East Side | 1.406496 |
| 15 | Manhattan Valley | 0.735000 |
| 16 | Midtown | 0.990000 |
| 17 | Morningside Heights | 0.627500 |
| 18 | Murray Hill | 0.787500 |
| 19 | Roosevelt Island | 1.150000 |
| 20 | Soho | 2.250000 |
| 21 | Tribeca | 2.940000 |
| 22 | Upper East Side | 1.145000 |
| 23 | Upper West Side | 1.225000 |
| 24 | Washington Heights | 0.590147 |

*Figure 9- Median cost per residential unit by neighborhood*

Manhattan neighborhood in this analysis were plotted using Folium and the latitude and longitude coordinated were taken from Foursquare database based on the neighborhood name.

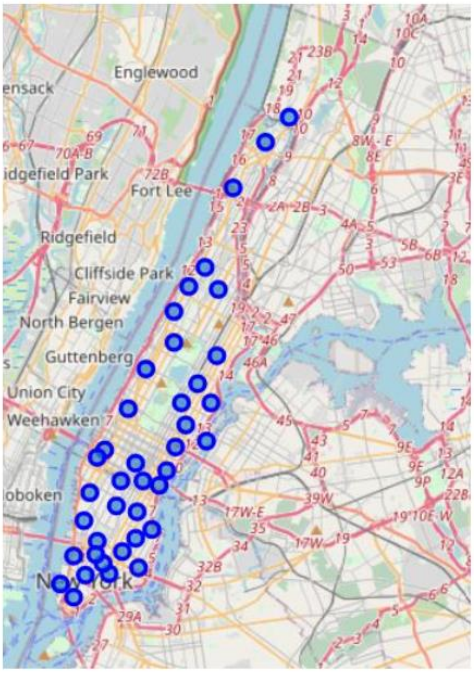
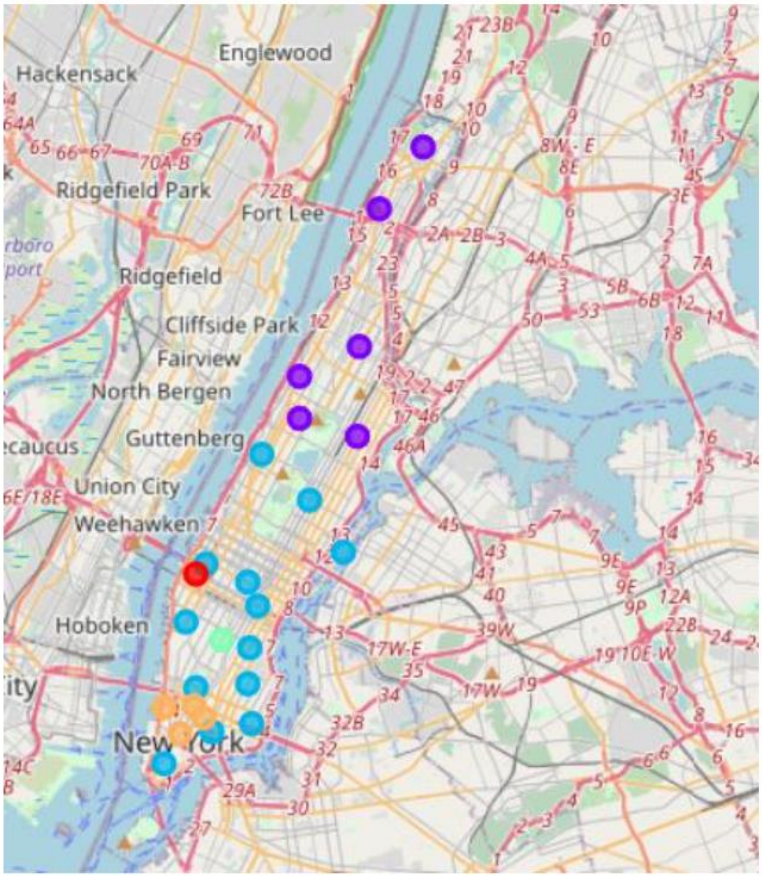



