



# NEW YORK COMPARATIVE ANALYSIS

Data Science Professional Capstone  
Project

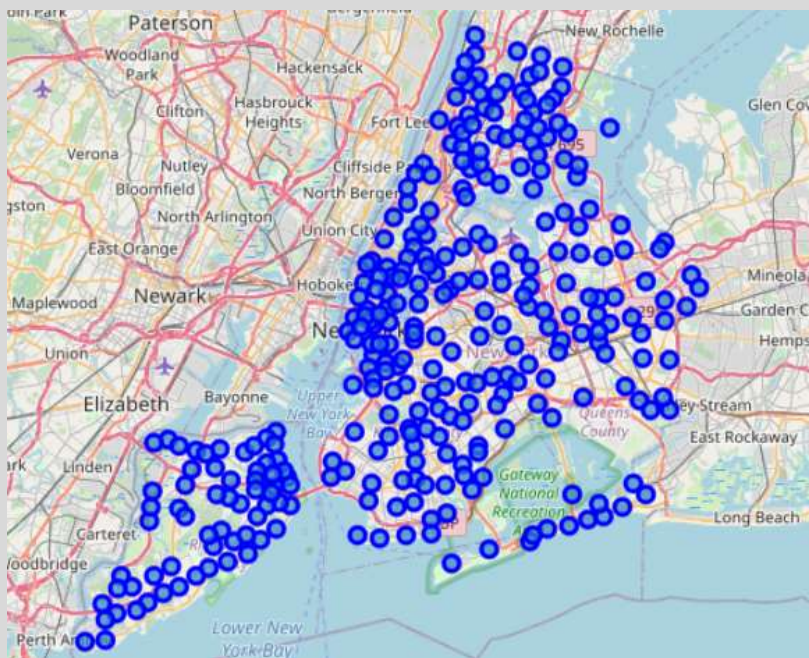
# Comparing Dissimilarity between Cities

- **Background:** A new family living in the Bushwick Neighborhood of Brooklyn, New York seeks to relocate to the tri-city area (Albany, Schenectady, Troy) of Upstate, New York. To ease their transition, this analysis reviews the similarity and differences between the three cities compared to their neighborhood of origin, Bushwick, Brooklyn.
- **Audience:** The analysis provides a point of reference for anyone seeking to compare the neighborhoods in Brooklyn, NY to the tri-city area, and a brief comparison of the similarities and differences between Albany, Schenectady, and Troy, NY.
- **Objective:** The goal of this analysis is to investigate Troy, Schenectady, and Albany, New York and determine which city provides the closest approximation to the living conditions of Bushwick, Brooklyn, NY USA.

