

Manhattan Real Estate Purchase Price correlation to Neighborhoods and surrounding Venues

Data description:

In this work an open source data of NYC property purchase through 2019 was used. The data is used by NYC tax collector to estimate the taxes per purchase. The data included 16473 property entries of purchases through 2019 in Manhattan only. The dataset included the neighborhood name, borough, building class category, tax class, block, lot, address, zip code, number of residential units under this purchase, number of commercial units under this purchase, total units, land square feet, gross square feet, year built, sale price and sale date.

In addition, 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 includes both residential commercial, lots, garages and hospitals. For simplicity only the residential properties were considered. In addition, few neighborhoods were consolidated based on the available neighborhood data on Foursquare and the NYC property purchase data.

Fig. 1 shows the top 14 data entries by building class category.

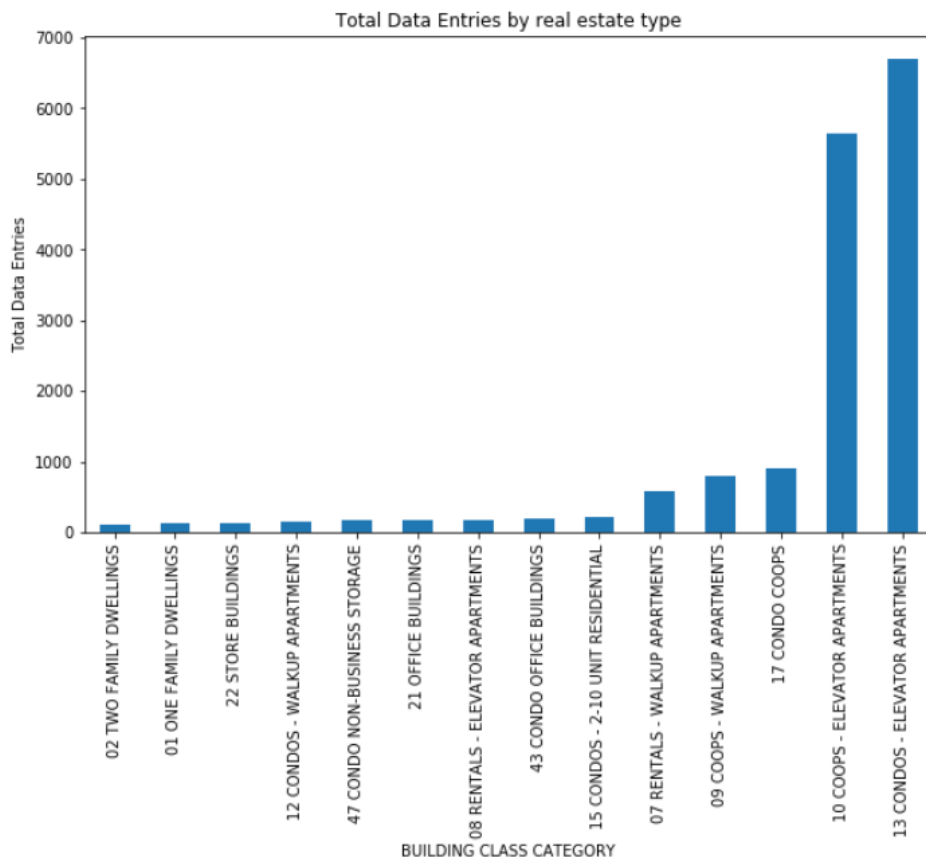


Figure 1 - Total real estate data entries in Manhattan by buliding class category

Figure 2. summarized all the data entries based on building class category for the top 14 building class category. Note the with the borough and hood columns in fact show to total sum of entries in this particular category