*Figure 10 Manattan neighborhoods*

The median cost of housing per neighborhood from Figure 9 was binned and classified into 5 clusters:

| Cluster and color | Median price per unit (in million dollars) |
|---|---|
| 1, purple | < 0.75 |
| 2, blue | <1.5 |
| 3, cyan | <2.25 |
| 4, orange | < 3 |
| 5, red | > 3 |

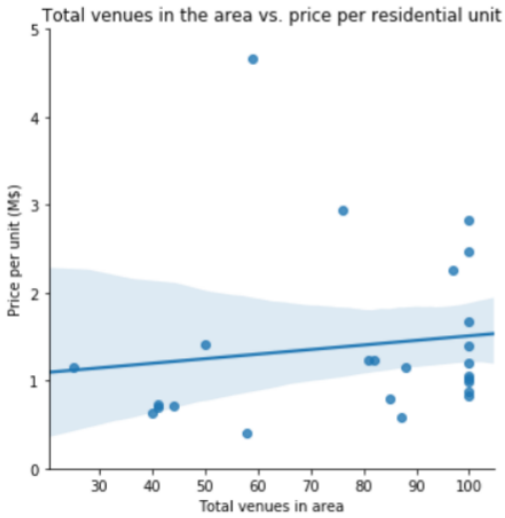
Figure 11 shows Manhattan map with the residential median cost clusters.




As seen from the images above, it is cheaper to buy a residential unit in Manhattan than in downtown or mid-town.

## Discussion continued

As indicated in data section; 2 datasets were used in this work. In this section, the total number of venues in a neighborhood was considered. Foursquare is limited to maximum of 100 venues per neighborhood. The venue category was also collected in addition to its location and name. Among the venues we can find restaurants, bars, cafes, parks, museums, theatres and more.




Furthermore, the venues data from Foursqaure was normalized in a way that every data entry was categorized by the venue type and given a 1 or 0 to quantify the category type. Then the data was grouped by the neighborhood and the average was taken. As a result if more of a certain neighborhood the weighted umber (frequency) will be larger. For example, figured 13 shows the most common venues in Central Harlem and in Chelsea. Whereas African restaurants are the most common venues in Central Harlen while coffee shops and art galleries are most common in Chelsea.

```
----Central Harlem----
                  venue  freq
0    African Restaurant  0.07
1     French Restaurant  0.05
2   American Restaurant  0.05
3  Gym / Fitness Center  0.05
4    Chinese Restaurant  0.05


----Chelsea----
                venue  freq
0         Coffee Shop  0.08
1         Art Gallery  0.08
2      Ice Cream Shop  0.04
3  American Restaurant  0.03
4                Café  0.03
```

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|
| 0 | Central Harlem | African Restaurant | Chinese Restaurant | French Restaurant | Gym / Fitness Center | Cosmetics Shop |
| 1 | Chelsea | Coffee Shop | Art Gallery | Ice Cream Shop | Café | Bakery |
| 2 | Chinatown | Chinese Restaurant | Bakery | Cocktail Bar | Optical Shop | Spa |
| 3 | Civic Center | Coffee Shop | Hotel | Cocktail Bar | Spa | French Restaurant |
| 4 | Clinton | Theater | Coffee Shop | Gym / Fitness Center | Italian Restaurant | Wine Shop |

k-means method was used to cluster the neighborhoods based on the 5 most common venues in each neighborhood. Similar neighborhoods with similar frequencies of venues types will be clustered together while dissimilar neighborhoods will be positioned in a different cluster. The data was combined with the median residential unit purchase cost

data frame. Each neighborhood has two types o classification: median residential purchase price per unit classification (1-5) and venues type cluster (0-4). Figure 14 shows the data frame after the 2 different classification of each neighborhood.

