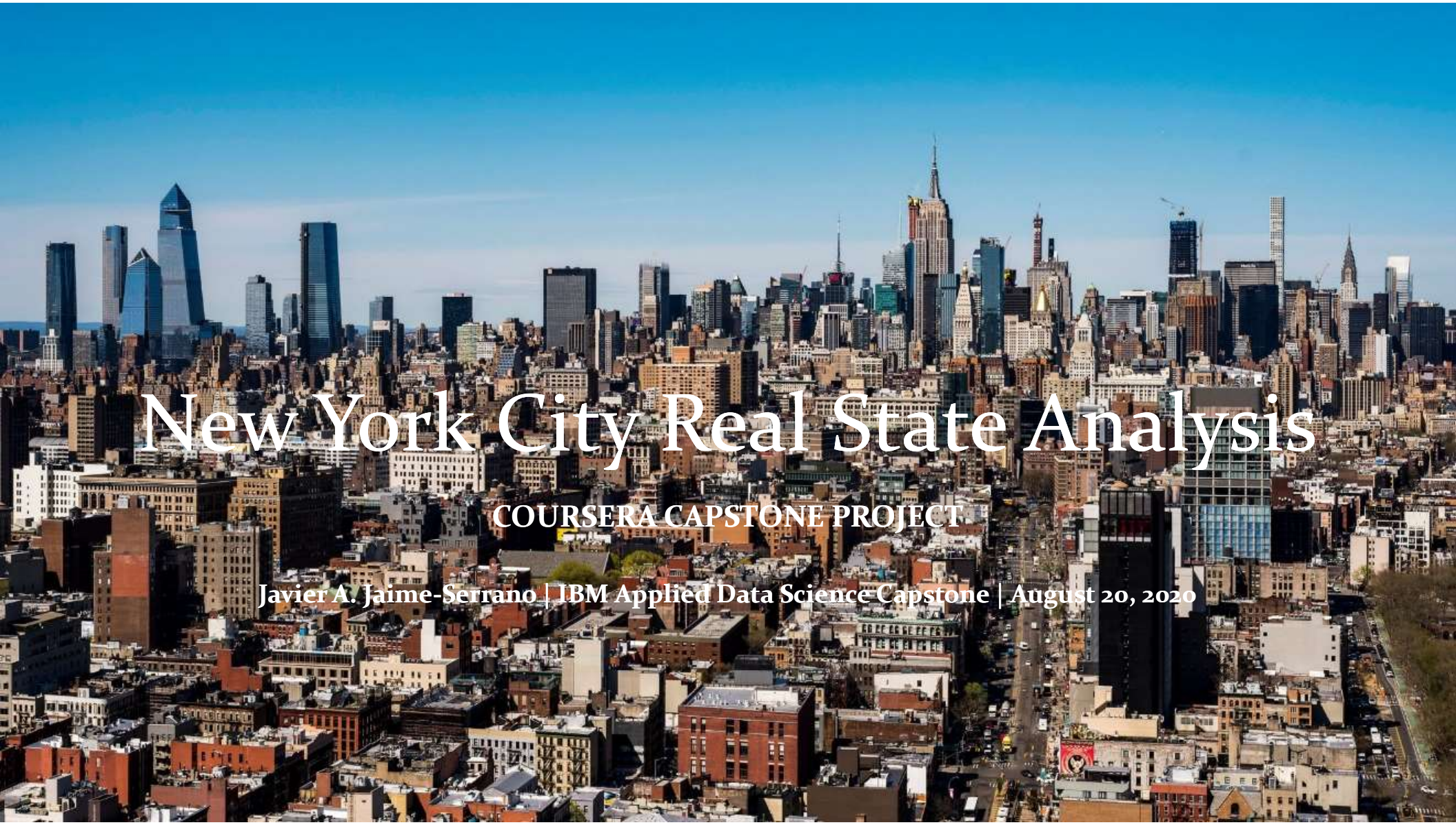


New York City Real State Analysis

COURSERA CAPSTONE PROJECT

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

Part 1: Introduction

Background

- New York City is the most populated city in the US (over 8 million)
- Largest Metropolitan Area (over 20 million)
- Five boroughs: Brooklyn, Queens, Manhattan, the Bronx, and Staten Island
- Total Value of Real State over $1 Trillion, 10% Year to Year increases

Business Problem

- How to select the best neighborhood for a given investment amount and the preferences venues of a potential investor.

