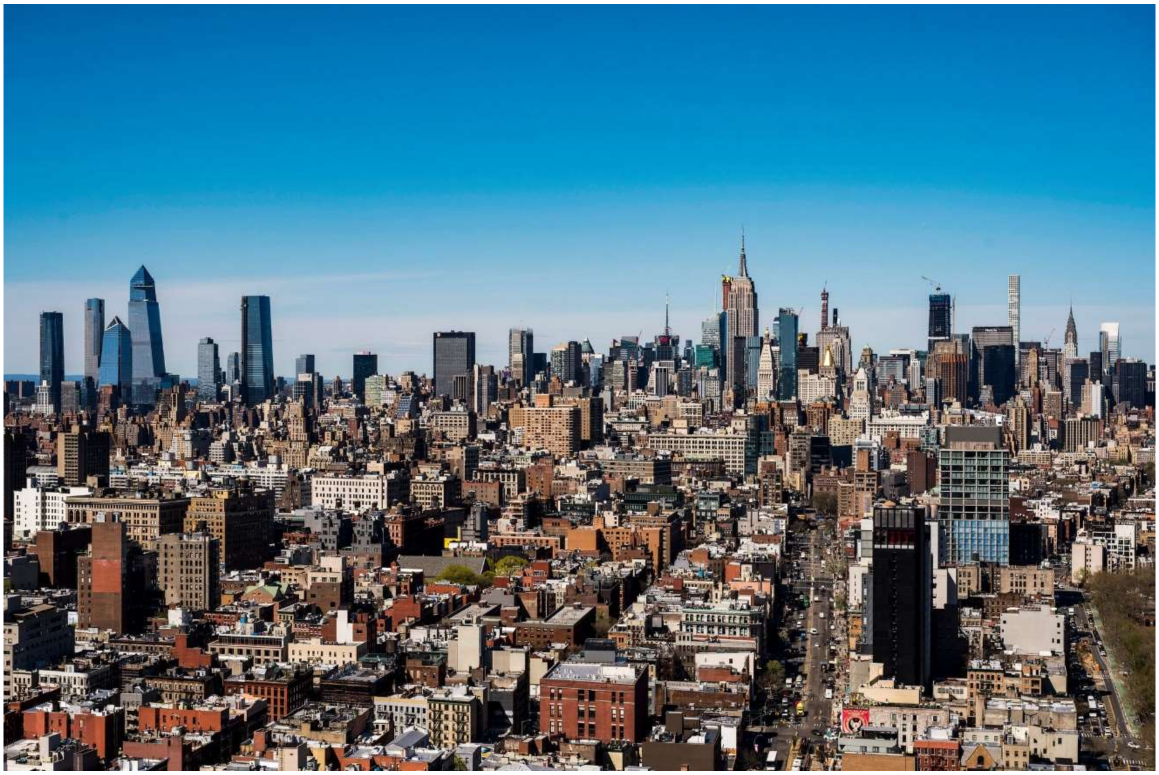

# New York City Real State Analysis

COURSERA CAPSTONE PROJECT

Javier A. Jaime-Serrano | IBM Applied Data Science Capstone | August 20, 2020

# Part 1: Introduction

## BACKGROUND

New York City (NYC) is the most populated city int the United States (US), with an estimated population of more than 8 million distributed over more than 300 square miles.

It is also the largest metropolitan area in the US with more than 20 million people, and is composed of five boroughs: Brooklyn, Queens, Manhattan, the Bronx, and Staten Island.

Real estate is a major part in the city economy, as the total value of all NYU passed the $1 trillion mark in thew 2017 fiscal year with an increase of more than 10% over the previous year. NYC is home to some of the world's most valuable real state. [1]

## BUSINESS PROBLEM

An investment on real state in NYC will come with some challenges, and any potential investor will need to be assured of the value of the investments. This value will come from selecting the best neighborhood for a target amount to be invested.

The best neighborhood will depend mainly of the proximity to the venues preferred by the potential investor. Other factors like crime statistics were not considered, assumed not relevant to the higher market segments analyzed in this report.

The question is: How to select the best neighborhood for a given investment amount and the venues preferences of a potential investor.

[1] New York City (https://en.wikipedia.org/wiki/New_York_City)

# Part 2: Data Acquisition

## DATA SOURCES

In order to segment the neighborhoods of NYC and explore them, we need the first dataset. This dataset was extracted from the NYU Spatial Data Repository [2] and contains the 5 boroughs and the 306 neighborhoods as well as the latitude and longitude coordinates of each neighborhood. This dataset was previously used in the course week 3 lab Segmenting and Clustering Neighborhoods in New York City.

The second dataset was extracted from Zillow Research Data [3] and contains a time series of the Zillow Home Value Index (ZHVI), this is a smoothed, seasonally adjusted measure of the typical home value and market changes across a given region and housing type.

The dataset selected from Zillow Research is for the mid-tier condo/coops (typical value in US dollars for homes that fall within $33^{rd}$ to $67^{th}$ percentile range for a given region). The mid-tier condo/coop ZHVI is assumed to be representative of the average home values of a highly dense region with a low percentage of single-family homes.

The third data set was extracted with the Foursquare Developer Application Programming Interface (API) [4] and includes the venues, categories and location data of the neighborhoods requested by the API within a given radius. This dataset was previously used in the curse week 2 lab Foursquare API.

