COURSERA CAPSTONE PROJECT – REAL ESTATE PURCHASE PRICE CORRELATION TO NEIGHBOURHOODS

SMRITI SHARMA

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.

INTRODUCTION

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 in 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.

INTRODUCTION - SUCCESS CRITERIA

Success Criteria: The success criteria of the project will be a price estimation to a stakeholder based on neighbourhoods.

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.

DATA

DATA PT.1

In this report, an open source data of New York property purchase through 2019 will be used.

The data is used by New York City tax collector to estimate the taxes per purchase.

The data includes 16473 property entries of purchases through 2019 in Manhattan only.

The dataset included the neighborhood name, borough, building class category, tax classes, block, lot, address, zip code, number of residential units under this purchase, number of commercial units under this purchase, total units, square feet, gross square feet, year built, sale price and sale date.

Foursquare open source dataset was used for collecting the venues per neighborhood. The venues per neighborhood were limited to 100 venues in the radius of 500 meters from geo-locators of the neighborhood.

The dataset incudes both residential commercial, lots, garages, and hospitals although only the residential properties will be considered.

Additionally, a few neighborhoods were consolidated based on the available neighborhood data on Foursquare and the NYC property purchase data.

DATA PT. 2

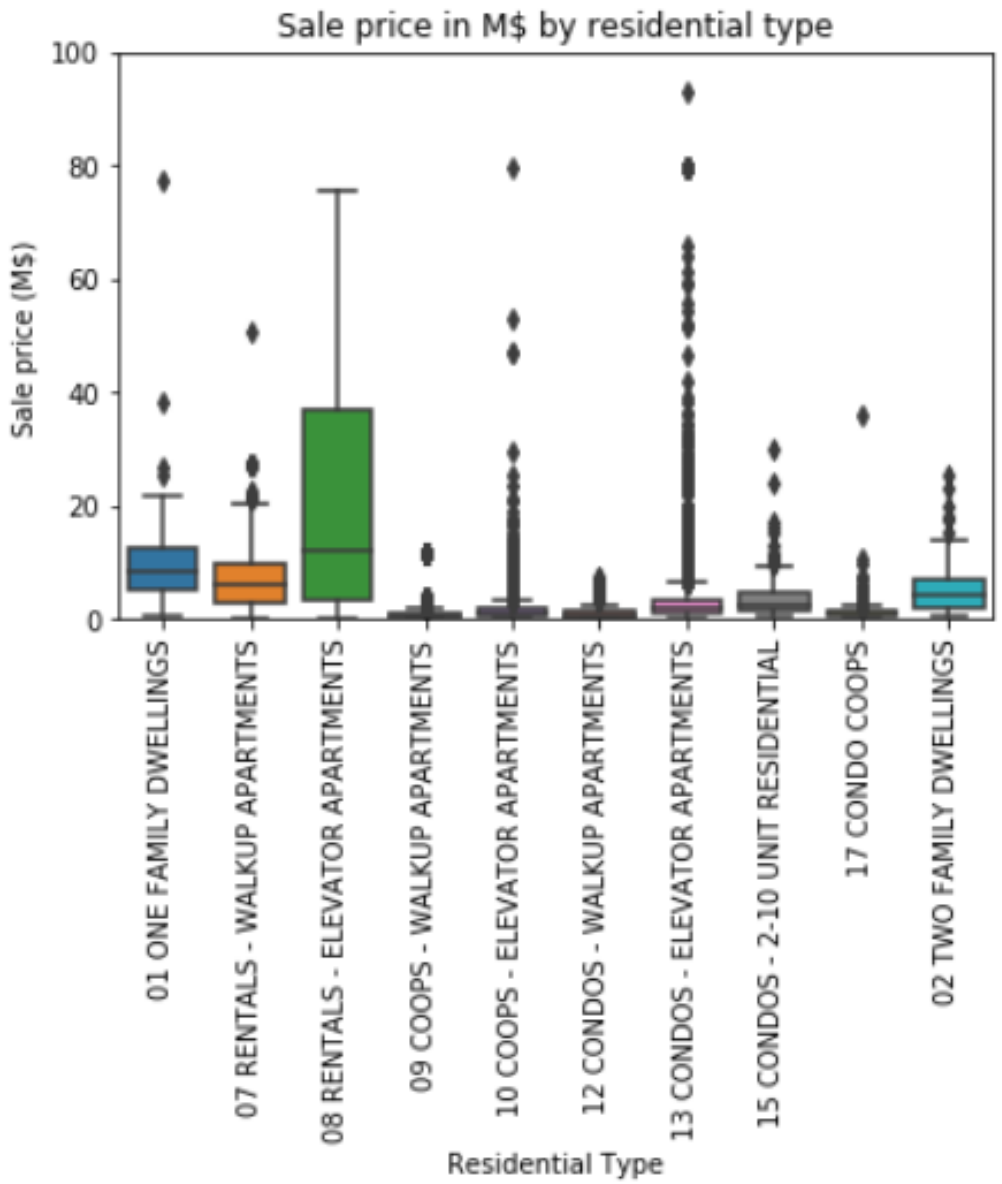
1. Numeric data such as sale price, and square feet will be converted to integers instead of. Sale price below $100,000 will be dropped
2. Non-residential units will not be considered – based on building class category
3. Sale price will be converted to million dollars for ease
4. If the number of units are missing; one unit will be assumed
5. Total sale price will be divided by number of units to reflect the cost per residential unit
6. The total price reflects the price of the entire building in a multi-unit building.
7. The cost a unit in a co-op reflects one-unit share in a co-op
8. Square feet per unit will be calculated where square feet data is available.