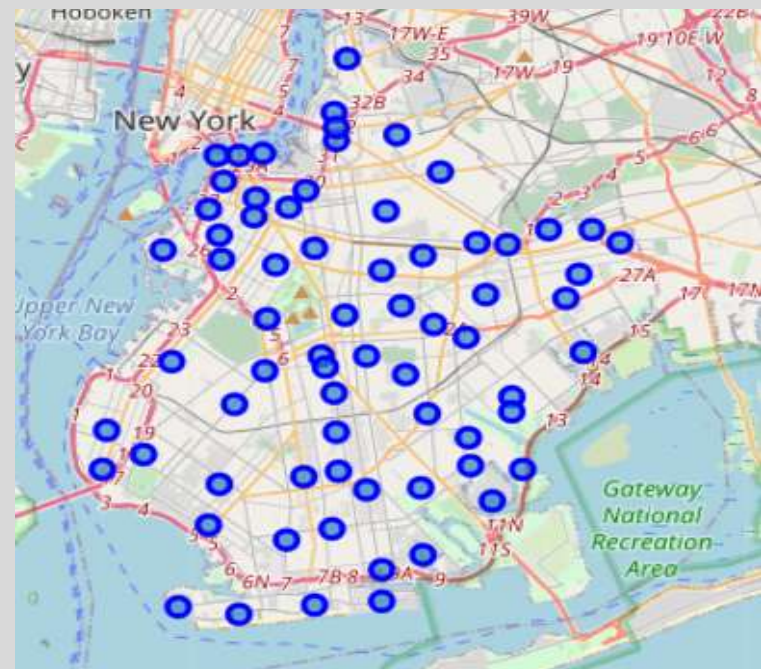
# Data Acquisition and Cleaning

- **City Geolocation Data:** Geolocation data for New York Cities' 5 boroughs, Albany, Schenectady, and Troy, NY used to identify, map, and search for venue information using the Foursquare Application Programming Interface (API)
- **New York City Neighborhood Data:** A list of all NYC neighborhoods and Boroughs acquired from the following linked location: [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)
- **Shops/Venues:** Foursquare API Data for all identified cities and neighborhoods.
  - Venue Name and Category by Neighborhood in New York City, and by City for all Upstate, NY Venues.
    - All venues were searched within 2000 meters of Upstate, NY cities (due to size of the neighborhood), and within 1Km of Bushwick due to the size of Bushwick
- **Housing Price/Cost Information:** Mean Single Family Housing price/cost information from Zillow data, obtained from the following linked location: <https://www.zillow.com/research/data/>
  - Monthly data from 1996 through 2020. Data cleaned and transformed from horizontal to vertical structure for analysis.
  - Housing data filtered based on Bushwick, Albany, Schenectady, Troy zip-codes, using zip-codes.com to locate all zip-codes associated with the neighborhood/city being analyzed.
  - Federal regional cost of living adjustment data – obtained from open source data on the Federal Housing and Finance Agency (FHFA) available at: <https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index-Datasets.aspx>

# Neighborhood Data



All NYC Neighborhoods

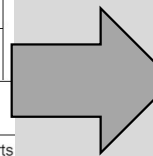


Brooklyn Neighborhoods Only

# Venue Data

- Venue data by Neighborhood was collected and transformed prior to analysis.
- Data transformation was conducted by Venue Category Type per neighborhood.
- Once transformation was complete, venue counts per venue category were developed, and the average per Neighborhood was used to normalize the data per neighborhood so that neighborhoods more densely populated with structures (such as Brooklyn) could be equally compared with the significantly less dense cities of Upstate, NY

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bay Ridge	40.625801	-74.030621	Pilo Arts Day Spa and Salon	40.624748	-74.030591	Spa
1	Schenectady	42.814243	-73.939569	Tara Kitchen	42.814826	-73.940049	Moroccan Restaurant
2	Schenectady	42.814243	-73.939569	Proctor's Theatre	42.811866	-73.942157	Performing Arts Venue



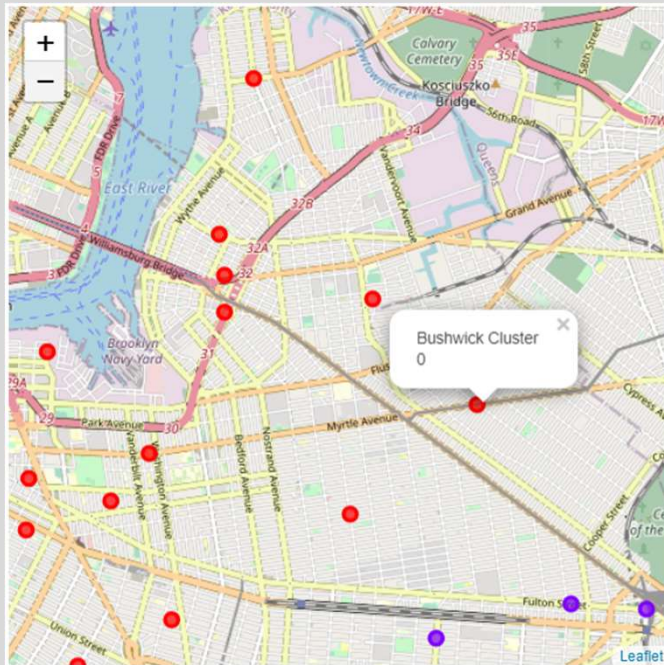
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
68	Windsor Terrace	Deli / Bodega	Diner	Plaza	Café	Park	Grocery Store	Antique Shop	Bagel Shop	Sushi Restaurant
69	Wingate	Fried Chicken Joint	Donut Shop	Hardware Store	Park	Bakery	Gym / Fitness Center	BBQ Joint	Juice Bar	Fast Food Restaurant
70	Albany	Café	Park	American Restaurant	Pub	Sandwich Place	New American Restaurant	Discount Store	Convenience Store	Deli / Bodega
71	Schenectady	Bar	Italian Restaurant	Convenience Store	Pharmacy	Donut Shop	Hotel	Diner	Pizza Place	Pub
72	Troy	Bar	Pizza Place	Convenience Store	Sandwich Place	Park	Bakery	Hot Dog Joint	Coffee Shop	Diner