## DATA CLEANING

Dataset 1 (from NYU) was extracted first in a json file, and filtered to the features key that contains all relevant data, this data was transformed to a pandas dataframe using loop script in Python (Programming language), see figure 1 for the first five rows of the dataframe.

| | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|
| 0 | Bronx | Wakefield | 40.894705 | -73.847201 |
| 1 | Bronx | Co-op City | 40.874294 | -73.829939 |
| 2 | Bronx | Eastchester | 40.887556 | -73.827806 |
| 3 | Bronx | Fieldston | 40.895437 | -73.905643 |
| 4 | Bronx | Riverdale | 40.890834 | -73.912585 |

Figure 1

Dataset 2 (from Zillow) was extracted first in a Comma Separated Values (CSV) file and transferred directly to a dataframe (read csv method) and filtered to show NYC only, not required columns were dropped and remaining columns were renamed keeping the neighborhood name as the key index, see figure 2 for the first 5 rows and the last 10 columns of the dataframe.

| Neighborhood | 2019-09 | 2019-10 | 2019-11 | 2019-12 | 2020-01 | 2020-02 | 2020-03 | 2020-04 | 2020-05 | 2020-06 |
|---|---|---|---|---|---|---|---|---|---|---|
| Upper West Side | 1233923.0 | 1226329.0 | 1228089.0 | 1234873.0 | 1232891.0 | 1228788.0 | 1218444.0 | 1218807.0 | 1214683.0 | 1210707.0 |
| Upper East Side | 929593.0 | 927350.0 | 926257.0 | 927430.0 | 927683.0 | 929749.0 | 925260.0 | 925438.0 | 925215.0 | 933518.0 |
| East New York | 341233.0 | 340846.0 | 340971.0 | 340998.0 | 342076.0 | 343255.0 | 344385.0 | 345332.0 | 345130.0 | 345802.0 |
| Washington Heights | 566566.0 | 560402.0 | 556665.0 | 552960.0 | 548861.0 | 544782.0 | 541350.0 | 538371.0 | 533715.0 | 531615.0 |
| Astoria | 513852.0 | 514333.0 | 513360.0 | 514066.0 | 514298.0 | 514380.0 | 513222.0 | 513961.0 | 514368.0 | 515600.0 |

Figure 2

Dataset 3 (from Foursquare) was extracted first in a json file, flattened (json normalize method), filtered to the venues and their categories, and keeping the location data, see figure 3 for the first five rows of the dataframe.

| | name | categories | lat | lng |
|---|---|---|---|---|
| 0 | The Bar Room at Temple Court | Hotel Bar | 40.711448 | -74.006802 |
| 1 | The Beekman, A Thompson Hotel | Hotel | 40.711173 | -74.006702 |
| 2 | Alba Dry Cleaner & Tailor | Laundry Service | 40.711434 | -74.006272 |
| 3 | Gibney Dance Center Downtown | Dance Studio | 40.713923 | -74.005661 |
| 4 | The Class by Taryn Toomey | Gym / Fitness Center | 40.712753 | -74.008734 |

Figure 3

## FEATURE SELECTION

After cleaning the data, we ended with 3 datasets: the first with 306 neighborhoods and the required location data, the second with 190 neighborhoods and ZHVI 296 months' time series and the third with 10,058 venues, neighborhoods, categories and location data.

The field in common in the 3 datasets is the neighborhood and was selected as the key index to link the 3 datasets. The number of venues and neighborhoods will need to be reduced after the exploratory data analysis.

[2] NYU Spatial Data Repository (https://geo.nyu.edu/catalog/nyu_2451_34572)

[3] Zillow Research Data (https://www.zillow.com/research/data/)

[4] Foursquare Developer (https://developer.foursquare.com/)

## Part 3: Exploratory Data Analysis

### NEIGHBORHOODS

For the first dataset we used folium library to create a map of NYC with a marker for each one of the neighborhoods on the dataset see figure 4 for the resulting map.


Figure 4

In the course week 3 Lab Segmenting and Clustering Neighborhoods in New York City, we used the k-means (machine learning) algorithm from the scikit learn library to segment and cluster NYC neighborhoods by venues similarity. The results from this exercise were inconclusive and were not repeated for this report, see figure 5 for a map with the clusters in different color markers for the Manhattan borough.

