

The Battle of Neighborhoods

Scope selection

- Why we selecting *New York City borough Queens and Toronto borough North York?*
- *1) Diversity*
- *2) Multicultural*
- *3)*
- Problems we want to solve?
- *1) expand food markets*
- *2) beeter understand local habitats*

Data acquisition and cleaning

- *New York neighbourhoods data from https://cocl.us/new_york_dataset*
- *Toronto neighbourhoods data from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M and http://cocl.us/Geospatial_data/Geospatial_Coordinates.csv*
- *Basic data format after cleaning :*
 - *Borough Neighborhood Latitude Longitude*

Foursquare data

- Use Explore endpoint based on section „food“

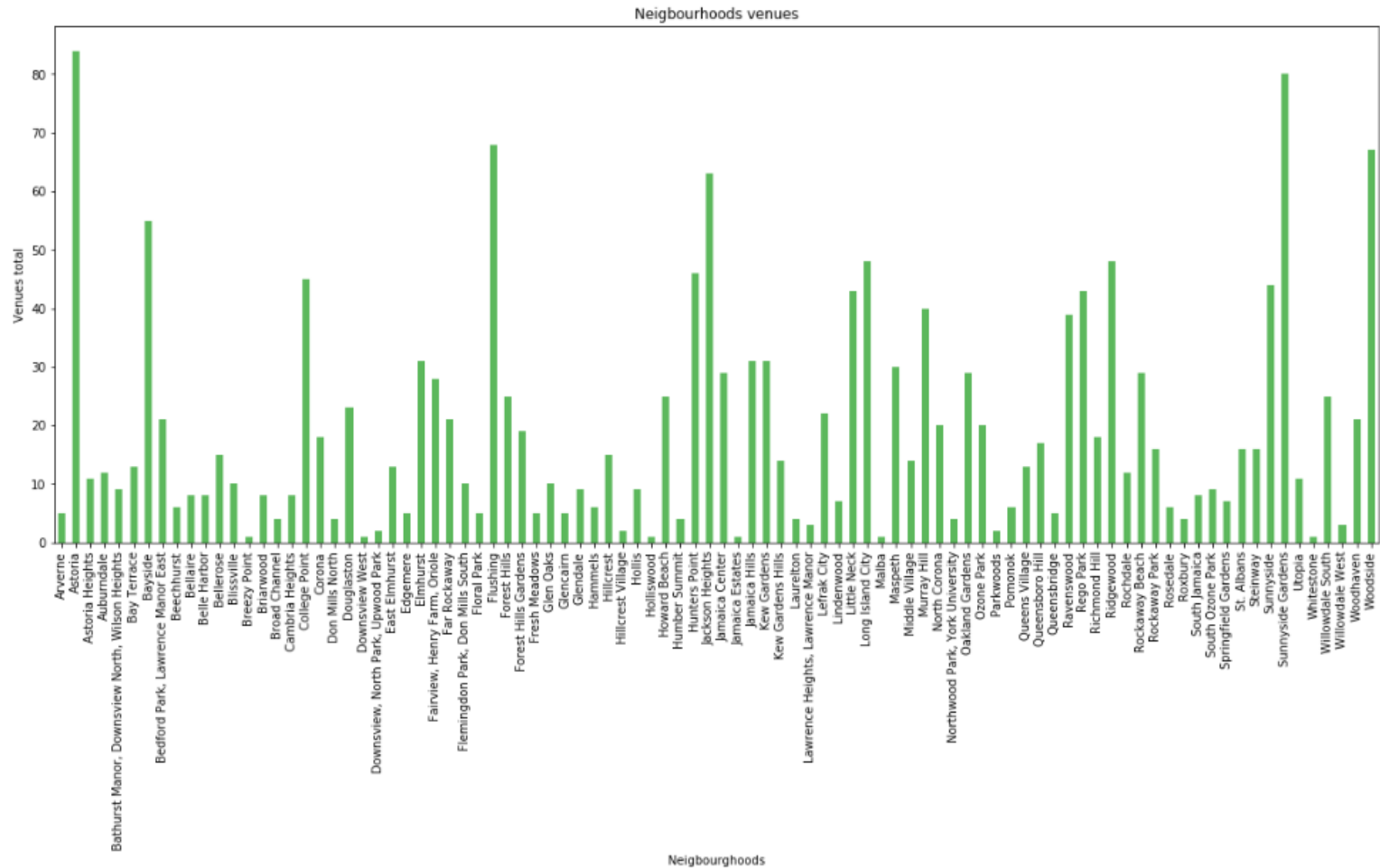
Parameters

Name	Example	Description
ll	44.3,37.2	required unless <code>near</code> is provided. Latitude and longitude of the user's location.
near	Chicago, IL	required unless <code>ll</code> is provided. A string naming a place in the world. If the <code>near</code> string is not geocodable, returns a <code>failed_geocode</code> error. Otherwise, searches within the bounds of the geocode and adds a geocode object to the response.
llAcc	10000.0	Accuracy of latitude and longitude, in meters.
alt	0	Altitude of the user's location, in meters.
altAcc	10000.0	Accuracy of the user's altitude, in meters.
