

# Evaluation of the Feasibility of Opening a Multi-Sport-Discipline Center in Manhattan (New York City)

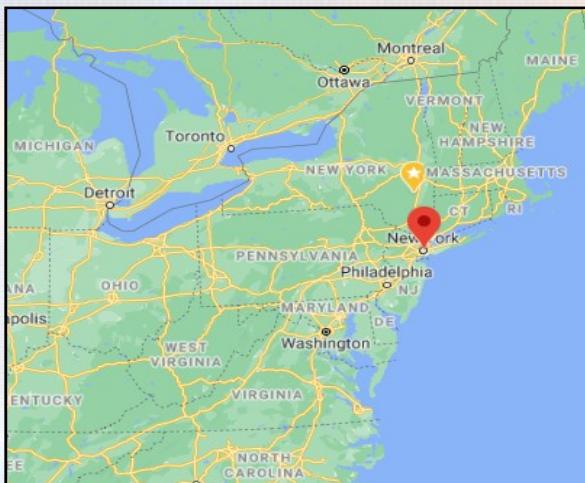
Enrico Tomassoli  
September 5th, 2020



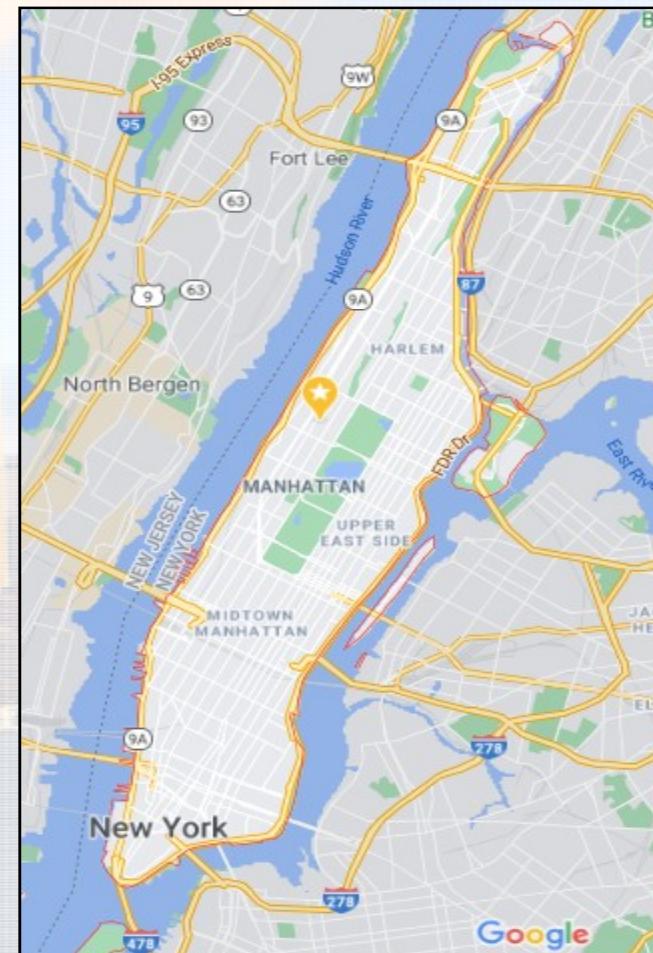
# Introduction and Business Problem

Evaluation of the feasibility of opening a multi-sport center in Manhattan, NY

Geographic Location



Manhattan, NY

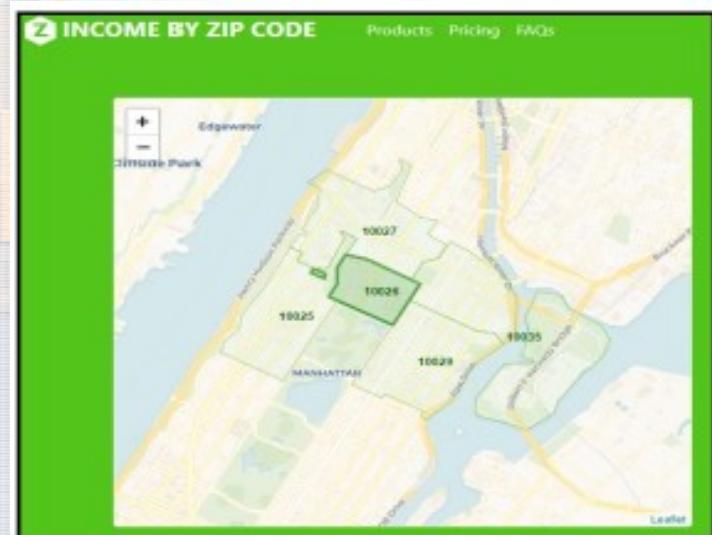


Zip Codes

Manhattan	Central Harlem	10026, 10027, 10030, 10037, 10039
	Chelsea and Clinton	10001, 10011, 10018, 10019, 10020, 10036
	East Harlem	10029, 10035
	Gramercy Park and Murray Hill	10010, 10015, 10017, 10022
	Greenwich Village and Soho	10012, 10013, 10014
	Lower Manhattan	10004, 10005, 10006, 10007, 10038, 10280
	Lower East Side	10002, 10003, 10009
	Upper East Side	10021, 10028, 10044, 10065, 10075, 10128
	Upper West Side	10023, 10024, 10025
	Inwood and Washington Heights	10031, 10032, 10033, 10034, 10040

# Raw Data

- A. Zip Codes and areas
- A. Foursquare and Sport Business Ratio
  - 1. Venue Names
  - 2. Venue Categories
- B. Json file for zip code boundaries
- C. Coordinate:
  - 1. Latitude
  - 2. Longitude
- D. Income:
  - 1. Median Household Income
  - 2. Average Household Income
  - 3. Per-Capita Income
- E. Area land by zip code
- F. Current Population
- G. School Test Performance
- H. Real Estate Price:
  - 1. Average Real Estate Asking Price
  - 2. Average Real Estate Sale Price