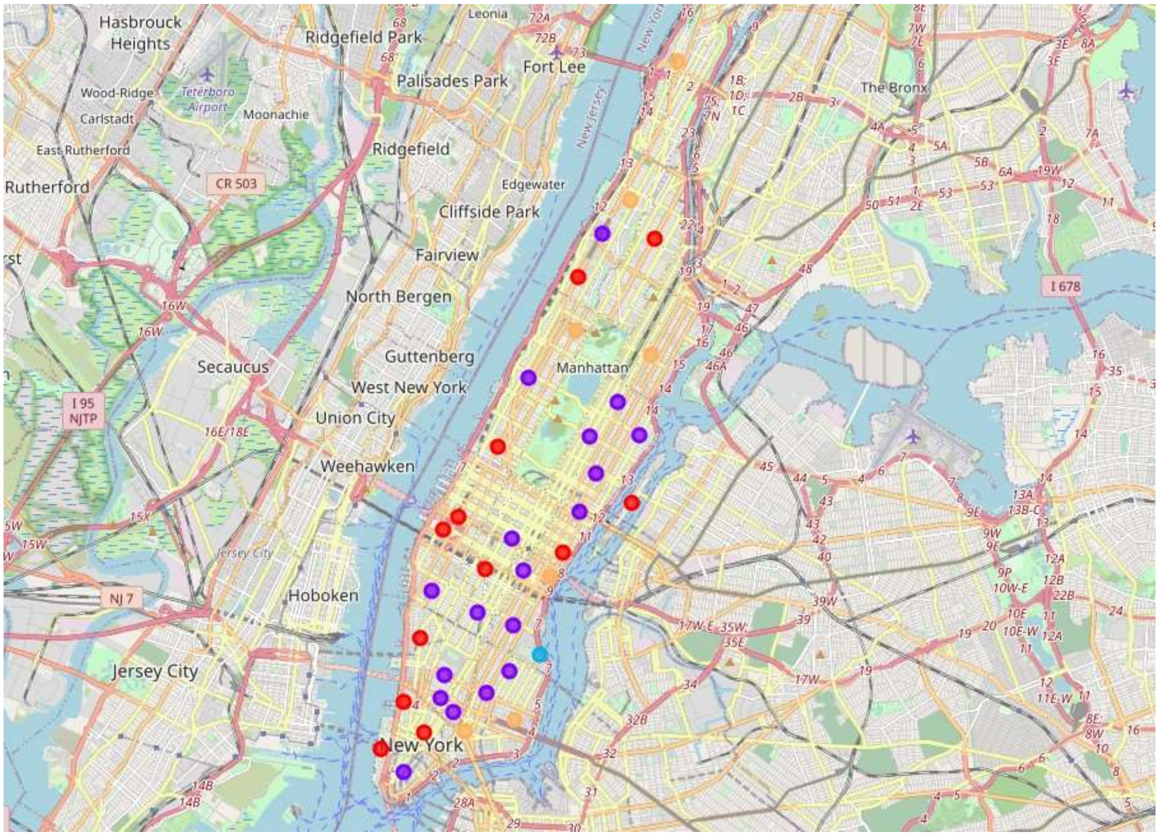


Figure 5

We decided to look for other variables to segment and rank the Neighborhoods.

## VALUE

From the second dataset we used the matplotlib library to create a line plot (graph), and selected the top 30 neighborhoods by the last available month of the ZHVI (Value Index), see figure 6 for a time series of the top 30 neighborhoods by the ZHVI (Value Index).

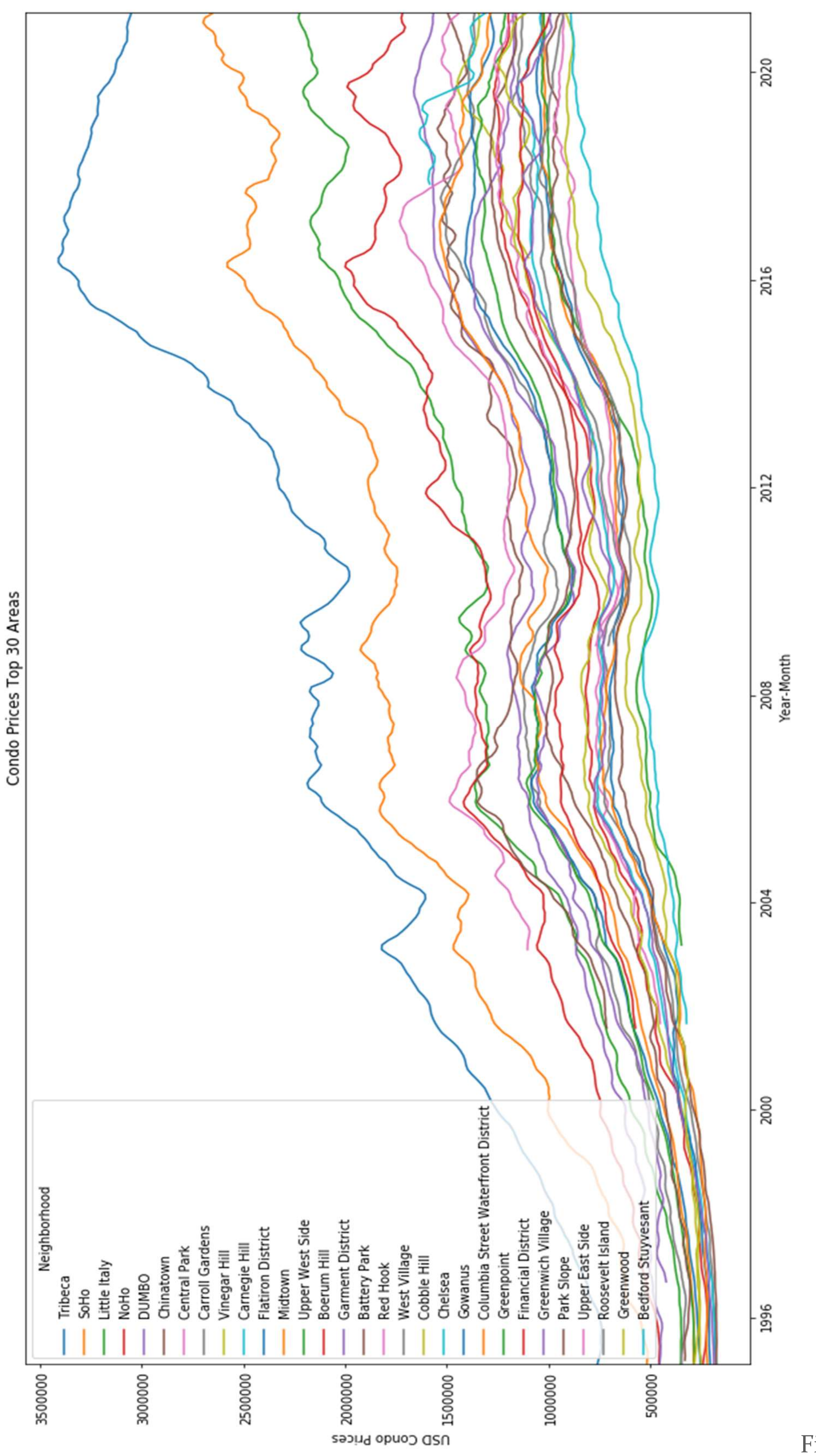


Figure 6

Figure 6

In order to select the best neighborhood and avoid recommending a property in a neighborhood that have declining property values, is important to analyze the trend.

As you can see in figure 6, the most expensive neighborhood (by ZHVI), Tribeca, peaked around 2016 and started a decline, other neighborhood like SoHo or Little Italy continue their trend up, at least until the most recent months, the value index (XHVI) is a lagging indicator and the effect of the COVID-19 Pandemic is not reflected in this values.

The more recent information suggests that the effects of the pandemic will apply a temporary discount of the rental price and create a better investment deals in NYC. The percentage discounts vary by proximity of the unused office space with the biggest discounts in Manhattan according to data from StreetEasy. The biggest rental reductions are in Flatiron (45%), Financial District (42%) and Midtown (40%) neighborhoods [5].