# Clustering Analysis

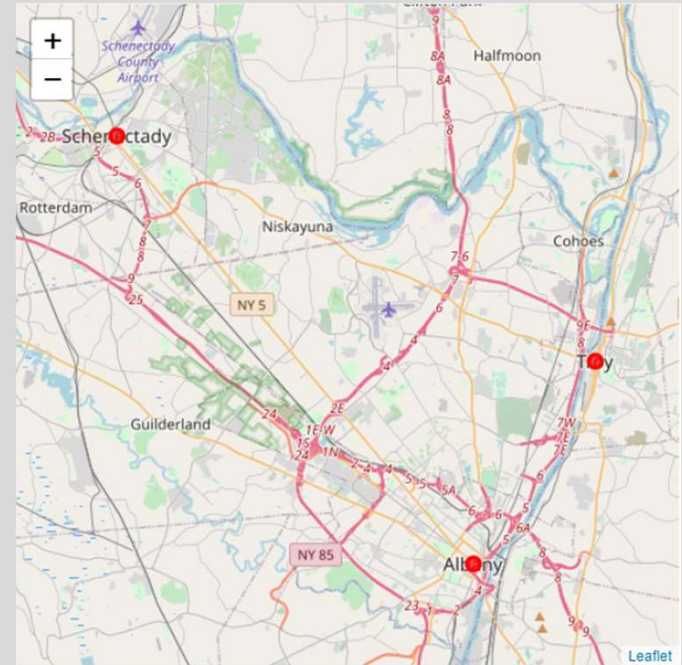
- **Data Analysis:** Venue Data by Neighborhood was clustered using unclassified k-means clustering at the following intervals to check for significant deltas in results:
  - Increasing the number of venues from 5 to 25 in 5 venue increments.
  - Clustering from 3 to 15 in of 3.
- **Clustering Analysis Results:** in all instances, Albany, Schenectady, and Troy were always included in the same cluster as our Neighborhood of Interest, Bushwick, Brooklyn, NY.



# Clustering Results



Brooklyn Clustering Results



Upstate, NY Clustering Results

- Maps cluster results are color coded, as shown Bushwick (left) is in the same cluster, based on venue category and frequency as Albany, Schenectady, and also Troy, NY.

# Housing Price Analysis

- **Home Prices:** Given that all four neighborhoods of interest were deemed similar, further analysis was explored to determine if one Upstate, NY neighborhood may be more similar to Brooklyn than the others based on Average Single Value Home Price.
- **Data:** Data was collected from the Zillow Single Family Homes Database by zipcode, and zipcodes with no values were dropped from the review.
  - Note in the DataFrame, BW = Bushwick, T = Troy, A = Albany, S = Schenectady

	Date	BW1	BW2	T1	T2	A1	A2	A3	A4	S1	S2
288	12/31/2019	1158112	501716	188144	155897	216845	203571	137034	156061	192182	141470
289	1/31/2020	1163999	504299	188315	156683	217246	203696	137209	156309	192923	142352
290	2/29/2020	1169158	507026	188989	158051	216242	202697	136998	155902	192894	142802
291	3/31/2020	1173964	509759	189791	159260	215339	201520	136846	155413	192478	142774
292	4/30/2020	1181757	511904	191002	160588	214635	200478	137469	155378	192474	143187



# Housing Price Analysis

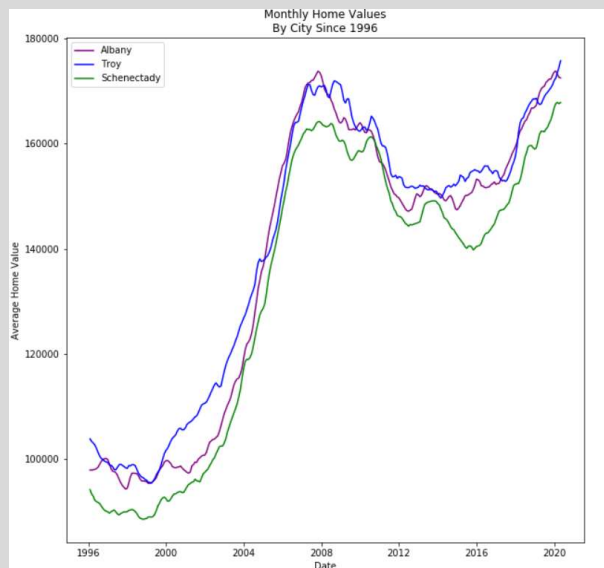
- **Adjusted Home Prices:** Average Home prices for the most recent data (April 2020) were adjusted based on FHFA home price index delta between Brooklyn, NY and the Albany Metro Region, which is identified as a single area. Average home prices in Brooklyn were identified as ~315% > Albany Metro Region Homes

Average Home Prices				
Date	Bushwick	Troy	Schenectady	Albany
2020-04-30	\$846,830	\$553,754	\$528,666	\$543,415

- As shown above, home values in Upstate, NY are ~35% lower than the Bushwick neighborhood of Brooklyn, NY, even after accounting for cost of living differences between the two cities.

