

# **Battle of the Neighborhoods – New York**

## **Introduction**

New York – The concrete jungle where dreams are made of. Everybody dreams to live in New York, but the place is huge, with numerous counties, and then numerous neighborhoods within each county. Analysis can be done on numerous data to compare different neighborhoods – e.g. crime rates, air quality, school quality etc. However, that will not be our focus today, as we explore using foursquare API to determine the venues in a county, and cluster neighborhoods in the county based on what type of facilities and venues.

Tompkins County has been regarded as one of the best counties to live in New York, and we will be conducting our analysis on the neighborhoods in Tompkins to determine places to live there. This same analysis can be replicated on other counties in New York or other states to segment the neighbourhood into clusters based on their nearby venues and facilities.

We will be using k-means-clustering as our method of clustering the neighborhoods in the county.

## **Data**

The dataset for this project is obtained from <https://simplemaps.com/data/us-cities>. The data contains many columns, namely City Name, County Code, County Name, Density, Id, Latitude, Longitude, Source, State Id, State Name, and Time Zone. I will remove some of these columns later that we will not be using in our computations.

## **Methodology**

The first step is to always perform exploratory data analysis by examining tables, plots of the data. We then segment the data into cities/neighborhoods in Tompkins county in New York as that is our focus.

To explore our neighborhoods in Tompkins county, we use the Foursquare API. We deploy the API to find venues within a 1km radius of each neighbourhood.

We then segment the neighbourhood into 4 clusters with k-means clustering.

## **Results**

Our Tompkins County data.

	Neighborhood	State	County	lat	lng	density
8860	South Hill	NY	Tompkins	42.4113	-76.4883	438.0
8981	Dryden	NY	Tompkins	42.4912	-76.2996	474.0
9012	Cayuga Heights	NY	Tompkins	42.4680	-76.4874	819.0
9048	Northeast Ithaca	NY	Tompkins	42.4703	-76.4623	837.0
9229	Lansing	NY	Tompkins	42.4901	-76.4856	299.0
9238	Trumansburg	NY	Tompkins	42.5410	-76.6618	505.0
9274	Forest Home	NY	Tompkins	42.4530	-76.4702	917.0
9390	East Ithaca	NY	Tompkins	42.4263	-76.4627	499.0
9470	Groton	NY	Tompkins	42.5875	-76.3630	524.0
9546	Newfield Hamlet	NY	Tompkins	42.3582	-76.5923	222.0
9759	Northwest Ithaca	NY	Tompkins	42.4706	-76.5414	156.0
9774	Freeville	NY	Tompkins	42.5114	-76.3457	187.0
9860	Ithaca	NY	Tompkins	42.4442	-76.5032	2221.0

The top 3 venues in each neighbourhood in Tompkins County.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Cayuga Heights	Shopping Mall	Café	Gym / Fitness Center
1	Dryden	Pharmacy	Grocery Store	Gas Station
2	East Ithaca	Trail	Cosmetics Shop	Park
3	Forest Home	Food Truck	College Lab	Theater
4	Freeville	Breakfast Spot	Ice Cream Shop	Sports Club
5	Groton	Pizza Place	Gas Station	Gift Shop
6	Ithaca	Coffee Shop	Thai Restaurant	Bookstore
7	Lansing	Clothing Store	Shoe Store	Japanese Restaurant
8	Newfield Hamlet	Yoga Studio	Deli / Bodega	Diner
9	Northeast Ithaca	Playground	Park	Shopping Mall
10	Northwest Ithaca	Farm	Gift Shop	Museum
11	South Hill	Concert Hall	Student Center	Park
12	Trumansburg	American Restaurant	Brewery	Golf Course

We then cluster our neighbourhoods with k-means clustering.

	Neighborhood	State	County	lat	lng	density	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
8860	South Hill	NY	Tompkins	42.4113	-76.4883	438.0	1	Concert Hall	Student Center	Park
8981	Dryden	NY	Tompkins	42.4912	-76.2996	474.0	1	Pharmacy	Grocery Store	Gas Station
9012	Cayuga Heights	NY	Tompkins	42.4680	-76.4874	819.0	3	Shopping Mall	Café	Gym / Fitness Center
9048	Northeast Ithaca	NY	Tompkins	42.4703	-76.4623	837.0	1	Playground	Park	Shopping Mall
9229	Lansing	NY	Tompkins	42.4901	-76.4856	299.0	0	Clothing Store	Shoe Store	Japanese Restaurant
9238	Trumansburg	NY	Tompkins	42.5410	-76.6618	505.0	1	American Restaurant	Brewery	Golf Course
9274	Forest Home	NY	Tompkins	42.4530	-76.4702	917.0	1	Food Truck	College Lab	Theater
9390	East Ithaca	NY	Tompkins	42.4263	-76.4627	499.0	1	Trail	Cosmetics Shop	Park
9470	Groton	NY	Tompkins	42.5875	-76.3630	524.0	1	Pizza Place	Gas Station	Gift Shop
9546	Newfield Hamlet	NY	Tompkins	42.3582	-76.5923	222.0	2	Yoga Studio	Deli / Bodega	Diner
9759	Northwest Ithaca	NY	Tompkins	42.4706	-76.5414	156.0	1	Farm	Gift Shop	Museum
9774	Freeville	NY	Tompkins	42.5114	-76.3457	187.0	2	Breakfast Spot	Ice Cream Shop	Sports Club
9860	Ithaca	NY	Tompkins	42.4442	-76.5032	2221.0	1	Coffee Shop	Thai Restaurant	Bookstore

Our cluster labels are generated through k-means clustering.

We can then establish which cluster has what types of venues.

	State	density	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
9229	NY	299.0	0	Clothing Store	Shoe Store	Japanese Restaurant

Cluster 0 is by itself, mainly with fashion stores.

	State	density	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
8860	NY	438.0	1	Concert Hall	Student Center	Park
8981	NY	474.0	1	Pharmacy	Grocery Store	Gas Station
9048	NY	837.0	1	Playground	Park	Shopping Mall
9238	NY	505.0	1	American Restaurant	Brewery	Golf Course
9274	NY	917.0	1	Food Truck	College Lab	Theater
9390	NY	499.0	1	Trail	Cosmetics Shop	Park
9470	NY	524.0	1	Pizza Place	Gas Station	Gift Shop
9759	NY	156.0	1	Farm	Gift Shop	Museum
9860	NY	2221.0	1	Coffee Shop	Thai Restaurant	Bookstore

Cluster 1 mainly consists of entertainment and leisure venues.

	State	density	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
9546	NY	222.0	2	Yoga Studio	Deli / Bodega	Diner
9774	NY	187.0	2	Breakfast Spot	Ice Cream Shop	Sports Club

Cluster 2 mainly consists of eateries.

	State	density	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
9012	NY	819.0	3	Shopping Mall	Café	Gym / Fitness Center

Cluster 3 is once again by its lonesome, with a shopping mall, café, and gyms.

## **Discussion**

From our analysis, it is clear that it is not easy to come to an easy conclusion about which neighbourhood is best. Coming from Singapore myself, although my area does not have much good food and eateries which is important to me, I still enjoy living here, and do not mind commuting to other areas for food. Similarly, many people who go to prestigious schools do not live near the area, and come from other neighbourhoods to attend this particular school, because their neighbourhood is more affordable to live in.

However, this analysis is a useful beginner level analysis of neighbourhood data, and shows its potential. In the future, I will be able to use these skills of using the foursquare API to conduct better analyses with geographical data.