See Figure 7 for Manhattan, Brooklyn, and Queens real estate status as of June 2020.

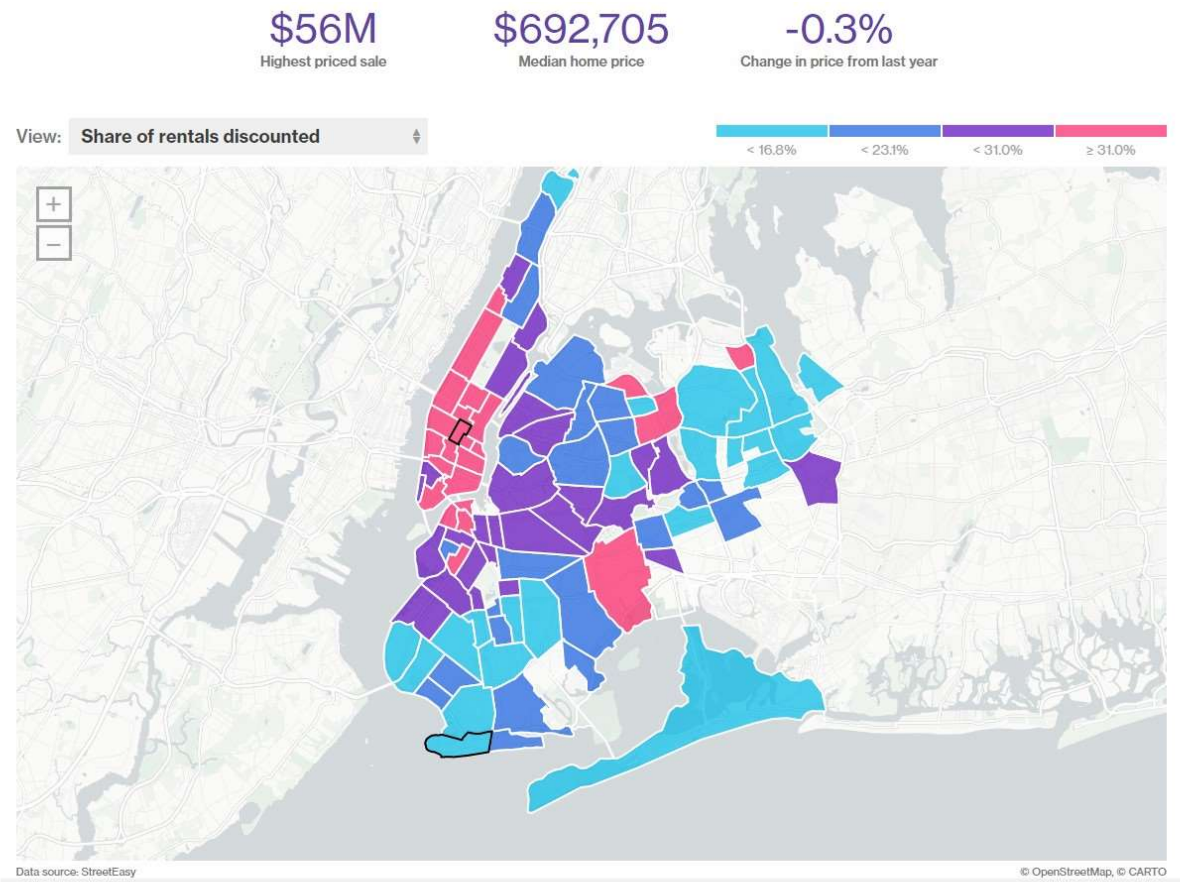


Figure 7

## VENUES

From the third dataset we selected the top 30 neighborhoods by the number of venues, and compared to the neighborhoods in the second dataset, see figure 8 for the top 5 neighborhood by venues (note that foursquare API caps the results to 100 per request).

| | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|
| **Neighborhood** | | | | | | |
| **Murray Hill** | 147 | 147 | 147 | 147 | 147 | 147 |
| **Chelsea** | 105 | 105 | 105 | 105 | 105 | 105 |
| **Lenox Hill** | 100 | 100 | 100 | 100 | 100 | 100 |
| **Little Italy** | 100 | 100 | 100 | 100 | 100 | 100 |
| **Chinatown** | 100 | 100 | 100 | 100 | 100 | 100 |

Figure 8

By combining the datasets we reduced or top neighborhood list to 20 and reduced the number of venues from more than 10,000 to close to 2,000, with 279 unique venues categories, see figure 9 for the top 5 venues in an example neighborhood.

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| **753** | Belmont | 40.857277 | -73.888452 | Tino's Delicatessen | 40.855882 | -73.887166 | Italian Restaurant |
| **754** | Belmont | 40.857277 | -73.888452 | Casa Della Mozzarella | 40.855440 | -73.887373 | Deli / Bodega |
| **755** | Belmont | 40.857277 | -73.888452 | Full Moon Pizzeria | 40.855506 | -73.887557 | Pizza Place |
| **756** | Belmont | 40.857277 | -73.888452 | DeLillo Pastry Shop | 40.855364 | -73.887198 | Dessert Shop |
| **757** | Belmont | 40.857277 | -73.888452 | Gino's Pastry Shop | 40.855648 | -73.888196 | Dessert Shop |

See also figure 10 for the updated time series of the now top 20 neighborhoods by the ZHVI (Value Index). As you can see the top neighborhood by venue continue their trend up or have a stable trend.