STEPS TO PROCESS DATA

	BOROUGH	Hood	BUILDING CLASS CATEGORY	ZIP CODE	TOTAL UNITS	LAND SQUARE FEET	GROSS SQUARE FEET	YEAR BUILT	SALE DATE	SALE PRICE
0	1	East Village	01 ONE FAMILY DWELLINGS	10009	2.0	2,090	3,680	1940.0	7/24/2019	3,200,000
1	1	East Village	01 ONE FAMILY DWELLINGS	10009	1.0	987	2,183	1860.0	9/25/2019	0
2	1	East Village	02 TWO FAMILY DWELLINGS	10009	2.0	1,510	4,520	1900.0	7/22/2019	0
3	1	East Village	03 THREE FAMILY DWELLINGS	10009	3.0	2,430	3,600	1899.0	4/30/2019	6,300,000
4	1	East Village	03 THREE FAMILY DWELLINGS	10009	3.0	2,375	5,110	1939.0	10/24/2019	0

EXAMPLE OF CLEAN UP

- In this work- neighborhood, sale price, building class category, total unit and gross square feet is considered
- Sale price =0 was dropped from the dataset.
- Only residential properties is considered. Commercial property sales were not considered.

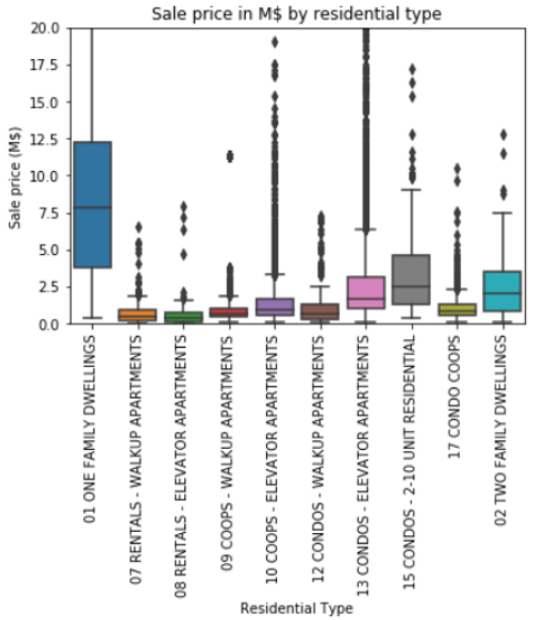


METHODOLOGY PT.1

- Rental properties show the most expensive sale prices
- Plausible cause: multi-units under the same building
- Potential solution: normalize the sale prices by the number of untis