# Preliminary Investigation of Raw Data

Correlation among raw data

Definition of the Correlation Matrix

Cases with correlation factor greater than 0.5 or lower than -0.5 considered

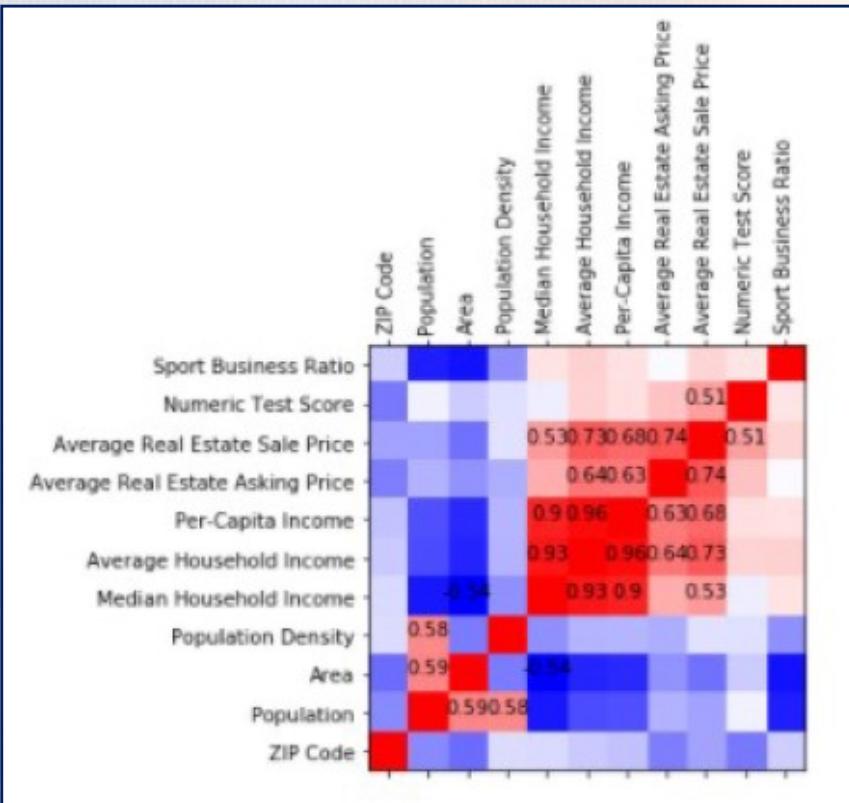
DataFrame with Raw Data (first 10 rows)

ZIP Code	String ZIP Code	Latitude	Longitude	Population	Area	Population Density	Median Household Income	Average Household Income	Per-Capita Income	Average Real Estate Asking Price	Average Real Estate Sale Price	Test Score	Numeric Test Score	Sport Business Ratio
10001	10001	40.750809	-73.986620	21102	0.614	34368.0	88526.0	151628.0	84765.0	1603814.0	1822618.0	Above Average	4	16
10002	10002	40.716469	-73.986183	81410	0.878	92722.0	35859.0	68315.0	32694.0	1194750.0	553000.0	Average	3	2
10003	10003	40.731701	-73.988838	56024	0.576	97264.0	112131.0	189885.0	92781.0	1587643.0	1646747.0	Above Average	4	5
10004	10004	40.703640	-74.014214	3089	0.560	5516.0	157645.0	218650.0	122165.0	878278.0	692808.0	Poor	1	4
10005	10005	40.706139	-74.008392	7135	0.071	100493.0	173333.0	208186.0	106702.0	995792.0	730625.0	Poor	1	8
10006	10006	40.709671	-74.013344	3011	0.092	32728.0	176250.0	214543.0	114611.0	1424800.0	932000.0	Poor	1	7
10007	10007	40.714352	-74.008011	6988	0.162	43136.0	246813.0	367343.0	147547.0	2135800.0	2350000.0	Excellent	5	12
10009	10009	40.727242	-73.980293	61347	0.617	99428.0	61548.0	95947.0	51655.0	963857.0	780469.0	Below Average	2	1
10010	10010	40.738918	-73.984161	31834	0.390	81626.0	117923.0	188064.0	93605.0	1477389.0	1870269.0	Average	3	6
10011	10011	40.740631	-74.000450	50984	0.661	77132.0	128613.0	207287.0	121991.0	2011029.0	1648559.0	Excellent	5	6