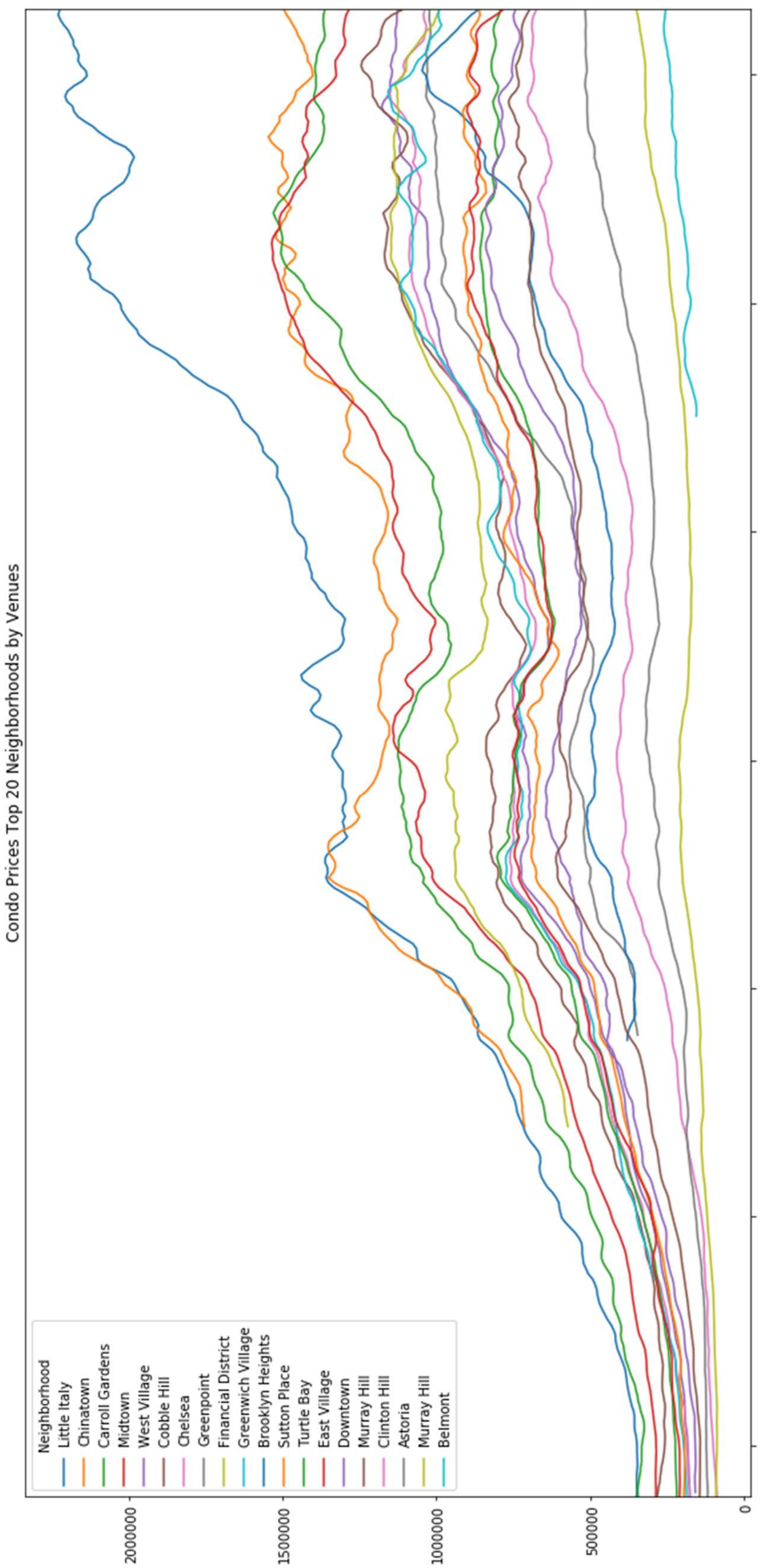
Condo Prices Top 20 Neighborhoods by Venues
USD Condo Prices
Year-Month
0
500000
1000000
1500000
2000000
1996
2000
2004
2008
2012
2016
2020
Neighborhood
Little Italy
Chinatown
Carroll Gardens
Midtown
West Village
Cobble Hill
Chelsea
Greenpoint
Financial District
Greenwich Village
Brooklyn Heights
Sutton Place
Turtle Bay
East Village
Downtown
Murray Hill
Clinton Hill
Astoria
Murray Hill
Belmont

Figure 10

We used the one hot encoding method to analyze each neighborhood. By expanding the dataframe, adding a dummy column for each venue and grouping rows by neighborhood and averaging the occurrences in each category we created a new table with 279 unique venue categories and 19 unique neighborhoods. See Figure 11 for part of the resulting table.

| | Neighborhood | Accessories Store | Adult Boutique | American Restaurant | Antique Shop | Arepa Restaurant | Argentinian Restaurant | Art Gallery | Arts & Crafts Store | Asian Restaurant |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Astoria | 0.00 | 0.00 | 0.010000 | 0.00 | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 1 | Belmont | 0.00 | 0.00 | 0.010417 | 0.00 | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 2 | Brooklyn Heights | 0.00 | 0.00 | 0.010000 | 0.00 | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.020000 |
| 3 | Carroll Gardens | 0.00 | 0.00 | 0.010000 | 0.00 | 0.00 | 0.000000 | 0.000000 | 0.010000 | 0.000000 |
| 4 | Chelsea | 0.00 | 0.00 | 0.038095 | 0.00 | 0.00 | 0.000000 | 0.047619 | 0.000000 | 0.009524 |
| 5 | Chinatown | 0.00 | 0.00 | 0.030000 | 0.00 | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.020000 |
| 6 | Clinton Hill | 0.00 | 0.00 | 0.000000 | 0.00 | 0.00 | 0.000000 | 0.010638 | 0.021277 | 0.000000 |
| 7 | Cobble Hill | 0.00 | 0.00 | 0.010309 | 0.00 | 0.00 | 0.010309 | 0.010309 | 0.010309 | 0.000000 |
| 8 | Downtown | 0.00 | 0.00 | 0.020000 | 0.01 | 0.00 | 0.000000 | 0.000000 | 0.020000 | 0.010000 |
| 9 | East Village | 0.00 | 0.00 | 0.010000 | 0.00 | 0.01 | 0.010000 | 0.010000 | 0.010000 | 0.000000 |
| 10 | Financial District | 0.00 | 0.00 | 0.030000 | 0.00 | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 11 | Greenpoint | 0.00 | 0.00 | 0.010000 | 0.00 | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 12 | Greenwich Village | 0.00 | 0.00 | 0.010000 | 0.00 | 0.00 | 0.000000 | 0.010000 | 0.000000 | 0.000000 |
| 13 | Little Italy | 0.00 | 0.00 | 0.000000 | 0.00 | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 14 | Midtown | 0.00 | 0.00 | 0.020000 | 0.00 | 0.00 | 0.000000 | 0.010000 | 0.000000 | 0.000000 |
| 15 | Murray Hill | 0.00 | 0.00 | 0.020408 | 0.00 | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.006803 |
| 16 | Sutton Place | 0.00 | 0.01 | 0.010000 | 0.00 | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.010000 |
| 17 | Turtle Bay | 0.00 | 0.00 | 0.020000 | 0.00 | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.010000 |
| 18 | West Village | 0.01 | 0.00 | 0.050000 | 0.00 | 0.00 | 0.000000 | 0.010000 | 0.000000 | 0.000000 |