| | Hood | PricePerUnit | Borough | Neighborhood_1 | Latitude | Longitude | Price Cluster | Venue | Neighborhood_2 | Venue Cluster | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Central Harlem | 0.710000 | Manhattan | Central Harlem | 40.815976 | -73.943211 | 1 | 44 | Central Harlem | 3.0 | African Restaurant | Chinese Restaurant | French Restaurant |
| 1 | Chelsea | 1.195000 | Manhattan | Chelsea | 40.744035 | -74.003116 | 2 | 100 | Chelsea | 3.0 | Coffee Shop | Art Gallery | Ice Cream Shop |
| 2 | Chinatown | 1.059500 | Manhattan | Chinatown | 40.715618 | -73.994279 | 2 | 100 | Chinatown | 3.0 | Chinese Restaurant | Bakery | Cocktail Bar |
| 3 | Civic Center | 2.466200 | Manhattan | Civic Center | 40.715229 | -74.005415 | 4 | 100 | Civic Center | 2.0 | Coffee Shop | Hotel | Cocktail Bar |
| 4 | Clinton | 0.871500 | Manhattan | Clinton | 40.759101 | -73.996119 | 2 | 100 | Clinton | 2.0 | Theater | Coffee Shop | Gym / Fitness Center |
| 5 | East Harlem | 0.701500 | Manhattan | East Harlem | 40.792249 | -73.944182 | 1 | 41 | East Harlem | 0.0 | Mexican Restaurant | Bakery | Thai Restaurant |
| 6 | East Village | 0.820000 | Manhattan | East Village | 40.727847 | -73.982226 | 2 | 100 | East Village | 3.0 | Bar | Mexican Restaurant | Coffee Shop |
| 7 | Financial District | 1.022250 | Manhattan | Financial District | 40.707107 | -74.010665 | 2 | 100 | Financial District | 2.0 | Coffee Shop | American Restaurant | Pizza Place |
| 8 | Flatiron | 1.675000 | Manhattan | Flatiron | 40.739673 | -73.990947 | 3 | 100 | Flatiron | 2.0 | Gym / Fitness Center | Café | Italian Restaurant |

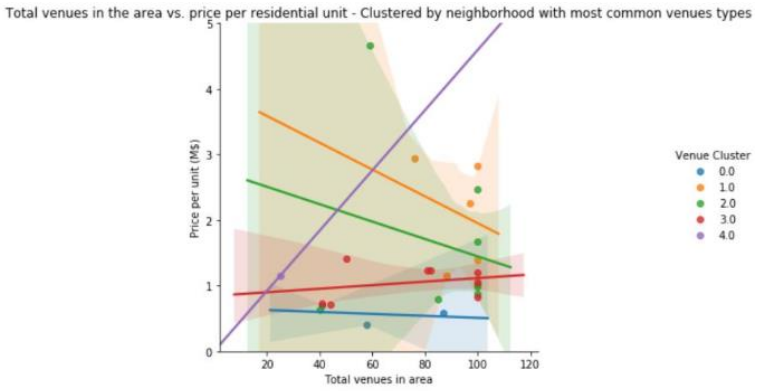



*Figure 15 total venues in area vs. price per residential unit grouped by venues types clusters*

Lastly, figure 15 shows the total number of units in a neighborhood vs. the median residential price per unit in Manhattan's neighborhood, clustered by the venues cluster. Neighborhoods with similar venues share the same cluster.

## Trends/possible conclusions

Through the data described in the pages above, the following trends can be concluded:

- Neighborhoods with venue cluster 0 typically have low end housing which are cheaper to buy. The cluster 0 neighborhood are location in north Manhattan which has more affordable housing and typically has low-income families.
- Neighborhoods with venue cluster 1 has mid-range housing that a relatively more expensive than cluster 0 yet all of them offer similar types of venues in a similar price-point.

- Neighborhoods with venue cluster 2 have larger price variation across the neighborhoods. In other words, if someone is interested in venues types that cluster 2 has to offer, one can choose to buy an apartment in a cheaper area that offers similar types of venues
- Venue cluster 1 neighborhood are high-end neighborhoods that is reflected in high-cost housing with a median cost of $ 2.11 million.

| | Hood | PricePerUnit | Borough | Neighborhood_1 | Latitude | Longitude | Price Cluster | Venue | Neighborhood_2 | Venue Cluster | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | East Harlem | 0.701500 | Manhattan | East Harlem | 40.792249 | -73.944182 | 1 | 41 | East Harlem | 0.0 | Mexican Restaurant | Bakery | Thai Restaurant | Deli / Bodega | Latin American Restaurant |
| 12 | Inwood | 0.607043 | Manhattan | Inwood | 40.867684 | -73.921210 | 1 | 58 | Inwood | 0.0 | Mexican Restaurant | Lounge | Restaurant | Bakery | Café |
| 24 | Washington Heights | 0.590147 | Manhattan | Washington Heights | 40.851903 | -73.936900 | 1 | 87 | Washington Heights | 0.0 | Café | Deli / Bodega | Bakery | Mobile Phone Shop | Latin American Restaurant |