# Preliminary Investigation of Raw Data

## Correlation among raw data

Correlation Matrix



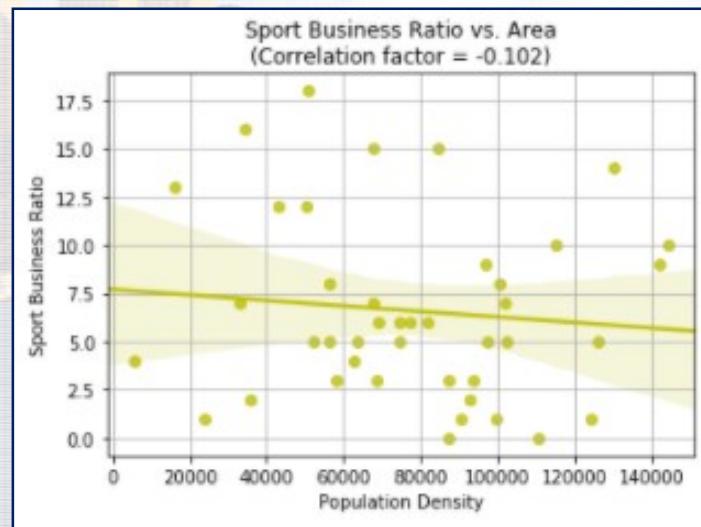
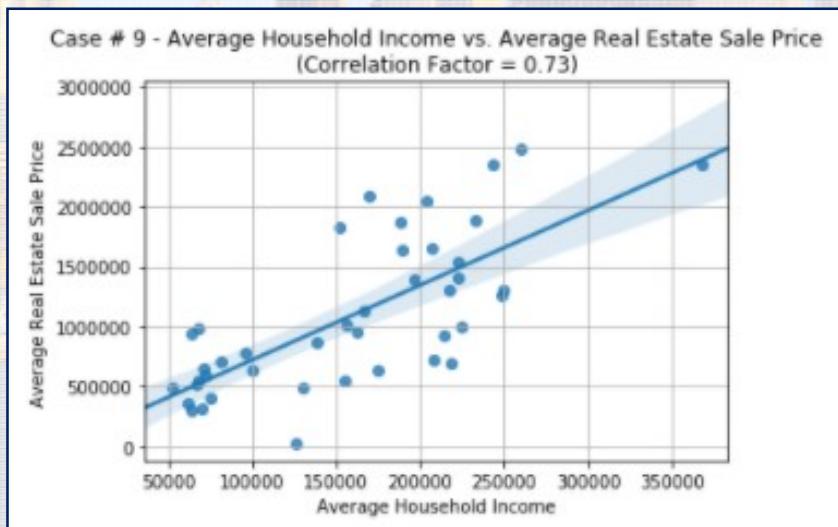
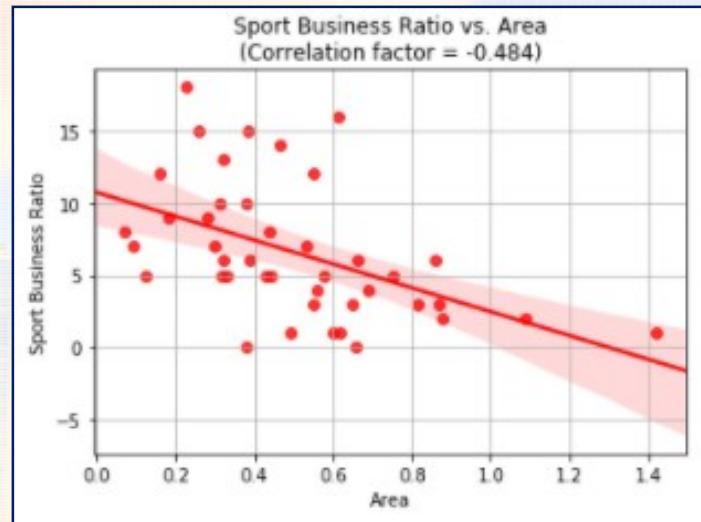
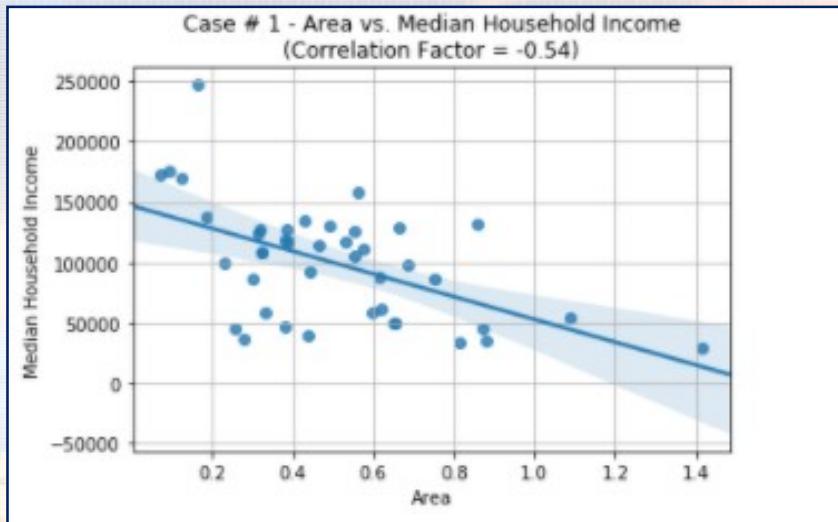
Significant Correlations

Number of unique correlations > 0.5 or < -0.5 = 13		
Case #	Corr	Parameters
1	-0.54	Area vs. Median Household Income
2	0.513	Average Real Estate Sale Price vs. Numeric Test Score
3	0.532	Average Real Estate Sale Price vs. Median Household Income
4	0.584	Population vs. Population Density
5	0.595	Area vs. Population
6	0.629	Average Real Estate Asking Price vs. Per-Capita Income
7	0.641	Average Household Income vs. Average Real Estate Asking Price
8	0.685	Average Real Estate Sale Price vs. Per-Capita Income
9	0.73	Average Household Income vs. Average Real Estate Sale Price
10	0.739	Average Real Estate Asking Price vs. Average Real Estate Sale Price
11	0.897	Median Household Income vs. Per-Capita Income
12	0.929	Average Household Income vs. Median Household Income
13	0.958	Average Household Income vs. Per-Capita Income

- The heavier the “red” the stronger the positive correlation
- The heavier the “blue” the stronger the negative correlation