Figure 11

[5] NYC Landlords Are Cutting Rents for Apartments Closer to Offices (https://www.bloomberg.com/graphics/property-prices/nyc/)

# Part 4: Modeling

## SELECTION

As we mentioned in Part 3, we decided not to use the k-means algorithm for segmenting and clustering the neighborhoods, as this was done in the course week 3 Lab Segmenting and Clustering Neighborhoods in New York City, with inconclusive results.

By analyzing figure 11 we realize that this information can be used in a recommendation system (like Netflix or Amazon recommend movies or books), but we could not get enough information on the preference of the venues from the foursquare API to do a collaborative filtering of the venues, so the model has to be based on the venue content.

So we decided to ask a potential client(tester) to select their venues preferences and simplify the content-based recommendation algorithm to a minimum viable model.

## DEPLOYMENT

Starting from the resulting venues by neighborhoods table shown in figure 11, we modified the functions and scripts used in the course week 3 Lab Segmenting and Clustering Neighborhoods in New York City, and sorted the venues and created a new dataframe with the top ten most common venues by neighborhood as fully shown in figure 12.

The table in figure 12 show the neighborhoods in alphabetical order, a hard copy was shown to a potential client (tester) and used select the top 10 must have venue categories by circling the selected venues from a reduced list.

The top 10 must have preferences provided by the potential client (tester) are: Gym, Mediterranean Restaurant, Mexican Restaurant, Pub, Restaurant, Sandwich Place, Spa, Steakhouse, Sushi Restaurant & Yoga Studio. The provided preferences are a must have and are not in a priority order, so these preferences were given equal weights.

A new table was produced that sort the neighborhoods by the value index (ZHVI), and a hard copy shown to the potential client (tester) with the average condo value index as partially shown in figure 13, so this table confirmed our assumptions that the top neighborhoods by venues are higher in the value index (pricier).

In a simplified (content-based) recommend system algorithm, we simple multiply the potential client (tester) preferences list times the grouped neighborhoods by venue table (figure 11), so given that the preferences have an equal weight of 1, we are simply applying a filter to the table.

A new total weight from the sum of the venues was calculated and multiplied to the value index (ZHVI), a new Ranking was produced dividing the results by the maximum weight.

| | 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 | Astoria | Bar | Middle Eastern Restaurant | Greek Restaurant | Hookah Bar | Indian Restaurant | Seafood Restaurant | Bakery | Mediterranean Restaurant | Café | Gourmet Shop |
| 1 | Belmont | Italian Restaurant | Pizza Place | Deli / Bodega | Bakery | Grocery Store | Dessert Shop | Donut Shop | Bank | Café | Shoe Store |
| 2 | Brooklyn Heights | Yoga Studio | Park | Deli / Bodega | Italian Restaurant | Cosmetics Shop | Gym | Bakery | Pizza Place | Mexican Restaurant | Ice Cream Shop |
| 3 | Carroll Gardens | Italian Restaurant | Coffee Shop | Bakery | Pizza Place | Cocktail Bar | Spa | Bar | Wine Shop | French Restaurant | Food & Drink Shop |
| 4 | Chelsea | Coffee Shop | Art Gallery | American Restaurant | Italian Restaurant | Bakery | Ice Cream Shop | Hotel | Seafood Restaurant | Sushi Restaurant | Nightclub |
| 5 | Chinatown | Chinese Restaurant | Bakery | Cocktail Bar | Vietnamese Restaurant | Dessert Shop | Spa | American Restaurant | Salon / Barbershop | Noodle House | Optical Shop |
| 6 | Clinton Hill | Italian Restaurant | Pizza Place | Thai Restaurant | Wine Shop | Japanese Restaurant | Mexican Restaurant | Chinese Restaurant | Diner | Indian Restaurant | Restaurant |
| 7 | Cobble Hill | Playground | Bar | Cocktail Bar | Coffee Shop | Pizza Place | Yoga Studio | Deli / Bodega | Italian Restaurant | Middle Eastern Restaurant | Bike Rental / Bike Share |
| 8 | Downtown | Burger Joint | Coffee Shop | Pizza Place | Sandwich Place | Grocery Store | Chinese Restaurant | Bar | French Restaurant | Performing Arts Venue | Cocktail Bar |
| 9 | East Village | Bar | Pizza Place | Mexican Restaurant | Ice Cream Shop | Cocktail Bar | Wine Bar | Vietnamese Restaurant | Coffee Shop | Vegetarian / Vegan Restaurant | Korean Restaurant |
| 10 | Financial District | Coffee Shop | Pizza Place | Cocktail Bar | Bar | Hotel | Gym / Fitness Center | Salad Place | Café | Steakhouse | Italian Restaurant |
| 11 | Greenpoint | Bar | Coffee Shop | Pizza Place | Cocktail Bar | Yoga Studio | Mexican Restaurant | French Restaurant | Deli / Bodega | Flower Shop | Thrift / Vintage Store |
| 12 | Greenwich Village | Italian Restaurant | Sushi Restaurant | Café | Clothing Store | Ice Cream Shop | Gym | Dessert Shop | Chinese Restaurant | Indian Restaurant | Bubble Tea Shop |
| 13 | Little Italy | Bakery | Café | Italian Restaurant | Bubble Tea Shop | Coffee Shop | Cocktail Bar | Salon / Barbershop | Pizza Place | Mediterranean Restaurant | Ice Cream Shop |
| 14 | Midtown | Coffee Shop | Clothing Store | Hotel | Bakery | Theater | Sporting Goods Shop | Sandwich Place | Bookstore | Steakhouse | Gym |
| 15 | Murray Hill | Korean Restaurant | Hotel | Coffee Shop | Japanese Restaurant | Sandwich Place | Pizza Place | Bar | Pub | Gym / Fitness Center | Bakery |
| 16 | Sutton Place | Italian Restaurant | Coffee Shop | Furniture / Home Store | Park | Pizza Place | Gym / Fitness Center | Bar | Beer Bar | Beer Garden | Chinese Restaurant |
| 17 | Turtle Bay | Coffee Shop | Italian Restaurant | Café | Sushi Restaurant | Park | Seafood Restaurant | Japanese Restaurant | French Restaurant | Deli / Bodega | Wine Bar |
| 18 | West Village | Italian Restaurant | American Restaurant | New American Restaurant | Cocktail Bar | Wine Bar | Park | Pizza Place | Jazz Club | Cosmetics Shop | Theater |