Part 2: Data Acquisition

Data Sources

- Dataset 1 from NYU Spatial Data Repository contains 5 boroughs, 306 neighborhoods as the location coordinates of each neighborhood.

- Dataset 2 from Zillow Research Data contains a time series of the Zillow Home Value Index (ZHVI) for mid-tier condo/coops.

- Dataset 3 from Foursquare API contains the venues, categories and location data of the neighborhoods requested.

Part 2: Data Acquisition

Cleaning Dataset 1 from NYU

	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

Part 2: Data Acquisition

Cleaning Dataset 2 from Zillow

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

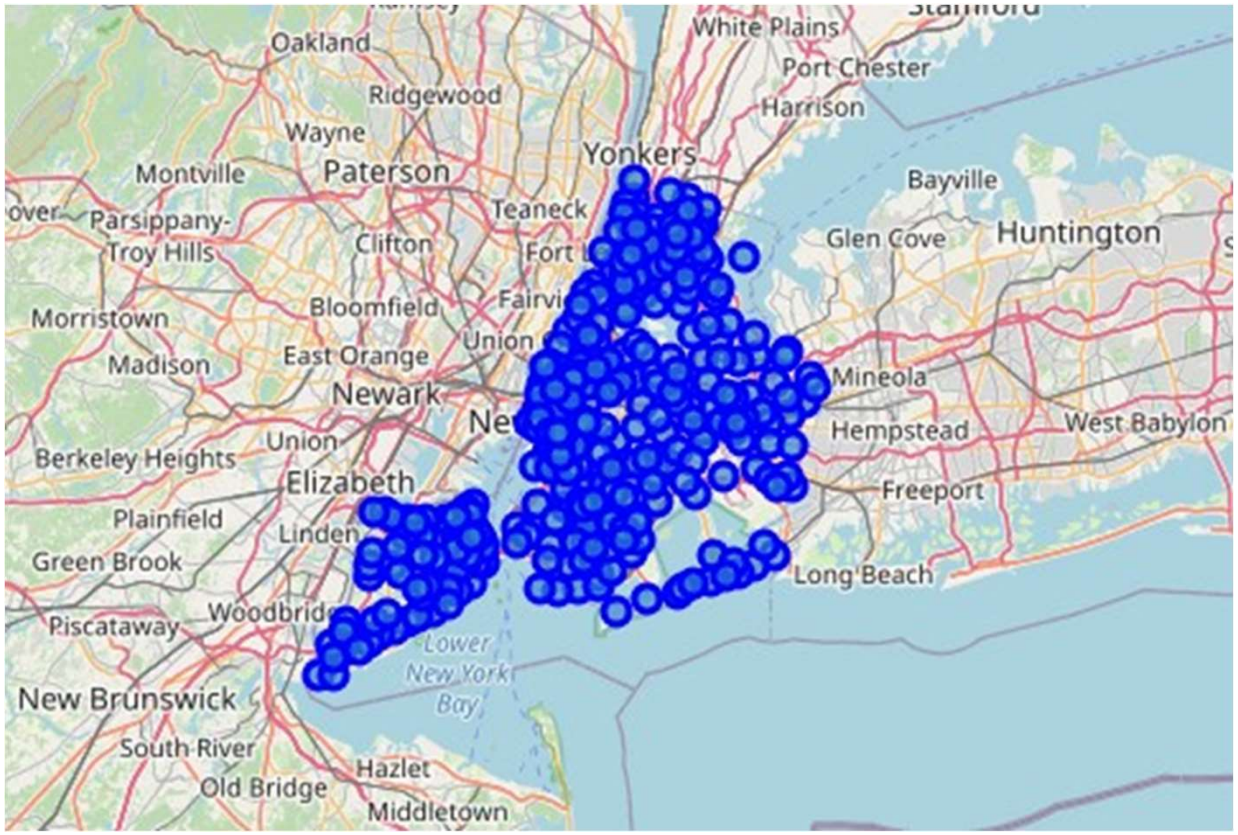
Part 2: Data Acquisition

Cleaning Dataset 2 from Zillow

	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

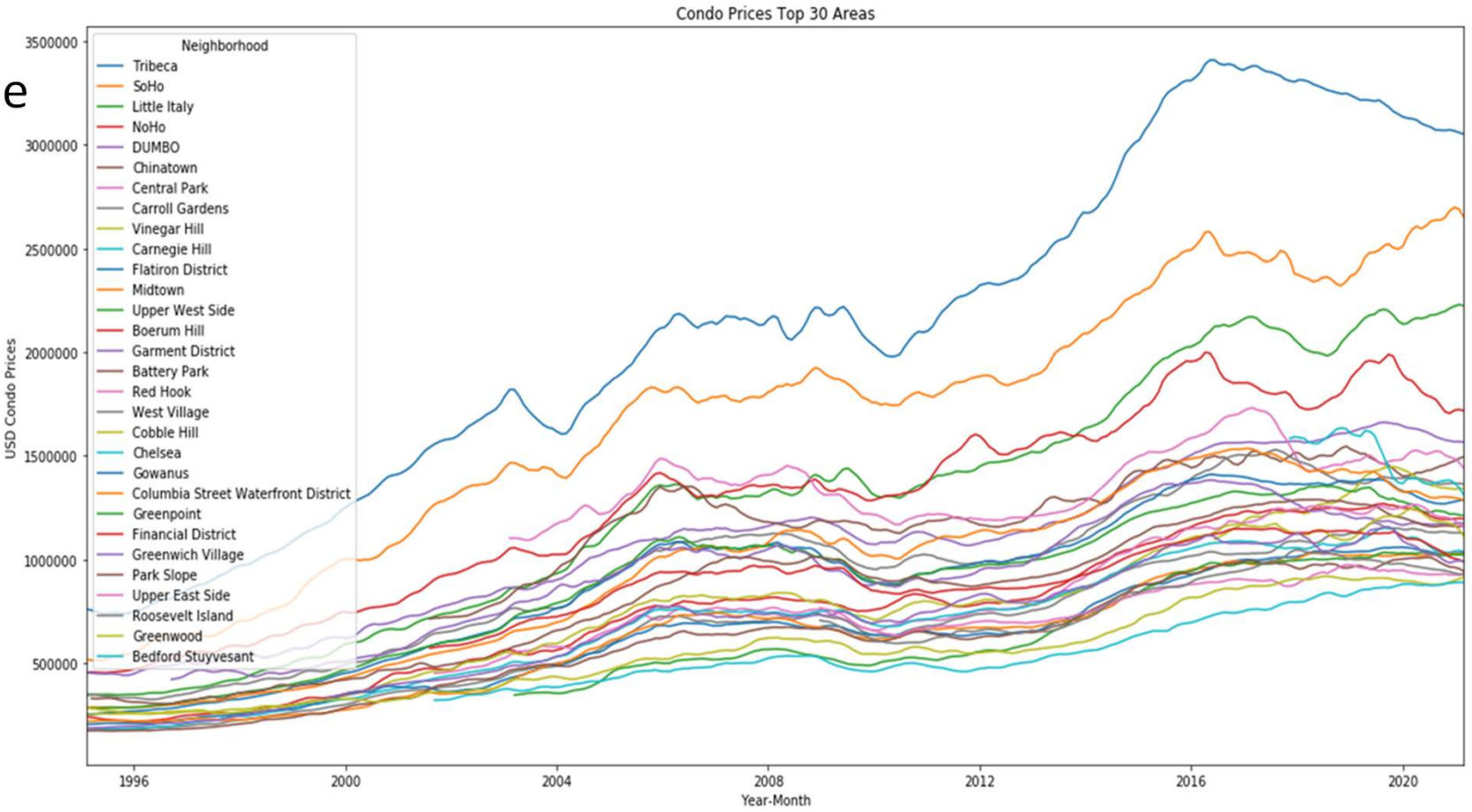
Part 3: Exploratory Data Analysis

Neighborhoods



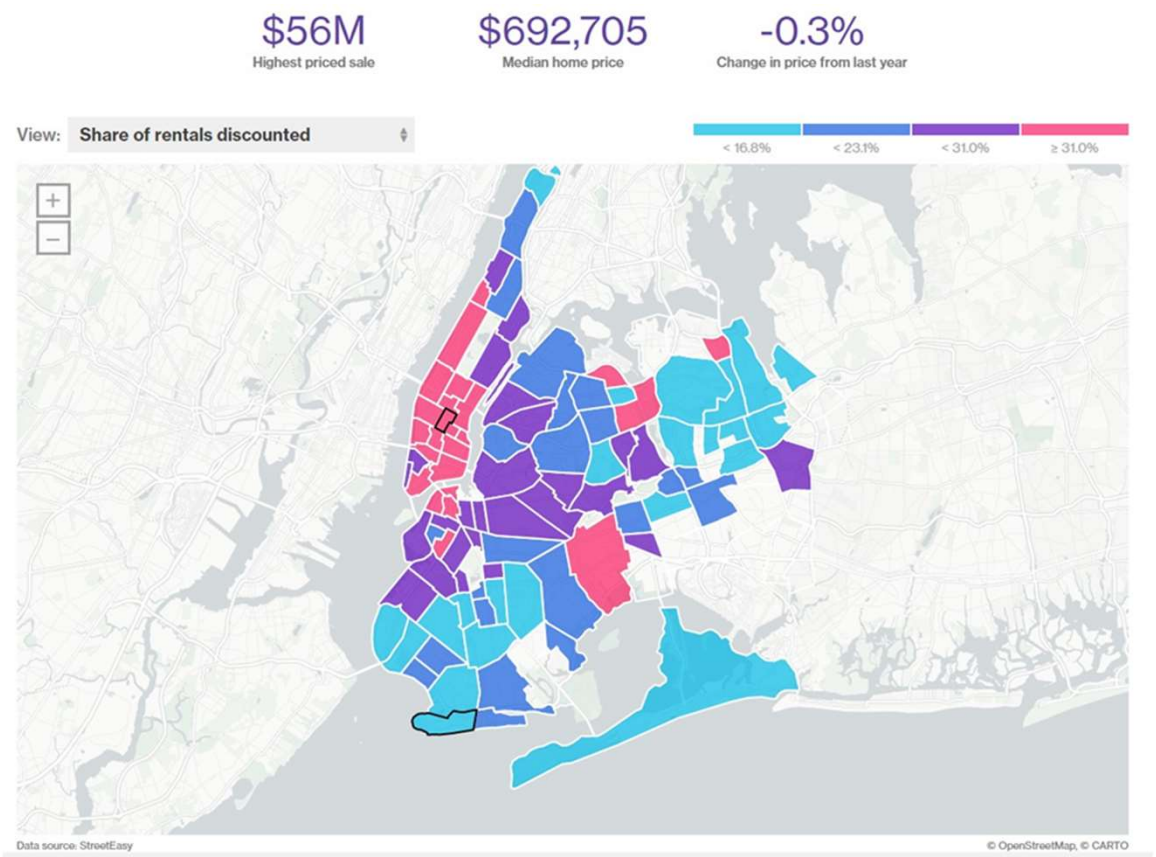
Part 3: Exploratory Data Analysis

Value



Part 3: Exploratory Data Analysis

Status



Part 3: Exploratory Data Analysis

Venues

	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

	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

Part 3: Exploratory Data Analysis

One hot Encoding venues

	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

Part 4: Modeling

Selection

- Not using k-means for segmenting and clustering
- Not enough data for collaborative filtering in a recommender system
- Ask potential client to select their top venue preferences
- Simplify a content-based recommendation algorithm to a minimum

Deployment

- Create new table with the top ten most common venues
- Ask potential client to circle their preference of top ten venues
- Filter the table with the new preferences
- Calculate new ranking from the sum of the venue's weights and the value index