radius	250	Radius to search within, in meters. If radius is not specified, a suggested radius will be used based on the density of venues in the area. The maximum supported radius is currently 100,000 meters.
section	food	One of <code>food</code> , <code>drinks</code> , <code>coffee</code> , <code>shops</code> , <code>arts</code> , <code>outdoors</code> , <code>sights</code> , <code>trending</code> , <code>nextVenues</code> (venues frequently visited after a given venue), or <code>topPicks</code> (a mix of recommendations generated without a query from the user). Choosing one of these limits results to venues with the specified category or property.
query	tacos	A term to be searched against a venue's tips, category, etc. The query parameter has no effect when a <code>section</code> is specified.

Foursquare venues results

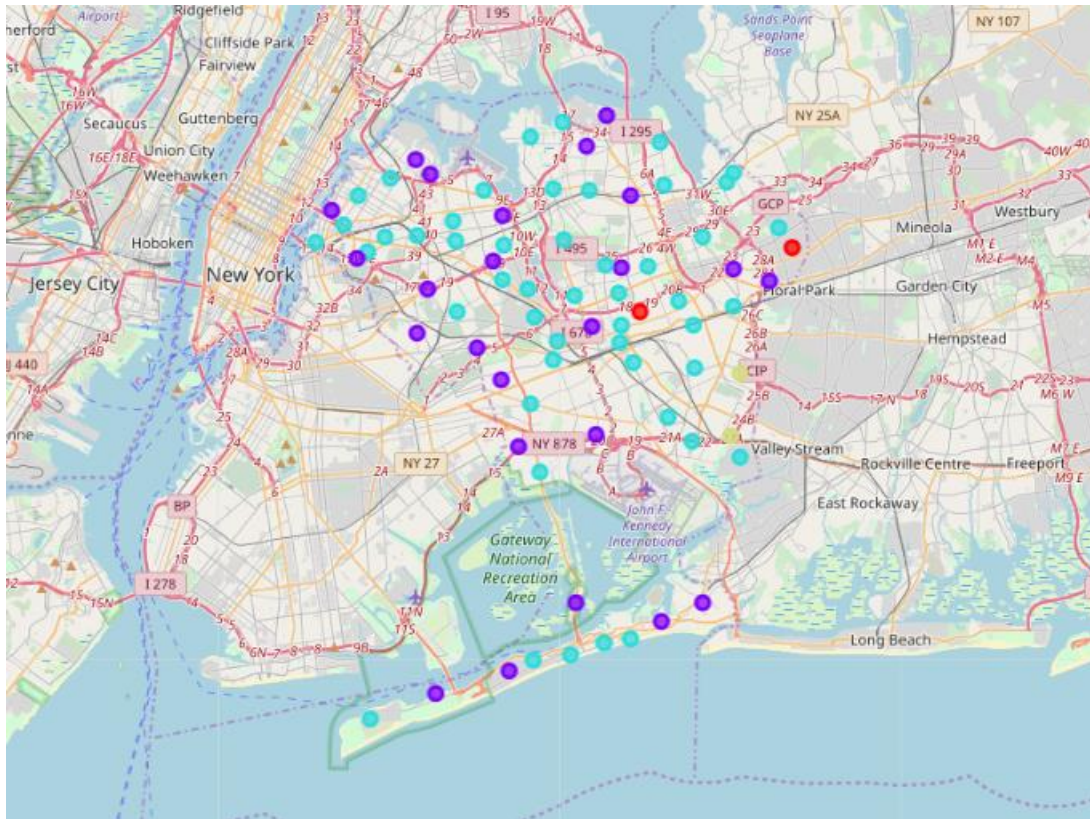
	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Arverne	5	5	5	5	5	5
Astoria	84	84	84	84	84	84
Astoria Heights	11	11	11	11	11	11
Auburndale	12	12	12	12	12	12
Bathurst Manor, Downsview North, Wilson Heights	9	9	9	9	9	9
Bay Terrace	13	13	13	13	13	13
Bayside	55	55	55	55	55	55
Bedford Park, Lawrence Manor East	21	21	21	21	21	21
Beechhurst	6	6	6	6	6	6
Bellaire	8	8	8	8	8	8
Belle Harbor	8	8	8	8	8	8
Bellerose	15	15	15	15	15	15