# Preliminary Investigation of Raw Data

## Correlation among raw data



# Presentation of Data

## Representation of data and finding using Folium maps

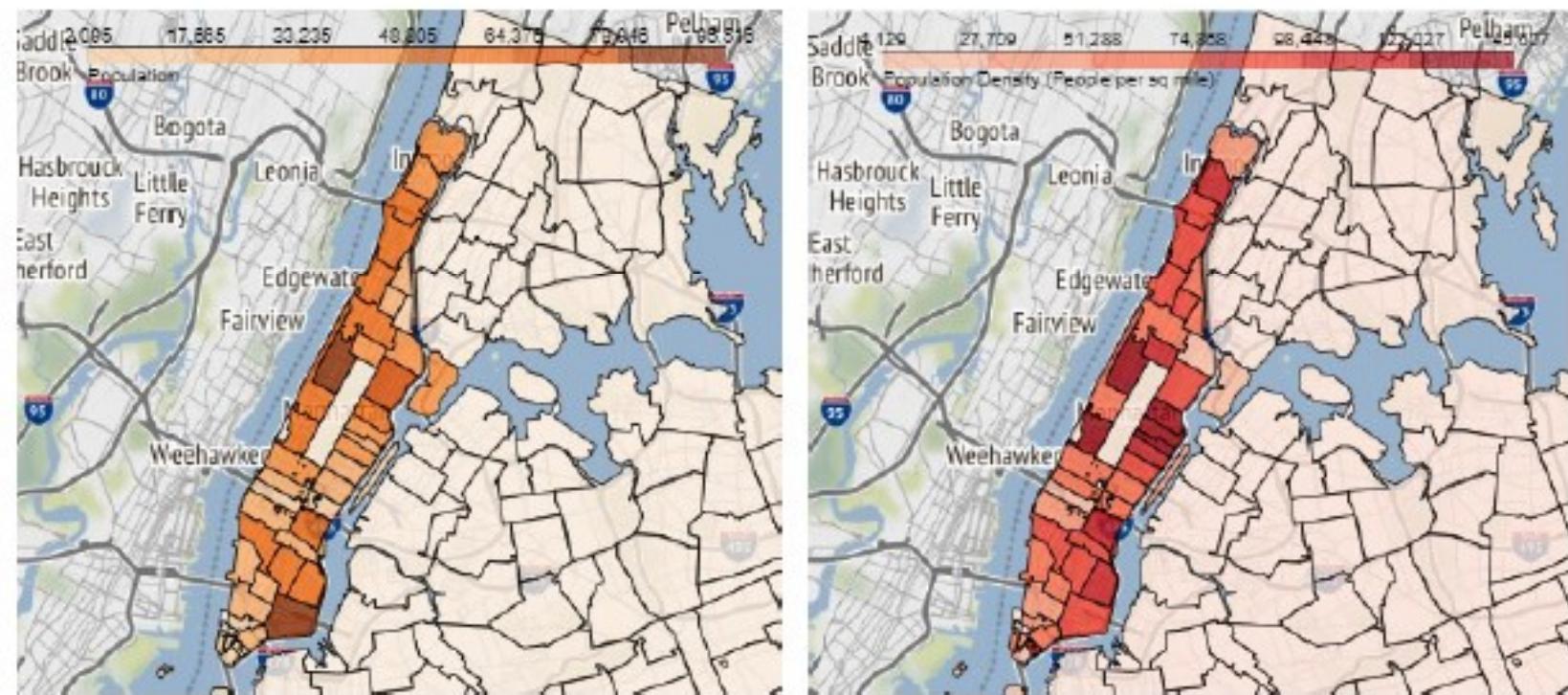
Folium map with markers reporting main Info



# Presentation of Data

## Representation of data and finding using Folium maps

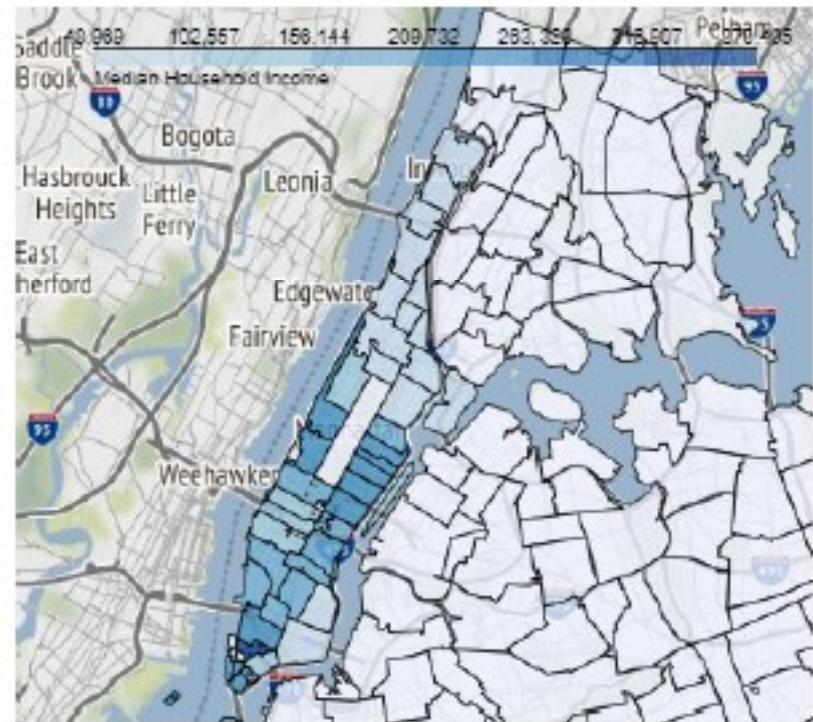
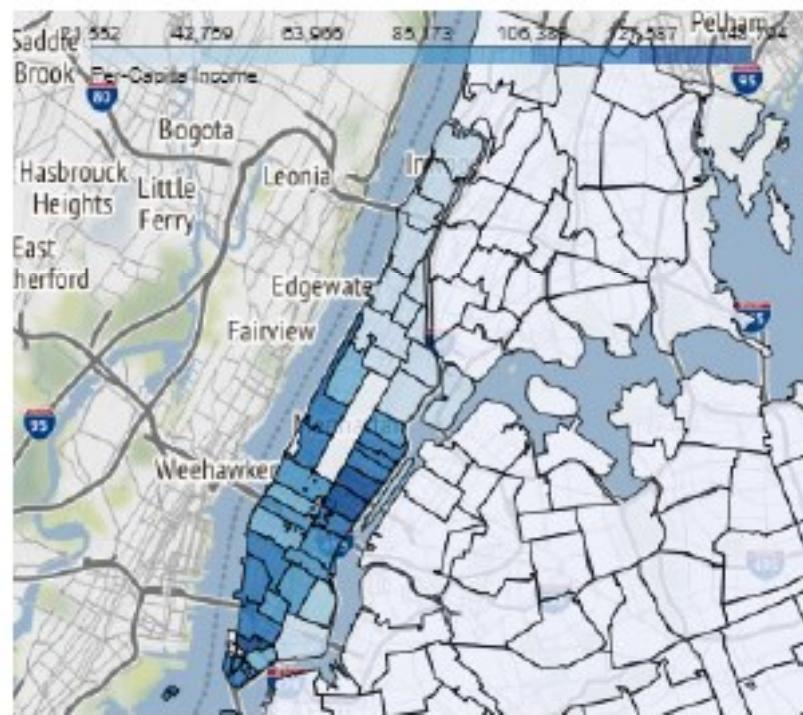
Population (left) and Population Density (right)