Figure 12

| | Neighborhood | condo value index($) | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Little Italy | 2224698.0 | Bakery | Café | Italian Restaurant | Bubble Tea Shop | Coffee Shop | Cocktail Bar | Salon / Barbershop | Pizza Place | Mediterranean Restaurant |
| 1 | Chinatown | 1494949.0 | Chinese Restaurant | Bakery | Cocktail Bar | Vietnamese Restaurant | Dessert Shop | Spa | American Restaurant | Salon / Barbershop | Noodle House |
| 2 | Carroll Gardens | 1362646.0 | Italian Restaurant | Coffee Shop | Bakery | Pizza Place | Cocktail Bar | Spa | Bar | Wine Shop | French Restaurant |
| 3 | Midtown | 1284458.0 | Coffee Shop | Clothing Store | Hotel | Bakery | Theater | Sporting Goods Shop | Sandwich Place | Bookstore | Steakhouse |
| 4 | West Village | 1125683.0 | Italian Restaurant | American Restaurant | New American Restaurant | Cocktail Bar | Wine Bar | Park | Pizza Place | Jazz Club | Cosmetics Shop |
| 5 | Cobble Hill | 1110841.0 | Playground | Bar | Cocktail Bar | Coffee Shop | Pizza Place | Yoga Studio | Deli / Bodega | Italian Restaurant | Middle Eastern Restaurant |
| 6 | Chelsea | 1037893.0 | Coffee Shop | Art Gallery | American Restaurant | Italian Restaurant | Bakery | Ice Cream Shop | Hotel | Seafood Restaurant | Sushi Restaurant |
| 7 | Greenpoint | 1022042.0 | Bar | Coffee Shop | Pizza Place | Cocktail Bar | Yoga Studio | Mexican Restaurant | French Restaurant | Deli / Bodega | Flower Shop |
| 8 | Financial District | 990869.0 | Coffee Shop | Pizza Place | Cocktail Bar | Bar | Hotel | Gym / Fitness Center | Salad Place | Café | Steakhouse |
| 9 | Greenwich Village | 990262.0 | Italian Restaurant | Sushi Restaurant | Café | Clothing Store | Ice Cream Shop | Gym | Dessert Shop | Chinese Restaurant | Indian Restaurant |
| 10 | Brooklyn Heights | 863929.0 | Yoga Studio | Park | Deli / Bodega | Italian Restaurant | Cosmetics Shop | Gym | Bakery | Pizza Place | Mexican Restaurant |
| 11 | Sutton Place | 860577.0 | Italian Restaurant | Coffee Shop | Furniture / Home Store | Park | Pizza Place | Gym / Fitness Center | Bar | Beer Bar | Beer Garden |
| 12 | Turtle Bay | 791213.0 | Coffee Shop | Italian Restaurant | Café | Sushi Restaurant | Park | Seafood Restaurant | Japanese Restaurant | French Restaurant | Deli / Bodega |
| 13 | East Village | 783746.0 | Bar | Pizza Place | Mexican Restaurant | Ice Cream Shop | Cocktail Bar | Wine Bar | Vietnamese Restaurant | Coffee Shop | Vegetarian / Vegan Restaurant |
| 14 | Downtown | 742593.0 | Burger Joint | Coffee Shop | Pizza Place | Sandwich Place | Grocery Store | Chinese Restaurant | Bar | French Restaurant | Performing Arts Venue |
| 15 | Murray Hill | 697072.0 | Korean Restaurant | Hotel | Coffee Shop | Japanese Restaurant | Sandwich Place | Pizza Place | Bar | Pub | Gym / Fitness Center |
| 16 | Murray Hill | 349385.0 | Korean Restaurant | Hotel | Coffee Shop | Japanese Restaurant | Sandwich Place | Pizza Place | Bar | Pub | Gym / Fitness Center |
| 17 | Clinton Hill | 674418.0 | Italian Restaurant | Pizza Place | Thai Restaurant | Wine Shop | Japanese Restaurant | Mexican Restaurant | Chinese Restaurant | Diner | Indian Restaurant |
| 18 | Astoria | 515600.0 | Bar | Middle Eastern Restaurant | Greek Restaurant | Hookah Bar | Indian Restaurant | Seafood Restaurant | Bakery | Mediterranean Restaurant | Café |
| 19 | Belmont | 254028.0 | Italian Restaurant | Pizza Place | Deli / Bodega | Bakery | Grocery Store | Dessert Shop | Donut Shop | Bank | Café |