	BUILDING CLASS CATEGORY	BOROUGH	Hood	ZIP CODE	TOTAL UNITS	L
0	01 ONE FAMILY DWELLINGS	132	132	132	132	
1	02 TWO FAMILY DWELLINGS	113	113	113	113	
2	07 RENTALS - WALKUP APARTMENTS	585	585	585	585	
3	08 RENTALS - ELEVATOR APARTMENTS	184	184	184	184	
4	09 COOPS - WALKUP APARTMENTS	789	789	789	0	
5	10 COOPS - ELEVATOR APARTMENTS	5649	5649	5649	0	
6	12 CONDOS - WALKUP APARTMENTS	146	146	146	146	
7	13 CONDOS - ELEVATOR APARTMENTS	6689	6689	6689	6490	
8	15 CONDOS - 2-10 UNIT RESIDENTIAL	216	216	216	215	
9	17 CONDO COOPS	896	896	896	0	
10	21 OFFICE BUILDINGS	177	177	177	177	
11	22 STORE BUILDINGS	140	140	140	140	
12	43 CONDO OFFICE BUILDINGS	194	194	194	194	
13	47 CONDO NON-BUSINESS STORAGE	167	167	167	167	

Figure 2-Top data entries sum by building class

Data clean-up and pre-arrangement:

The dataset was pre-arranged. The columns that are not useful for this work were dropped. Fig. 3 shows a sample the real estate dataset before the columns drop and Fig. 4 shows the columns after the columns drop.

	BOROUGH	NEIGHBORHOOD	BUILDING CLASS CATEGORY	TAX CLASS AT PRESENT	BLOCK	LOT	EASE- MENT	BUILDING CLASS AT PRESENT	ADDRESS	APARTMENT NUMBER	...	COMMERCIAL UNITS	TOTAL UNITS	LAND SQUARE FEET	GROSS SQUARE FEET	YEAR BUILT
0	1	ALPHABET CITY	01 ONE FAMILY DWELLINGS	1	376	43	NaN	S1	743 EAST 6TH STREET	NaN	...	1.0	2.0	2,090	3,680	1940.0
1	1	ALPHABET CITY	01 ONE FAMILY DWELLINGS	1	390	61	NaN	A4	189 EAST 7TH STREET	NaN	...	0.0	1.0	987	2,183	1860.0
2	1	ALPHABET CITY	02 TWO FAMILY DWELLINGS	1	404	1	NaN	B9	166 AVENUE A	NaN	...	0.0	2.0	1,510	4,520	1900.0
3	1	ALPHABET CITY	03 THREE FAMILY DWELLINGS	1	377	56	NaN	C0	263 EAST 7TH STREET	NaN	...	0.0	3.0	2,430	3,600	1899.0
4	1	ALPHABET CITY	03 THREE FAMILY DWELLINGS	1	393	9	NaN	C0	604 EAST 11TH STREET	NaN	...	0.0	3.0	2,375	5,110	1939.0

Figure 3. Dataset sample before columns drop

	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

Figure 4. Dataset sample after columns drop

The following actions were done to pre-arrange and clean the dataset:

- Numeric data such as sale price, Land square feet and gross square feet were converted to integers instead of strings and the commas were removed.
- Sale price below 100K\$ was dropped
- Non-residential units were not considered – based on building class category
- Sale price was converted into M\$
- When number of units was missing – 1 unit was assumed
- Total sale price was divided by number of units to reflect the cost per residential unit. The total price reflects the price of the entire building in a multi-units buildings. The cost of a unit in a co-op reflect 1 unit share in a co-op.
- Square feet (SF) per unit was calculated where square feet data was available.

For simplicity, only residential properties data entries were considered. The following property types are available in the following analysis:

- One family dwelling (single family dwelling)
- Two family dwelling
- Condos – Walkup apartment (no elevator)
- Condos – Elevator apartments
- Condos – 2-10 units
- Condo-Co-op
- Co-ops- Elevator apartment
- Co-op- Walkup apartment
- Rentals (housing purchases for rental purposes) -Elevator
- Rentals (walkup apartments)

As a result of the above actions, the dataset included total of 12,232 entries.

```
TrueSale.shape
```

```
(12232, 10)
```