# Housing Prices in Upstate, NY

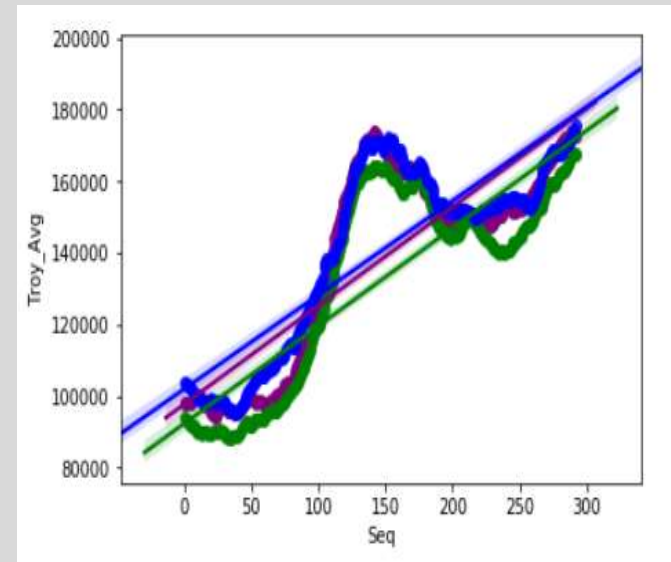
- As previously shown, housing prices in Upstate, NY are very similar. As such, an endeavor was undertaken to determine if either of the three cities are either:
  - Appear to be rising at a different rate than the other two Upstate, NY cities
  - Are the home values statistically different from one another



The chart on left compares home prices between all Upstate, NY cities since 1996.

The chart at right examines a linear regression of home prices to determine similarity of slope.

- Linear Model Scores = .69 for all three cities



# Housing Prices in Upstate, NY

- **Mann-Whitney Testing:** In order to determine if there is any statistically significant difference between the three cities in Upstate, NY, a Mann-Whitney U test was conducted to determine if there was any statistically significant difference between Schenectady (the lowest cost city to purchase a home) vs Albany, and Troy, NY – which are very similarly priced. This test is used due to the fact that the underlying data is both limited and of an unknown distribution.
  - Null Hypothesis 1: The home prices of Albany, NY are the same as Schenectady, NY
  - Null Hypothesis 2: The home prices of Troy, NY are the same as Schenectady, NY
- **Results:** As shown below, in both instances, there was a failure to reject the null, thus there is no statistically significant difference between any of home values of Albany, Schenectady, or Troy, NY.

```
Statistics=3.00000000, p=0.40848066  
Same distribution (fail to reject H0)  
Given failure to reject the Null Hypothesis, it is assumed the values of homes in Albany vs. Schenectady Homes are the same.  
Statistics=2.00000000, p=0.34926768  
Same distribution (fail to reject H0)  
Given failure to reject the Null Hypothesis, it is assumed the values of homes in Troy vs. Schenectady Homes are the same.
```

# Conclusions and Next Steps

- A clustering analysis revealed that the Bushwick Neighborhood in Brooklyn, NY is comparably similar to all three Upstate, NY cities of interest, Albany, Schenectady, and Troy, NY. Thus, a further investigation to separate the three Upstate, NY cities was conducted in an attempt to identify one city that may be significantly different from the other two.
- Results of housing price analyses produced no statistically significant difference in home values between Albany, Schenectady, and Troy, NY; indicating that all three may be considered equivalent based on housing value and difference in home value from each other and from Bushwick, Brooklyn.
- **Next Steps:** The next logical step to separate the three Upstate, NY cities is to delve into further information that may be used distinguish their characteristics from each other. Recommended next steps are:
  - A review of violent crimes by neighborhood/city comparing Bushwick with all Upstate, NY cities
  - An analysis of demographic makeup of neighborhood and cities