SALE PRICE PER UNIT VS. RESIDENTIAL TYPE

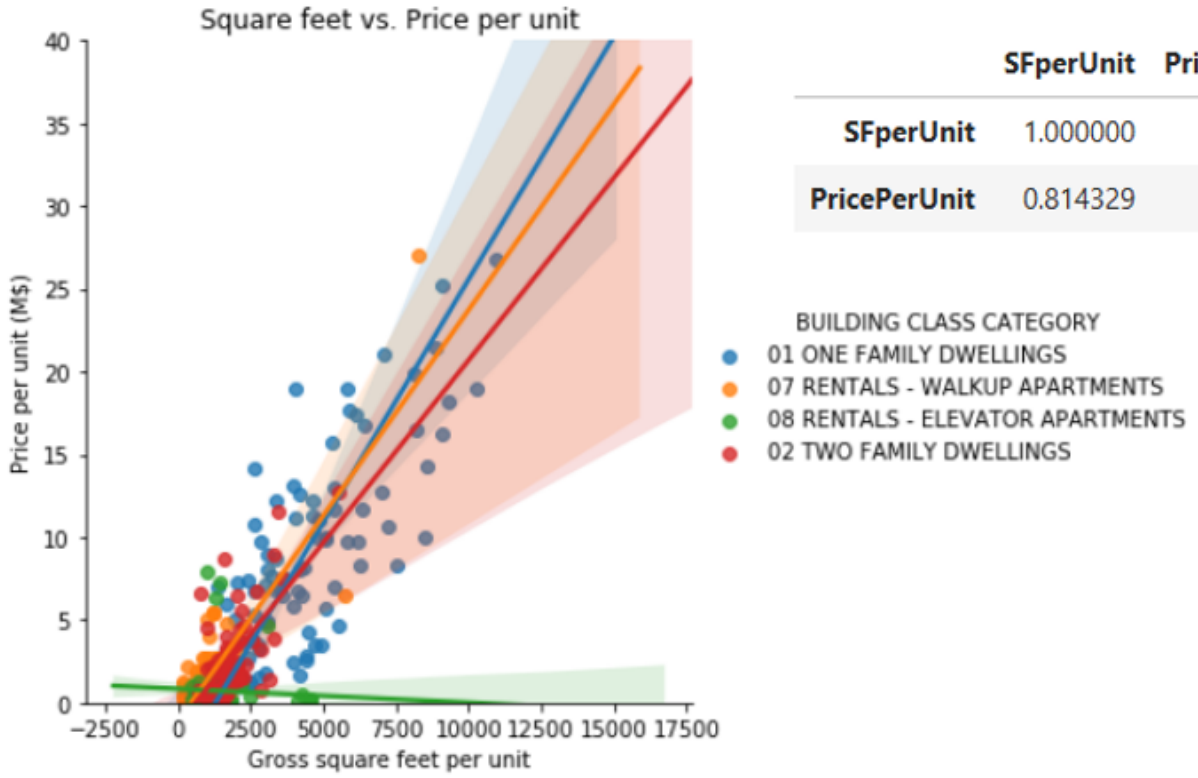
- Single family and 2 family dwelling is the most expensive type and therefore probably why it is the least common as well.
- Elevator condos and co-ops are worth more than non-elevator ones; adding roughly $200,000 (co-op) to its value
- Elevator did not however add value to any sort of rental properties



	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

SQUARE FEET VS. REAL ESTATE PRICE PER UNIT

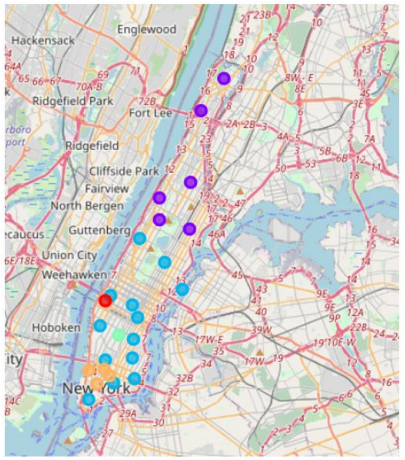
- There is a strong linear correlation between the size of the property to its price with r^2 0.814
- Slope = 0.002 (implying that 1 square feet in Manhattan costs roughly 2000$)