*Figure 17. Venue Cluster 0 neighborhoods*

| | Hood | PricePerUnit | Borough | Neighborhood_1 | Latitude | Longitude | Price Cluster | Venue | Neighborhood_2 | Venue Cluster | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | Greenwich Village | 1.400 | Manhattan | Greenwich Village | 40.726933 | -73.999914 | 2 | 100 | Greenwich Village | 1.0 | Italian Restaurant | Café | Sushi Restaurant | Gym | Comedy Club |
| 13 | Little Italy | 2.825 | Manhattan | Little Italy | 40.719324 | -73.997305 | 4 | 100 | Little Italy | 1.0 | Italian Restaurant | Bakery | Mediterranean Restaurant | Spa | Pizza Place |
| 20 | Soho | 2.250 | Manhattan | Soho | 40.722184 | -74.000657 | 4 | 97 | Soho | 1.0 | Italian Restaurant | Sandwich Place | Mediterranean Restaurant | Clothing Store | Coffee Shop |
| 21 | Tribeca | 2.940 | Manhattan | Tribeca | 40.721522 | -74.010683 | 4 | 76 | Tribeca | 1.0 | Italian Restaurant | American Restaurant | Park | Wine Bar | Greek Restaurant |
| 22 | Upper East Side | 1.145 | Manhattan | Upper East Side | 40.775639 | -73.960508 | 2 | 88 | Upper East Side | 1.0 | Italian Restaurant | Coffee Shop | Bakery | Gym / Fitness Center | Yoga Studio |

*Figure 18. Venue Cluster 1 neighborhoods*

| | Hood | PricePerUnit | Borough | Neighborhood_1 | Latitude | Longitude | Price Cluster | Venue | Neighborhood_2 | Venue Cluster | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | Civic Center | 2.466200 | Manhattan | Civic Center | 40.715229 | -74.005415 | 4 | 100 | Civic Center | 2.0 | Coffee Shop | Hotel | Cocktail Bar | Spa | French Restaurant |
| 4 | Clinton | 0.871500 | Manhattan | Clinton | 40.759101 | -73.996119 | 2 | 100 | Clinton | 2.0 | Theater | Coffee Shop | Gym / Fitness Center | Italian Restaurant | Wine Shop |
| 7 | Financial District | 1.022250 | Manhattan | Financial District | 40.707107 | -74.010665 | 2 | 100 | Financial District | 2.0 | Coffee Shop | American Restaurant | Pizza Place | Café | Italian Restaurant |
| 8 | Flatiron | 1.675000 | Manhattan | Flatiron | 40.739673 | -73.990947 | 3 | 100 | Flatiron | 2.0 | Gym / Fitness Center | Café | Italian Restaurant | Mediterranean Restaurant | Gym |
| 11 | Hudson Yards | 4.666812 | Manhattan | Hudson Yards | 40.756658 | -74.000111 | 5 | 59 | Hudson Yards | 2.0 | Hotel | Italian Restaurant | Gym / Fitness Center | American Restaurant | Coffee Shop |
| 16 | Midtown | 0.990000 | Manhattan | Midtown | 40.754691 | -73.981669 | 2 | 100 | Midtown | 2.0 | Hotel | Coffee Shop | Bakery | Theater | Cuban Restaurant |
| 17 | Morningside Heights | 0.627500 | Manhattan | Morningside Heights | 40.808000 | -73.963896 | 1 | 40 | Morningside Heights | 2.0 | Park | American Restaurant | Coffee Shop | Bookstore | Sandwich Place |
| 18 | Murray Hill | 0.787500 | Manhattan | Murray Hill | 40.748303 | -73.978332 | 2 | 85 | Murray Hill | 2.0 | Hotel | Sandwich Place | Coffee Shop | Gym / Fitness Center | Japanese Restaurant |