Figure 13

## RESULTS

A neighborhood recommendation ranking with the top neighborhood with a ranking of 1 was produced from the deployment of the simplified venue-based recommender system algorithm, see figure 14 for the results table.

| | condo value index($) | Total | Ranking |
|---|---|---|---|
| Neighborhood | | | |
| Little Italy | 2224698.0 | 0.09 | 1.00 |
| Midtown | 1284458.0 | 0.15 | 0.96 |
| Chinatown | 1494949.0 | 0.09 | 0.67 |
| Financial District | 990869.0 | 0.13 | 0.64 |
| Greenwich Village | 990262.0 | 0.13 | 0.64 |
| Greenpoint | 1022042.0 | 0.12 | 0.61 |
| Cobble Hill | 1110841.0 | 0.10 | 0.57 |
| Clinton Hill | 674418.0 | 0.16 | 0.54 |
| Brooklyn Heights | 863929.0 | 0.11 | 0.47 |
| Murray Hill | 697072.0 | 0.13 | 0.45 |
| West Village | 1125683.0 | 0.08 | 0.45 |
| Sutton Place | 860577.0 | 0.10 | 0.43 |
| Turtle Bay | 791213.0 | 0.10 | 0.40 |
| Carroll Gardens | 1362646.0 | 0.05 | 0.34 |
| Astoria | 515600.0 | 0.13 | 0.33 |
| Downtown | 742593.0 | 0.09 | 0.33 |
| Chelsea | 1037893.0 | 0.05 | 0.25 |
| East Village | 783746.0 | 0.06 | 0.23 |
| Murray Hill | 349385.0 | 0.13 | 0.23 |
| Belmont | 254028.0 | 0.06 | 0.08 |

Figure 14

As shown in Figure 14, the results are as expected, with the total column showing the weights of the preferred venues, and the new ranking produced by these preferences.

Figure 15 show a new map with the selected neighborhoods markers.

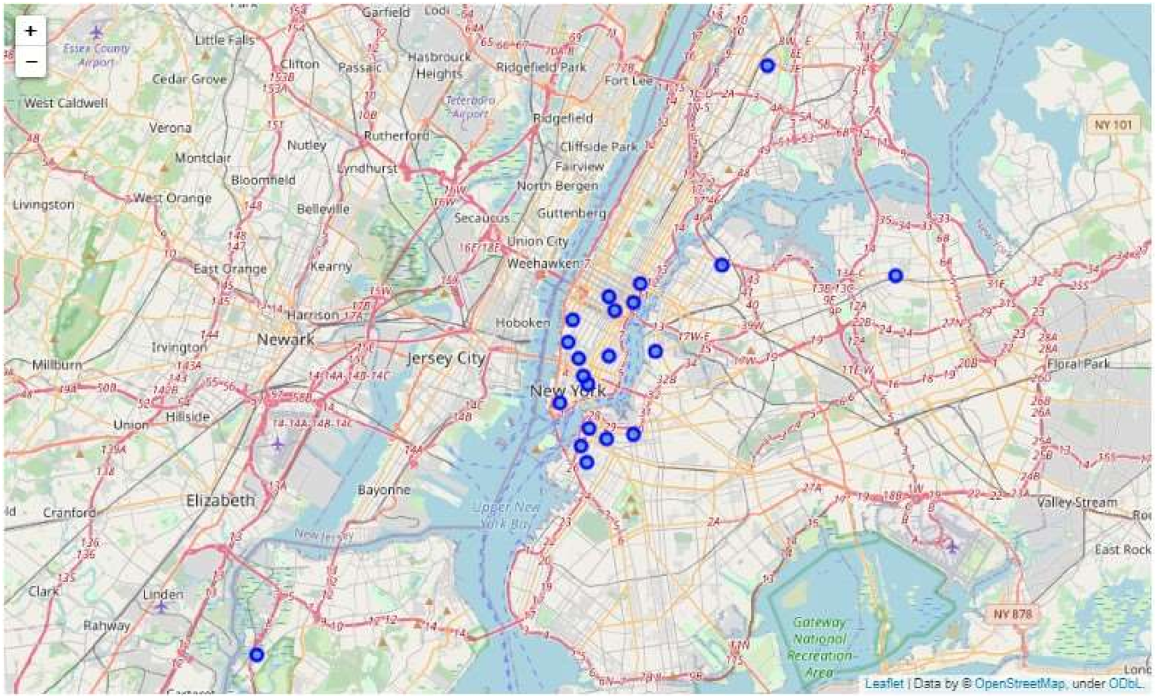


Figure 15

## Part 5: Conclusions

The results of the model differ from the previous rankings calculated by only one variable, either the value index of the maximum number of venues. By adding the potential client (tester) preferences to the calculation we are essentially personalizing the ranking, as this is the intention of a recommender system.

The value of the results is in the differentiation of the rankings, and when a maximum investment cap and their personal venue preferences is provided, we can give to our potential clients the top neighborhoods to consider for an investment. This answer the business question from Part 1: How to select the best neighborhood for a given investment amount and the preferences venues of a potential investor.

## Part 6: Future Work

The rankings are subject to change by the fluctuation of the condo value index, for this report we decided not to use the Zillow forecasts, or the discounts provided by StreetEasy as shown in figure 7. This are theoretical discount and will not be realized until a purchase is made, so we recommend apply them with caution at updating the model.

This model can be applied to other cities, suburbs, or lower market segments, and in these cases the variables that influence the model will differ, we can use crime, or economic data to generate a city-segment specific model.

A full fledge recommender system can be generated by a collaborative filtering of the preferred venues with data acquired from Foursquare or Google.