	SFperUnit	PricePerUnit
SFperUnit	1.000000	0.814329
PricePerUnit	0.814329	1.000000

```
PricePerUnit=Slope*SFPerUnit + Intercept
Intercept: [-1.49311099]
Slope: [[0.00237466]]
```

MANHATTAN REAL ESTATE PRICE BY CLUSTERS

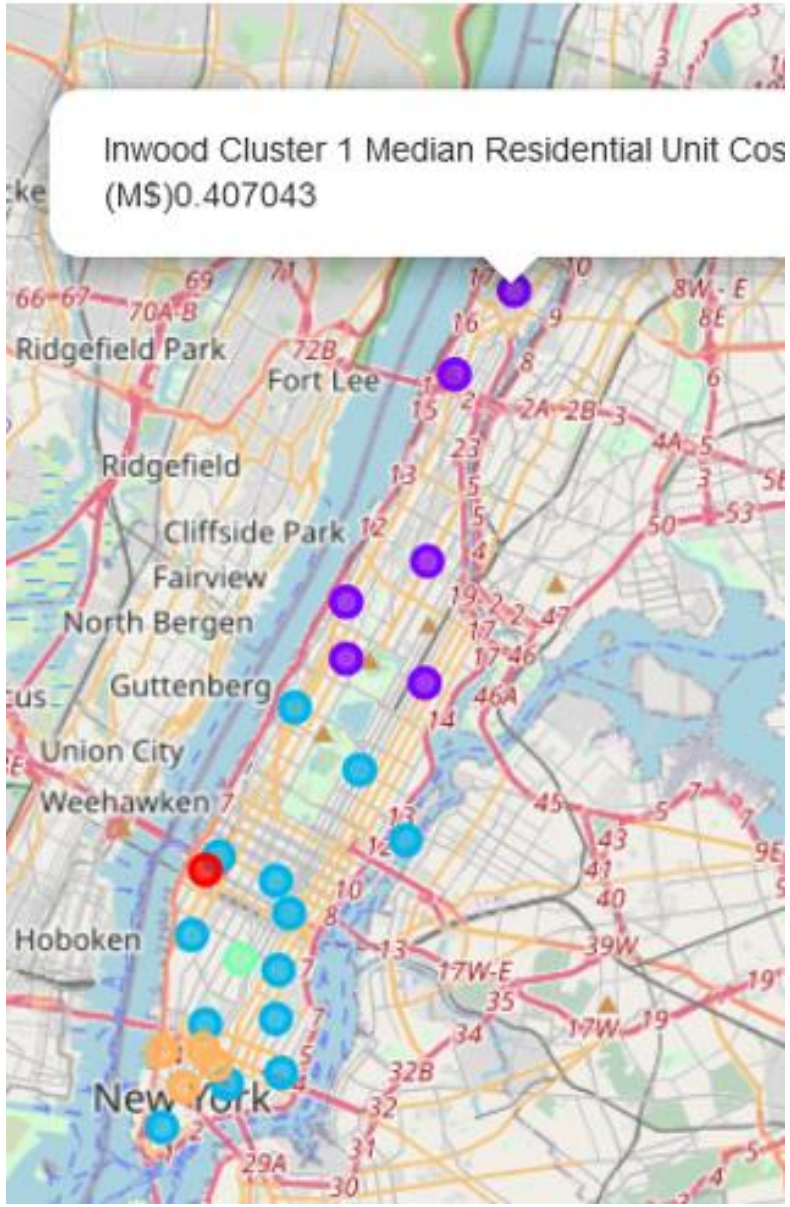


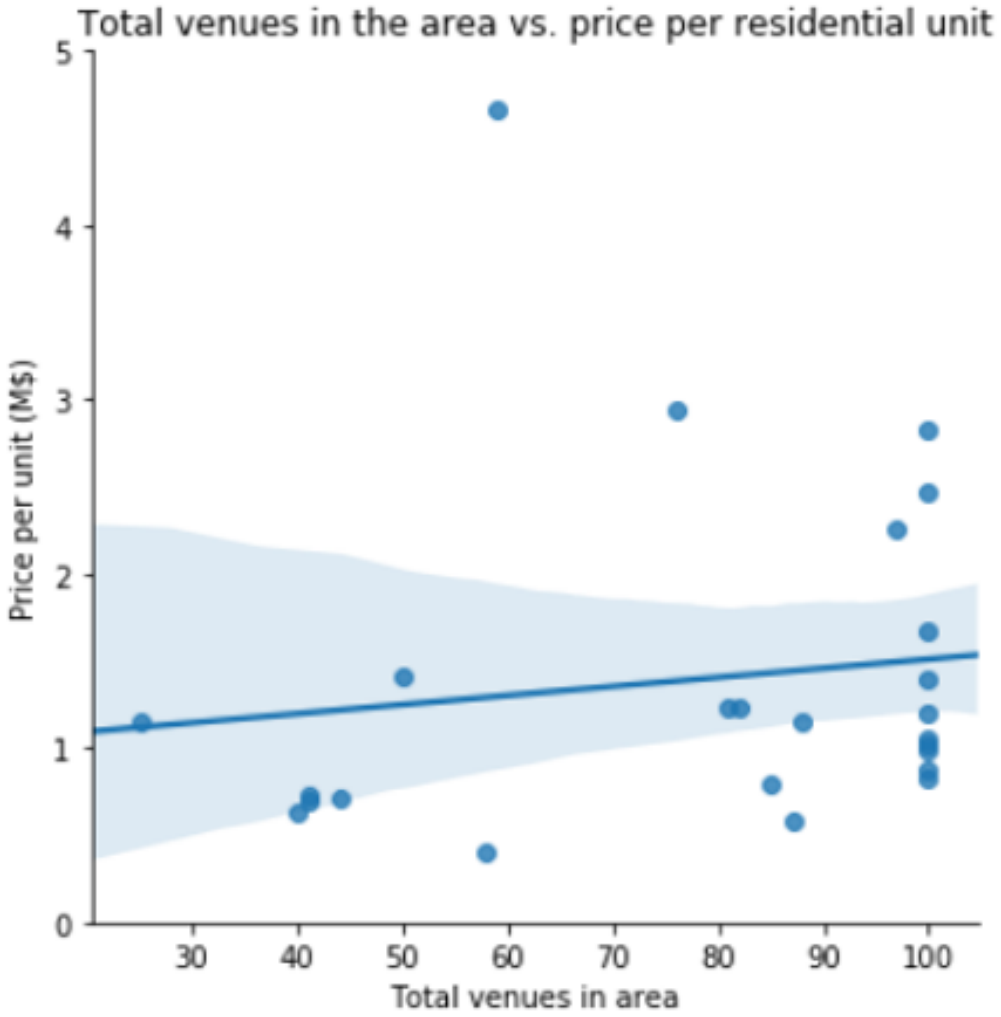
Cluster and color	Median price per unit (in million dollars)
1, purple	< 0.75
2, blue	<1.5
3, cyan	<2.25
4, orange	< 3
5, red	> 3

- North Manhattan: more affordable housing, residents are typically lower-class working income families
- Financial district area: expensive high-end housing. Typically residents work in the surrounding financial institutions