# Presentation of Data

## Representation of data and finding using Folium maps

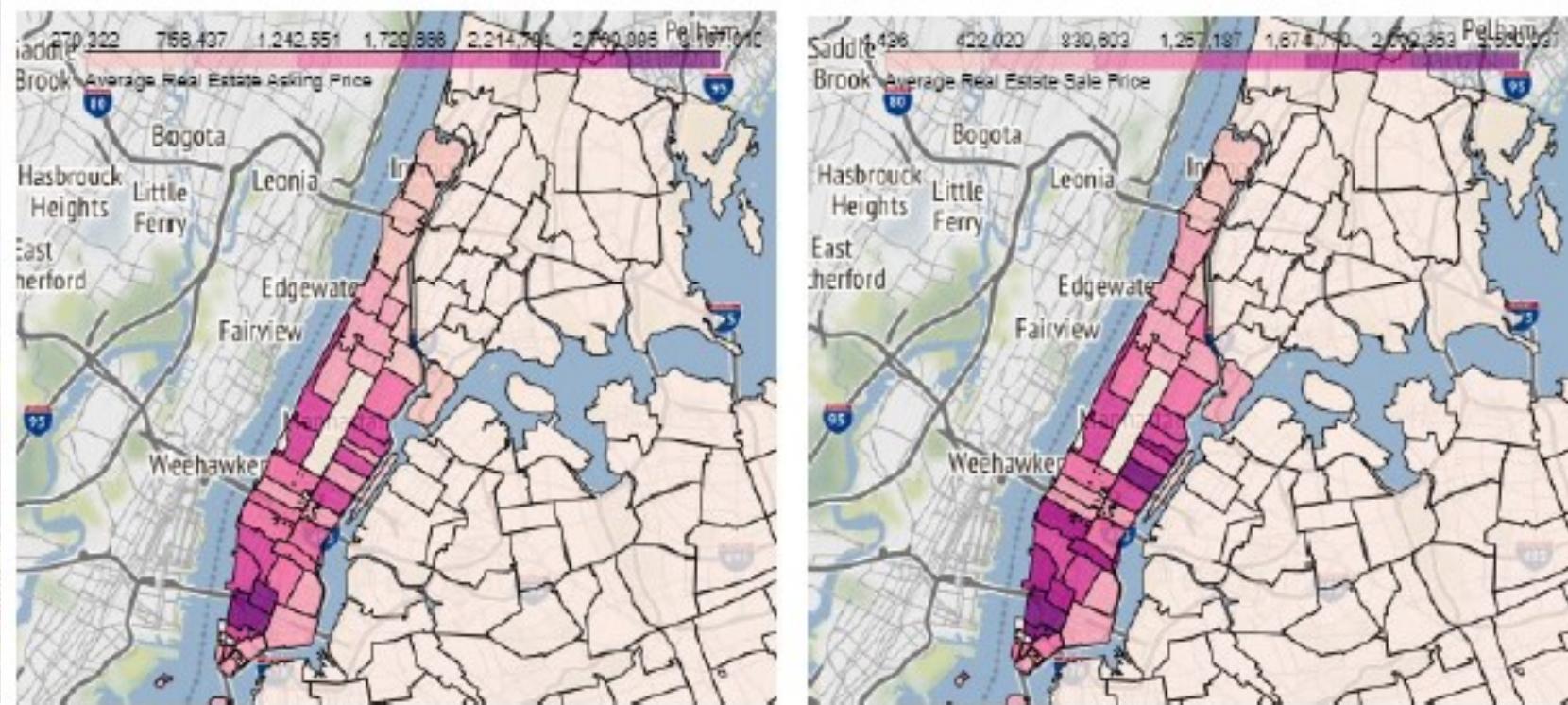
Per-Capita (left) and Median Household Income (right)



# Presentation of Data

## Representation of data and finding using Folium maps

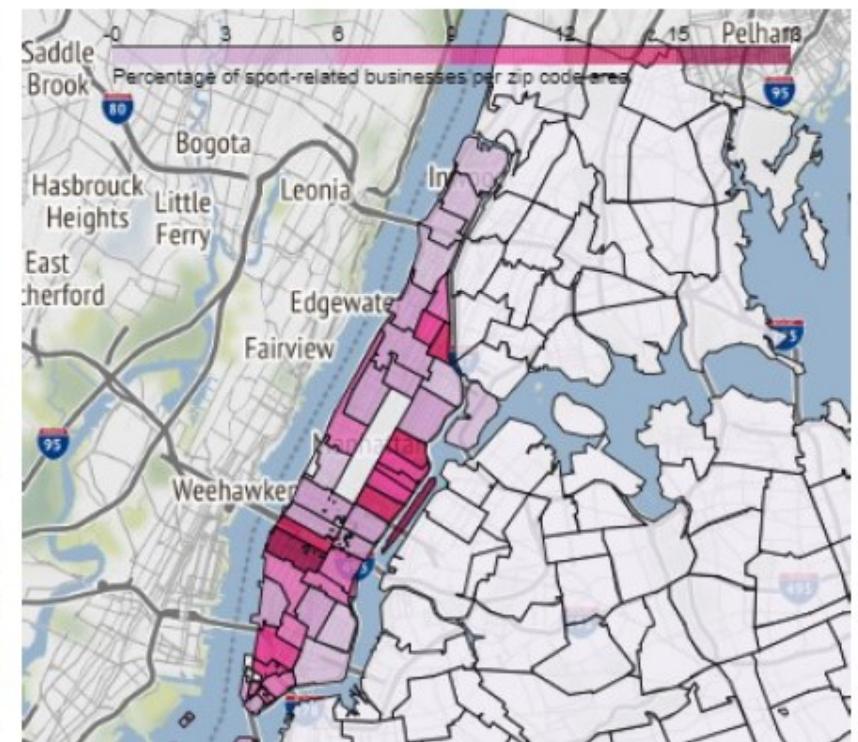
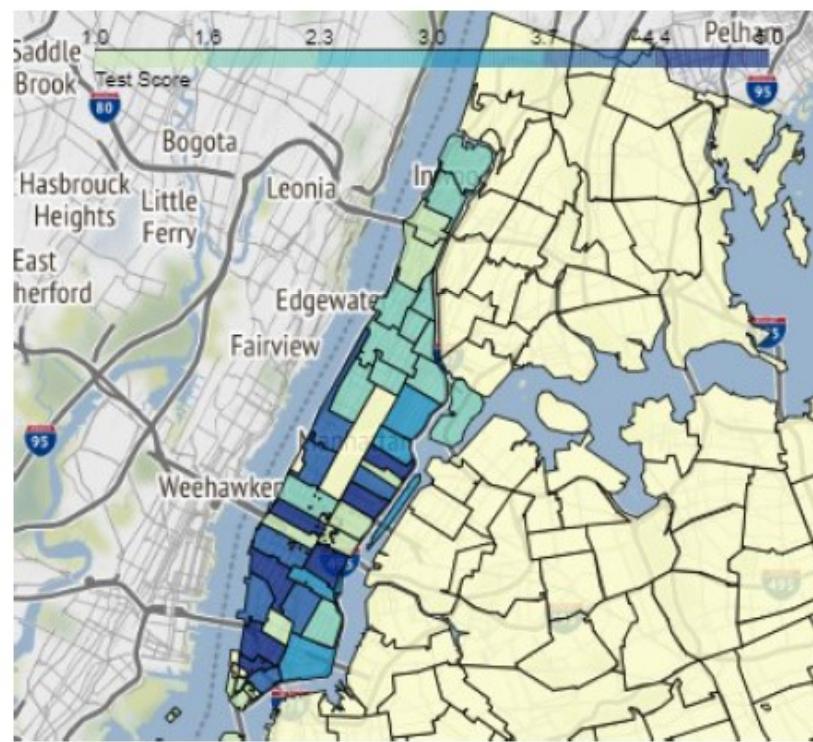
Average Real Estate Asking (left) and Sell (right) Price