Part 4: Modeling

	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

Part 4: Modeling

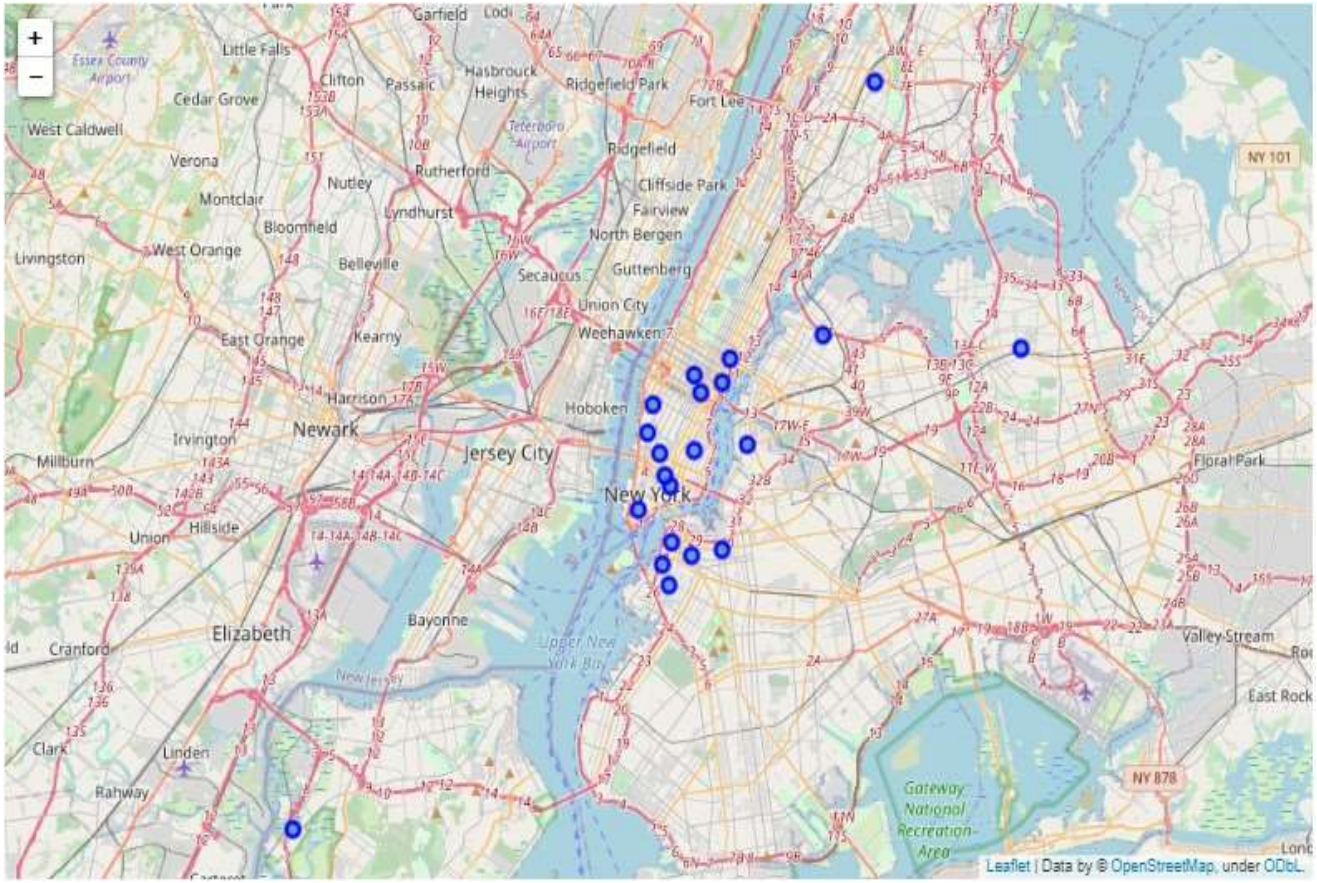
Neighborhood	condo value index($)	Total	Ranking
Little Italy	2224698.0	0.09	1.00
Midtown	1284458.0	0.15	0.96
Chinatown	1494949.0	0.09	0.67
Financial District	990869.0	0.13	0.64
Greenwich Village	990262.0	0.13	0.64
Greenpoint	1022042.0	0.12	0.61
Cobble Hill	1110841.0	0.10	0.57
Clinton Hill	674418.0	0.16	0.54
Brooklyn Heights	863929.0	0.11	0.47
Murray Hill	697072.0	0.13	0.45

Results

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

Part 4: Modeling

Results



Part 5: Conclusions

- Result differ from preliminary ranking by single variables
- By adding the potential client preferences we personalize the ranking
- The Value of the results is in the differentiation of the ranking
- We answer the business question: How to select the best neighborhood for a given investment amount and the preferences venues of a potential investor

Part 6: Future Work

- Ranking subject to change will need to be updated with fresh forecasts or apply recent discounts

- The model can be applied in other cities or segments, new variables will need to be added to improve the model

- A collaborative filtering recommender system can be generated by preferred venue data acquired from Foursquare or Google