MORE CLUSTERED MAP IMAGES

- There are 5 clusters by median price, check table in previous slide



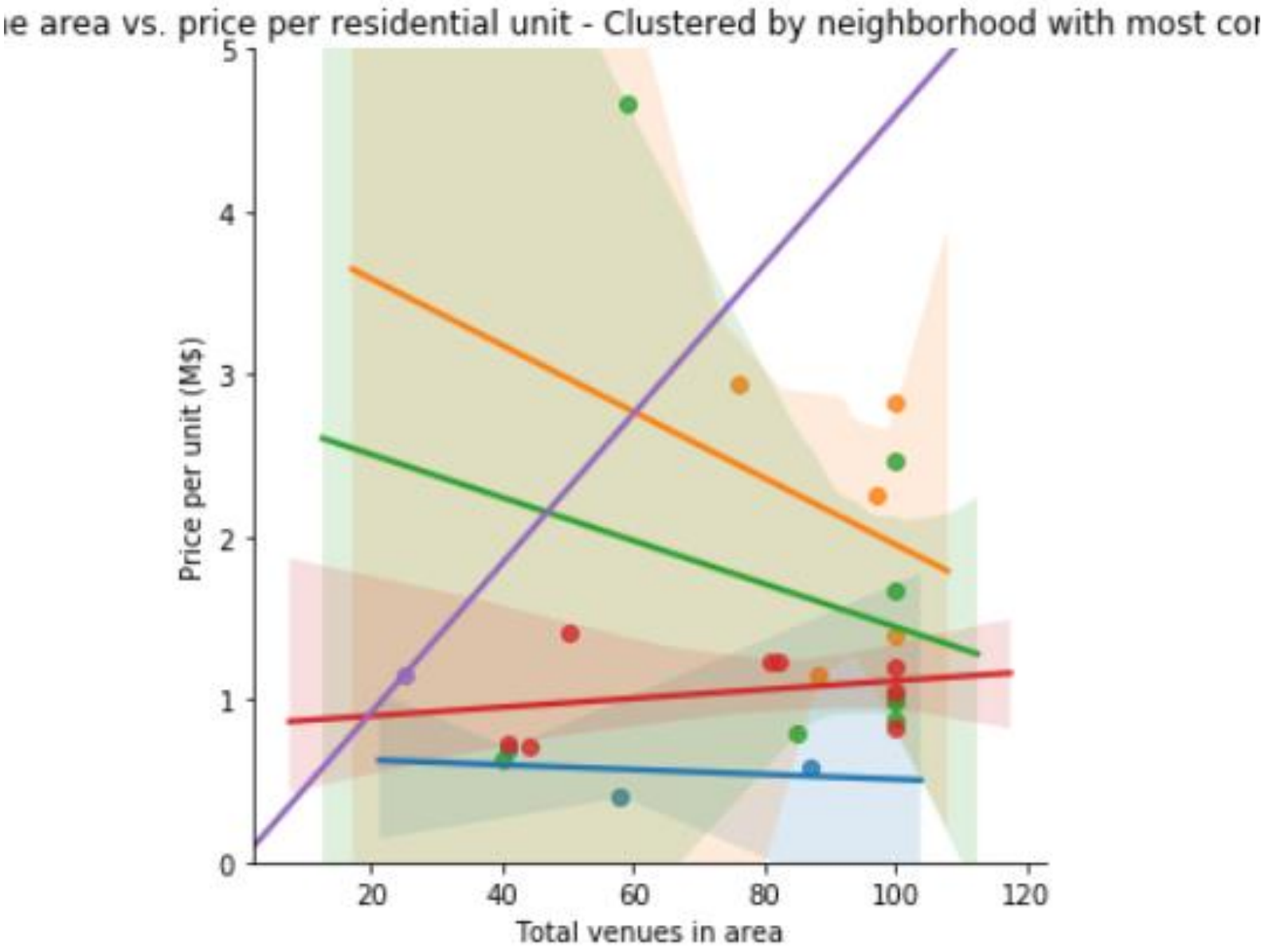


CORRELATION BETWEEN TOTAL NUMBER OF VENUES IN NEIGHBORHOOD TO ITS HOUSING COST

- Foursquare was used to extract the total amount of venues per neighborhood
- Slight correlation between the two factor.

NEIGHBOURHOOD VENUE TYPE CLUSTERING VS. HOUSING COST

- Second type of clustering in this report
- Top 5 frequent venues in each neighbourhood were considered and neighbourhood were clustered accordingly.
- Clear separation in price vs. venue type cluster



248]:

	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	0.407043	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

Let's checkout the mean of price per unit by neighborhood cluster type

CLUSTER 0

[250]:

	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

CLUSTER 1

251]:

	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

CLUSTER 2

[252]:

	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	Café	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	Café	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

CLUSTER 3

[253]:

	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	Plaza	School	Gym	Greek Restaurant

CLUSTER 4

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 .

CONCLUSION/DISCUSSION

THANK YOU FOR PAYING ATTENTION!

THE END