# Presentation of Data

## Representation of data and finding using Folium maps

Test Score Performance (left) and Business Sport Ration (right)



# Parameters and Normalization for Clustering

Parameters used for clustering analysis using K-Mean Method

- Population
- Population Density
- Average Household Income
- Median Household Income
- Per-Capita Income
- Average Real Estate Sale Price
- Numeric Test Score and Sport
- Business Ratio

DataFrame with parameters used for clustering (first 10 rows)

	Population	Population Density	Median Household Income	Average Household Income	Per-Capita Income	Average Real Estate Sale Price	Numeric Test Score	Sport Business Ratio
0	21102	34368.0	88526.0	151628.0	84765.0	1822618.0	4	16
1	81410	92722.0	35859.0	68315.0	32694.0	553000.0	3	2
2	56024	97264.0	112131.0	189885.0	92781.0	1646747.0	4	5
3	3089	5516.0	157645.0	218650.0	122165.0	692808.0	1	4
4	7135	100493.0	173333.0	208186.0	106702.0	730625.0	1	8
5	3011	32728.0	176250.0	214543.0	114611.0	932000.0	1	7
6	6988	43136.0	246813.0	367343.0	147547.0	2350000.0	5	12
7	61347	99428.0	61548.0	95947.0	51655.0	780489.0	2	1
8	31834	81626.0	117923.0	188064.0	93605.0	1870269.0	3	6
9	50984	77132.0	128613.0	207287.0	121991.0	1648559.0	5	6

# Parameters and Normalization for Clustering

Normalization of parameters for clustering analysis using K-Mean Method

DataFrame with parameters used for clustering (first 10 rows)

	Population	Population Density	Median Household Income	Average Household Income	Per-Capita Income	Average Real Estate Sale Price	Numeric Test Score	Sport Business Ratio
0	21102	34368.0	88526.0	151628.0	84765.0	1822618.0	4	16
1	81410	92722.0	35859.0	68315.0	32694.0	553000.0	3	2
2	56024	97264.0	112131.0	189885.0	92781.0	1646747.0	4	5
3	3089	5516.0	157645.0	218650.0	122165.0	682808.0	1	4
4	7135	100493.0	173333.0	208186.0	108702.0	730625.0	1	8
5	3011	32728.0	178250.0	214543.0	114811.0	932000.0	1	7
6	6988	43136.0	246813.0	367343.0	147547.0	2350000.0	5	12
7	61347	99428.0	81548.0	95947.0	51655.0	780489.0	2	1
8	31834	81626.0	117923.0	188064.0	93805.0	1870289.0	3	6
9	50884	77132.0	128613.0	207287.0	121991.0	1648559.0	5	6

$$\text{Norm} = \frac{\text{Parameter} - \text{Min}}{\text{Max} - \text{Min}}$$

DataFrame with normalized parameters used for clustering (first 10 rows)

	Population	Population Density	Median Household Income	Average Household Income	Per-Capita Income	Average Real Estate Sale Price	Numeric Test Score	Sport Business Ratio
0	0.198	0.208	0.270	0.316	0.497	0.730	0.75	0.889
1	0.856	0.629	0.027	0.051	0.079	0.213	0.50	0.111
2	0.579	0.661	0.379	0.437	0.561	0.659	0.75	0.278
3	0.001	0.000	0.589	0.528	0.797	0.270	0.00	0.222
4	0.045	0.685	0.661	0.495	0.673	0.286	0.00	0.444
5	0.000	0.196	0.674	0.515	0.736	0.388	0.00	0.389
6	0.043	0.271	1.000	1.000	1.000	0.945	1.00	0.667
7	0.637	0.677	0.145	0.139	0.231	0.306	0.25	0.056
8	0.315	0.549	0.405	0.431	0.568	0.750	0.50	0.333
9	0.524	0.516	0.455	0.492	0.795	0.659	1.00	0.333

# Parameters and Normalization for Clustering

## Normalization of parameters for clustering analysis using K-Mean Method

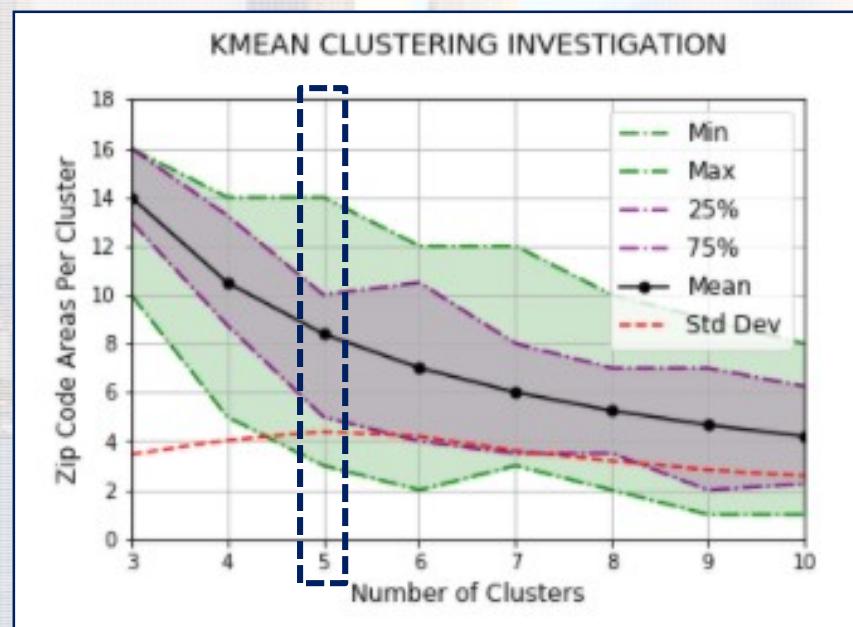
## DataFrame with statistical info for normalized parameters