Food venues counts



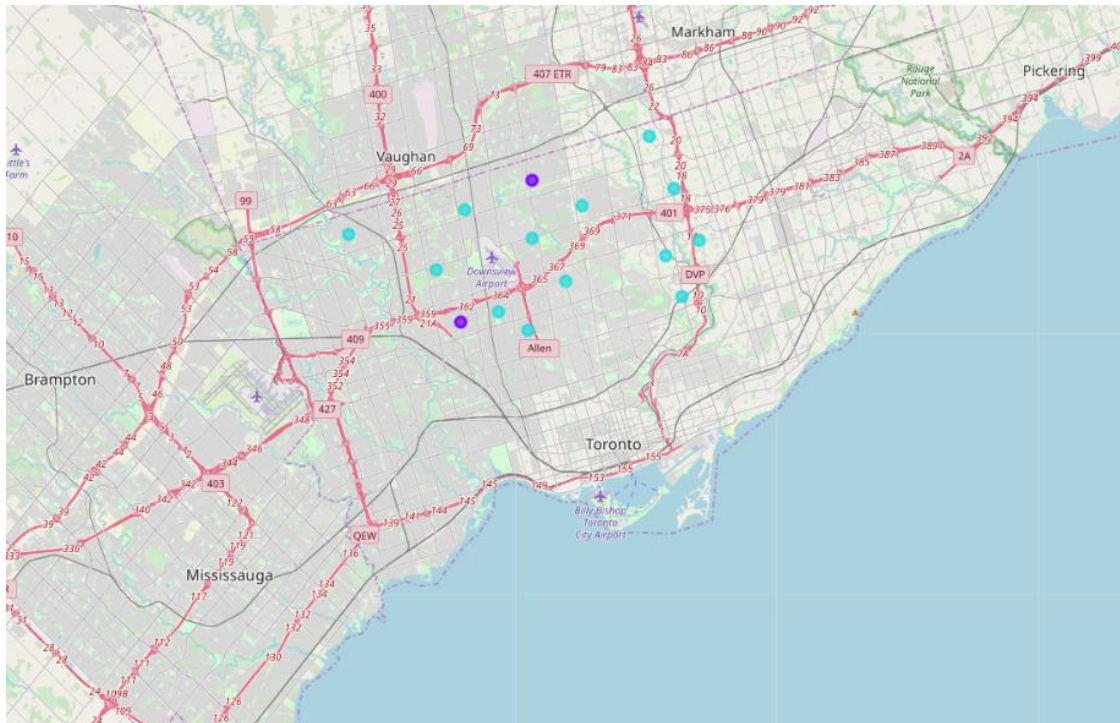
Results 4 k mens clusters

- *New York City borough Queens*

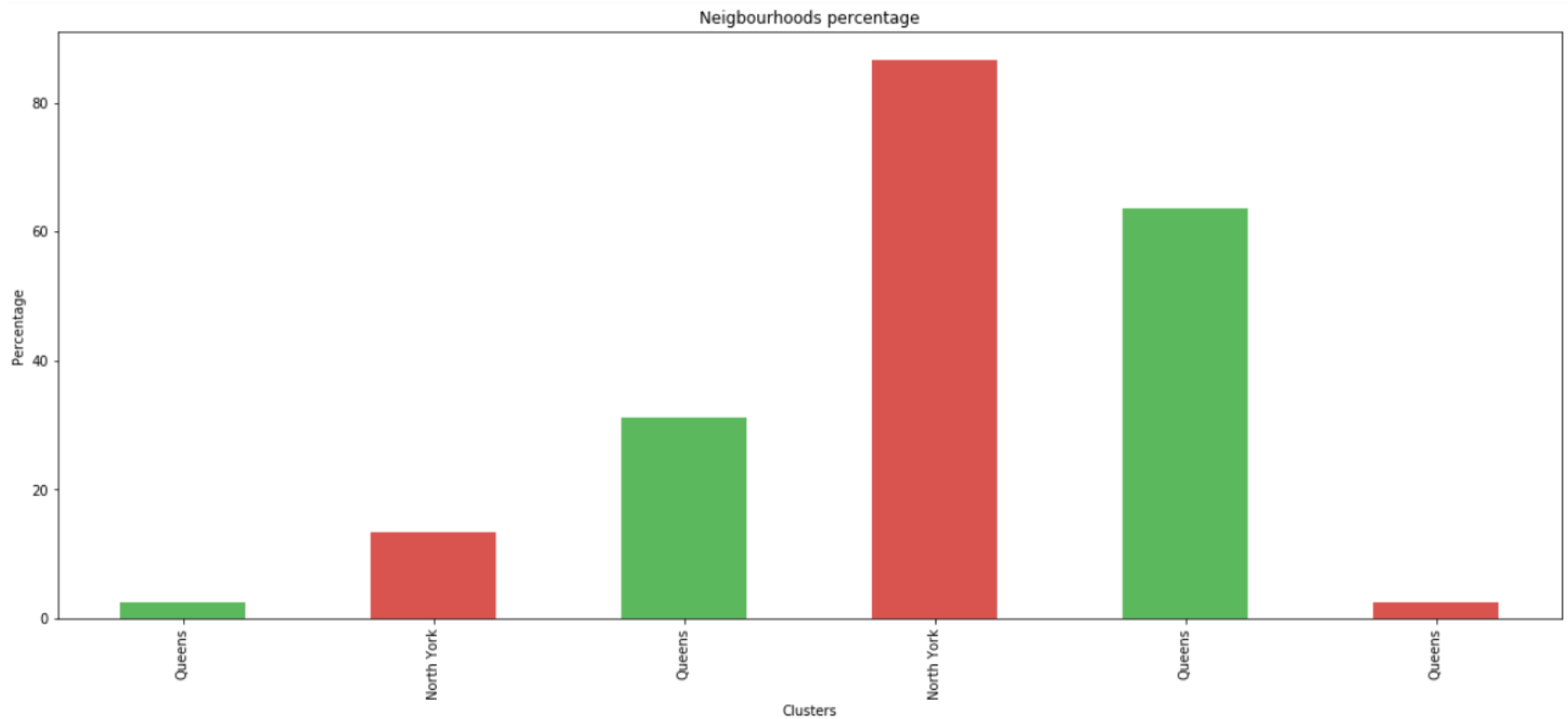


Results 4 k mens clusters

- *Toronto borough North York*



Cluster representation in percents



Conclusions and future directions

- Built useful model to choose which food industry venue better suits
- Accuracy of the clustering has room for improvement
- Four square API works good in USA but in another countries lack data