*Figure 19 . Venue Cluster 2 neighborhoods*

| | Hood | PricePerUnit | Borough | Neighborhood_1 | Latitude | Longitude | Price Cluster | Venue | Neighborhood_2 | Venue Cluster | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Central Harlem | 0.710000 | Manhattan | Central Harlem | 40.815976 | -73.943211 | 1 | 44 | Central Harlem | 3.0 | African Restaurant | Chinese Restaurant | French Restaurant | Gym / Fitness Center | Cosmetics Shop |
| 1 | Chelsea | 1.195000 | Manhattan | Chelsea | 40.744035 | -74.003116 | 2 | 100 | Chelsea | 3.0 | Coffee Shop | Art Gallery | Ice Cream Shop | Cafe | Bakery |
| 2 | Chinatown | 1.059500 | Manhattan | Chinatown | 40.715618 | -73.994279 | 2 | 100 | Chinatown | 3.0 | Chinese Restaurant | Bakery | Cocktail Bar | Optical Shop | Spa |
| 6 | East Village | 0.820000 | Manhattan | East Village | 40.727847 | -73.982226 | 2 | 100 | East Village | 3.0 | Bar | Mexican Restaurant | Coffee Shop | Cocktail Bar | Pizza Place |
| 9 | Gramercy | 1.225000 | Manhattan | Gramercy | 40.737210 | -73.981376 | 2 | 82 | Gramercy | 3.0 | Bagel Shop | Coffee Shop | Bar | Pizza Place | American Restaurant |
| 14 | Lower East Side | 1.406496 | Manhattan | Lower East Side | 40.717807 | -73.980890 | 2 | 50 | Lower East Side | 3.0 | Chinese Restaurant | Art Gallery | Pharmacy | Cafe | Cocktail Bar |
| 15 | Manhattan Valley | 0.735000 | Manhattan | Manhattan Valley | 40.797307 | -73.964286 | 1 | 41 | Manhattan Valley | 3.0 | Coffee Shop | Yoga Studio | Mexican Restaurant | Bar | Pizza Place |
| 23 | Upper West Side | 1.225000 | Manhattan | Upper West Side | 40.787658 | -73.977059 | 2 | 81 | Upper West Side | 3.0 | Italian Restaurant | Bar | Dessert Shop | Indian Restaurant | Wine Bar |

*Figure 20 . Venue Cluster 3 neighborhoods*

| | Hood | PricePerUnit | Borough | Neighborhood_1 | Latitude | Longitude | Price Cluster | Venue | Neighborhood_2 | Venue Cluster | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 19 | Roosevelt Island | 1.15 | Manhattan | Roosevelt Island | 40.76216 | -73.949168 | 2 | 25 | Roosevelt Island | 4.0 | Park | Pizza | School | Gym | Greek Restaurant |

*Figure 21. Venue Cluster 4 neighborhoods*

## Summary

The focus/research question of this project will be to find the correlation between the type and number of venues in a specific neighborhood in Manhattan to its residential real estate cost. In this work, I investigated the correlation between residential real estate median cost per unit vs. the type of venues that are located in the neighborhoods of Manhattan.

As the first and most important step, the impact of different residential properties types on the median purchase point was shown along with the correlation between the size of the real estate to its cost. Furthermore, in Manhattan, every square feet of residential unit cost roughly $2000. The impact of elevators on the cost of real estate value and the difference in price between co-op and condo was also explained and justified in the report.

Secondly, the neighborhoods by their mean residential real-estate cost were analyzed and the fact that north Manhattan is cheaper than mid-town or down-town was also justified. In one of the report sections, it was also mad clear that the new developing areas of Hudson Yards have the highest purchasing costs.

Additionally, a possible connection between the number of venue sin neighborhood to the cost of its housing was also shown and were classified into their venue types using k-means.

Finally, suggestions were made based on the most common venue types and the price point of housing. If someone per se would like to live in an area with certain venues this person can decide based on the price point. Additionally, some clusters like cluster 2 have

had high variations in prices meaning that one can find cheaper houses and the similar surrounding venue experience with lesser value.