# K-Mean Clustering

Preliminary investigation to select most appropriate number of clusters

1. Analysis done for number of cluster equal to 3 up to 10
2. For each analysis min, max, 75%, 25%, average and standard deviation related to number of items (zip code areas) per cluster are considered
3. We picked number of clusters equal to 5, where each cluster contains an average of approximately 8 items, with a min of 3 and a max of 14.

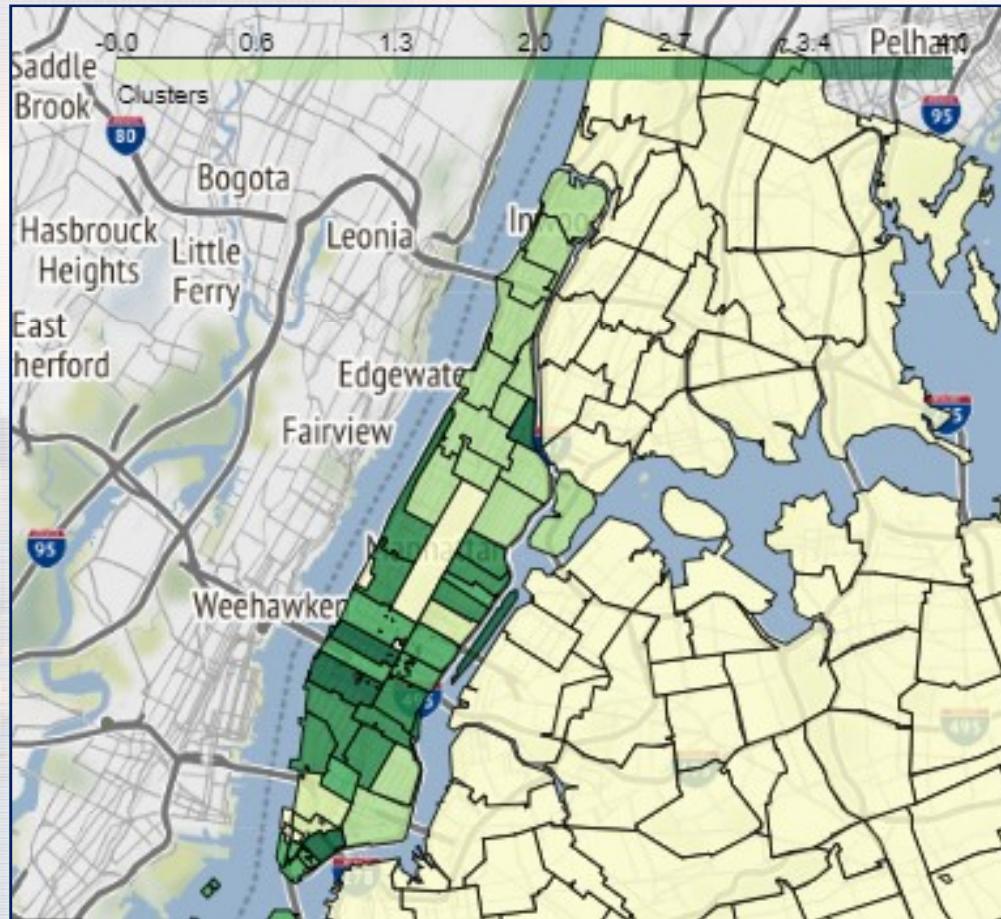
Preliminary Clustering Evaluation



# K-Mean Clustering

Results of clustering shown using Folium map

Clustering Results (with 5 clusters)



# K-Mean Clustering

Results of clustering shown using DataFrame

Clustering Results (with 5 clusters)

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Average
Population	2.231933e+04	52933.785714	16789.7	49449.0	19038.0	3.727638e+04
Population Density	5.932533e+04	83588.857143	62412.9	99865.0	55351.4	7.732752e+04
Median Household Income	1.598767e+05	49290.357143	139387.1	122183.1	82400.4	9.993807e+04
Average Household Income	2.900497e+05	78941.857143	201042.9	206549.1	122253.2	1.586315e+05
Per-Capita Income	1.294887e+05	34106.142857	110643.6	111765.3	62774.2	8.104552e+04
Average Real Estate Sale Price	2.394282e+06	593094.928571	1072727.7	1528863.7	832568.8	1.087260e+06
Numeric Test Score	5.000000e+00	2.000000	1.1	4.1	3.6	2.690476e+00
Sport Business Ratio	1.300000e+01	3.071429	6.6	6.8	12.2	6.595238e+00

Clustering Results Normalized to Average

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Population	1.21	1.54	0.48	0.46	0.59
Population Density	1.21	1.10	0.87	0.74	0.69
Median Household Income	1.17	0.51	1.43	0.66	1.42
Average Household Income	1.23	0.51	1.30	0.60	1.61
Per-Capita Income	1.30	0.43	1.40	0.56	1.46
Average Real Estate Sale Price	1.28	0.56	1.04	0.46	2.07
Numeric Test Score	1.55	0.74	0.41	0.74	1.77
Sport Business Ratio	0.99	0.37	0.82	2.00	2.06

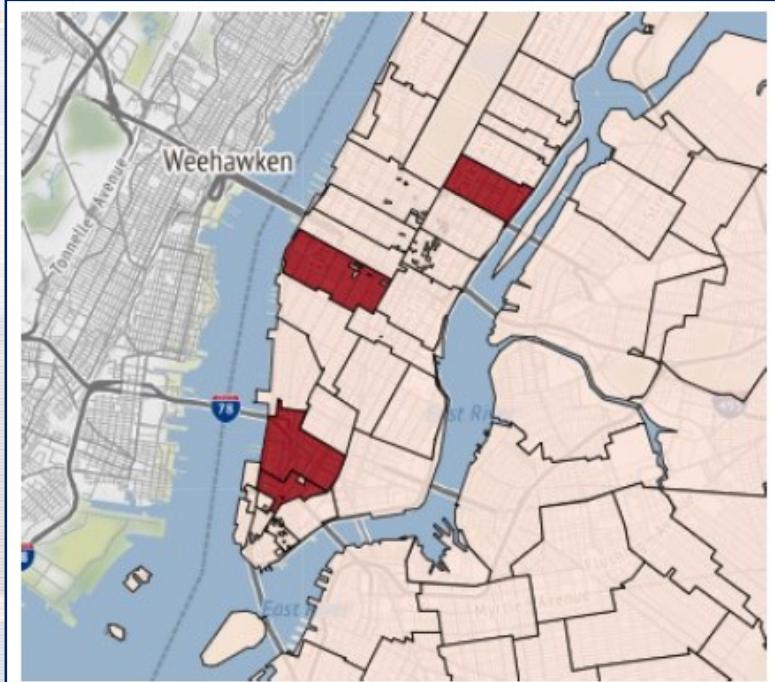
# Cluster Similarity and Recommendation

Existing data for similar neighborhood

Normalized data available for another location where owner already run a successful business

From existing data identify what cluster the existing data belong to

Population = 0.05  
Population density = 0.30,  
Median Household Income = 0.90  
Average Household Income 0.95  
Per-Capita Income = 0.85  
Average Real Estate Sale Price = 0.95  
Numeric Test Score = 0.85  
Sport Business Ration = 0.70



Similar Zip Code Areas  
10001, 10007, 10013, 10065



# Cluster Similarity and Recommendation

## Sport-Related Activities in similar neighborhoods

### ZIP CODE 10001 (Similar Area no. 1 )

	Venue Category	Venue
0	Boxing Gym	Renzo Gracie Academy
1	Gym / Fitness Center	Fly Fitness NYC
2	Gym / Fitness Center	Foxy Fitness and Pole
3	Gym	Crossfit Hell's Kitchen
4	Boxing Gym	iLoveKickboxing
5	Gym	Orange Theory Fitness
6	Dance Studio	You Should Be Dancing...! / Club 412
7	Dance Studio	Piel Canela Dancers
8	Dance Studio	Banana Skirt Productions
9	Dance Studio	Joel Salsa NY
10	Dance Studio	Pearl Studios
11	Dance Studio	Ripley-Grier Studios
12	Yoga Studio	AntiGravity® Aerial Yoga NYC Headquarters
13	Yoga Studio	Sivananda Yoga Vedanta Center New York
14	Martial Arts School	Marcelo Garcia Brazilian Jiu-Jitsu Academy
15	Tennis Court	Midtown Tennis Club
16	Basketball Stadium	Madison Square Garden

### ZIP CODE 10007 (Similar Area no. 2 )

	Venue Category	Venue
0	Gym / Fitness Center	The Class by Taryn Toomey
1	Gym / Fitness Center	Exceed Physical Culture
2	Gym	Equinox Tribeca
3	Gym	CrossFit 212 TribeCa
4	Boxing Gym	Church Street Boxing Gym
5	Gym	The Helena Gym
6	Gym / Fitness Center	Y7 Studio - Tribeca
7	Dance Studio	Gibney Dance Center Downtown
8	Dance Studio	Downtown Dance Factory
9	Cycle Studio	SoulCycle TriBeCa
10	Martial Arts School	Modern Martial Arts NYC Tribeca
11	Martial Arts School	T. Kang Taekwondo

### ZIP CODE 10013 (Similar Area no. 3 )

	Venue Category	Venue
0	Gym / Fitness Center	Barry's Bootcamp
1	Boxing Gym	UFC NYC - SoHo
2	Gym / Fitness Center	Y7 Studio - Tribeca
3	Gym / Fitness Center	AQUA Studio NY
4	Gym	Physique 57
5	Dance Studio	Dance With Me SoHo
6	Yoga Studio	Lyons Den Power Yoga
7	Design Studio	Michele Varian Shop
8	Pilates Studio	New York Pilates - Soho
9	Yoga Studio	SKY TING YOGA
10	Yoga Studio	Y7 Studio - Soho
11	Cycle Studio	Flywheel Sports

### ZIP CODE 10065 (Similar Area no. 4 )

	Venue Category	Venue
0	Gym / Fitness Center	Barry's Bootcamp
1	Gym	The Fhitting Room Penthouse
2	Gym / Fitness Center	SLT E 67
3	Gym	Equinox East 63rd Street
4	Gym	Equinox East 61st Street
5	Gym / Fitness Center	FlyBarre
6	Gym	TS Fitness
7	Gym	ICE NYC
8	Gym / Fitness Center	Crunch Fitness - 59th Street
9	Cycle Studio	Flywheel Sports
10	Cycle Studio	SoulCycle East 63rd
11	Yoga Studio	Earth Yoga NYC
12	Cycle Studio	Flywheel Sports
13	Massage Studio	Dusnee Thai Spa
14	Design Studio	D&D Building

# Summary

- Data Collection and Raw Data Manipulation
- Preliminary Exploration of Raw Data (Correlations)
- Representation of Raw Data Using Folium Maps
- Parameters of Interest used in the Study
- Normalization of Parameters used for Clustering
- K-Mean Clustering for Zip Code Areas in Manhattan, NY
- Representation of Clustering using Folium map
- Info Related to Each Cluster
- Use of Existing Data (other location) to Identify similar Neighborhoods
- Presentation of results for Each Similar Zip Code Area

# Conclusions

- Similar Study Can Be Used in Different Cities Using Same Methodology
- Other type of Business Can Be Considered (Restaurants, Café, Ect.)
- Useful Recommendation to Identify Most Similar Area in Comparison to An Existing Business Condition