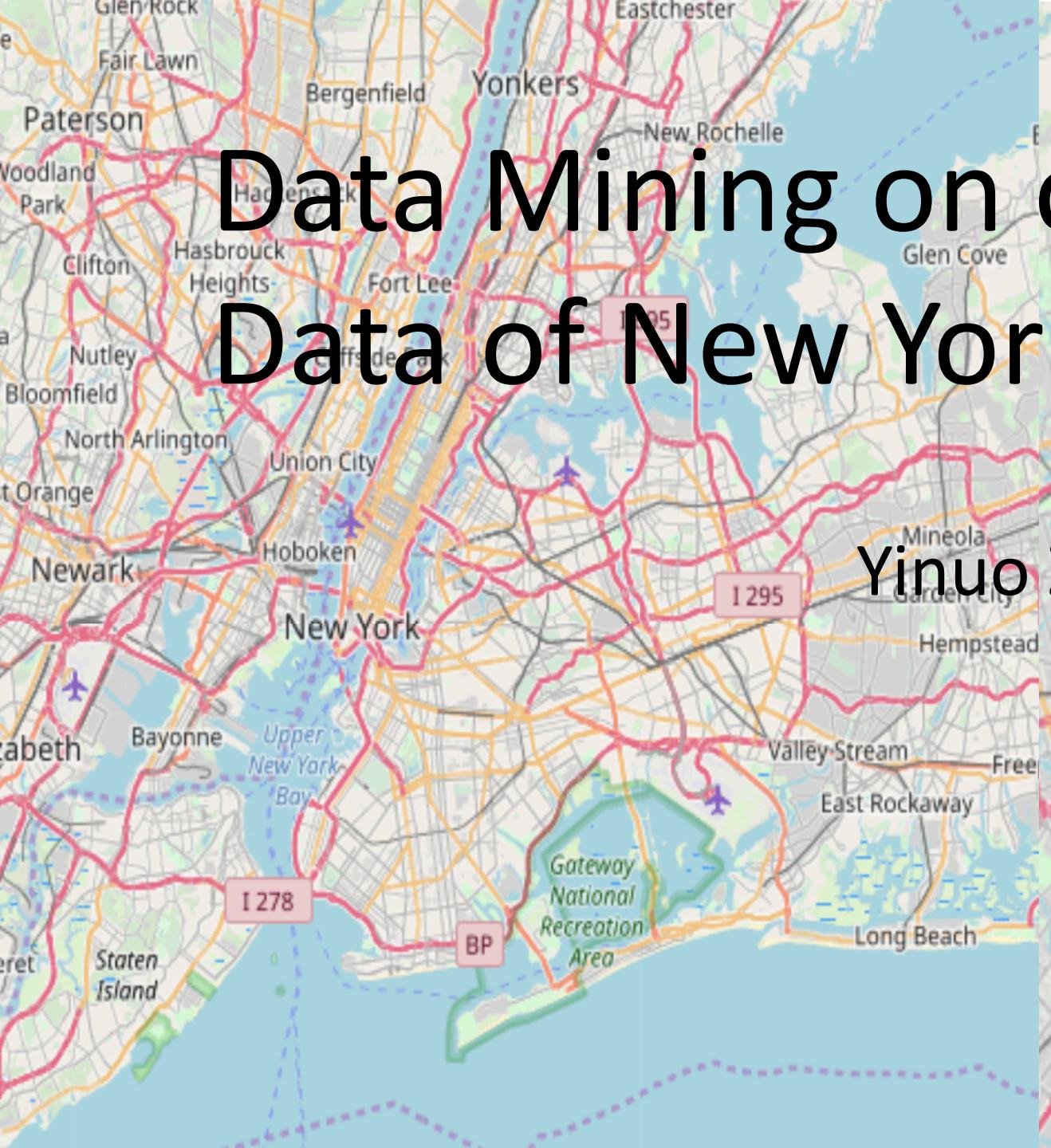


Data Mining on Geographical Data of New York and Toronto

Yinuo Zeng



Goal

- Investigate the similarity of New York City and City of Toronto and find the common characteristics of them.
- Answer business problems like which area in a city should be invested and what kind of business activities in this area are suggested

Data

- New York City data (<https://geo.nyu.edu/catalog/> nyu_2451_34572)
 - 306 observations and 3 features
- New York City venues and venues categories data
 - 10233 observations and 7 features
- City of Toronto data (<https://www.toronto.ca/>)
 - 138 observation and 3 features
- City of Toronto venues and venues categories data
 - 2130 observations and 7 features
- Other features (Foursquare API)

EDA

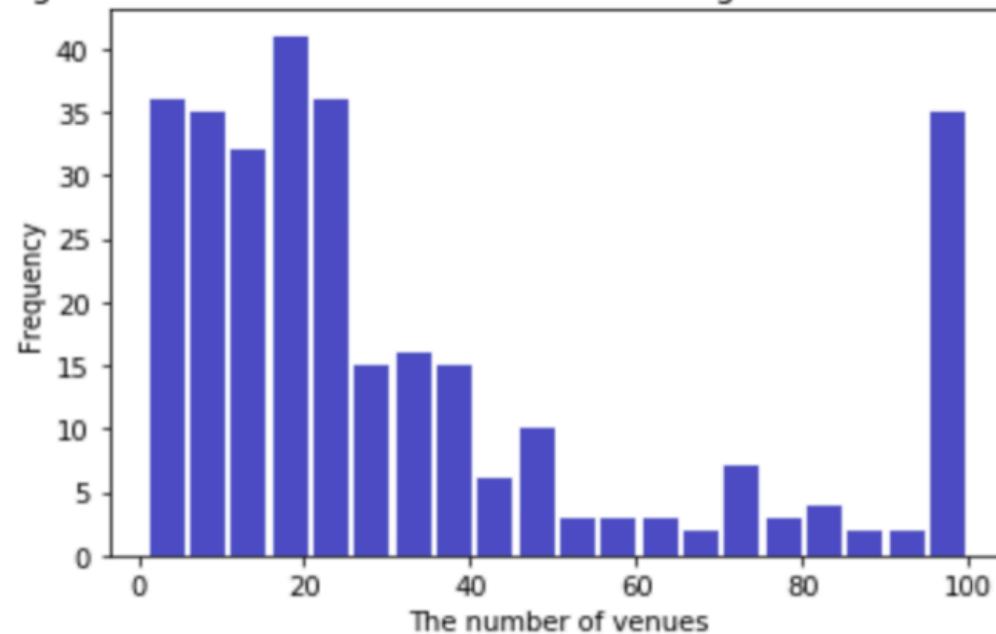
Table of the number of Venues in New York City

Venues	Counts
Dunkin'	151
Chase Bank	91
Subway	68
Rite Aid	55
Starbucks	42
T-Mobile	37
Baskin-Robbins	35
CVS pharmacy	32
Popeyes Louisiana Kitchen	31
Walgreens	30

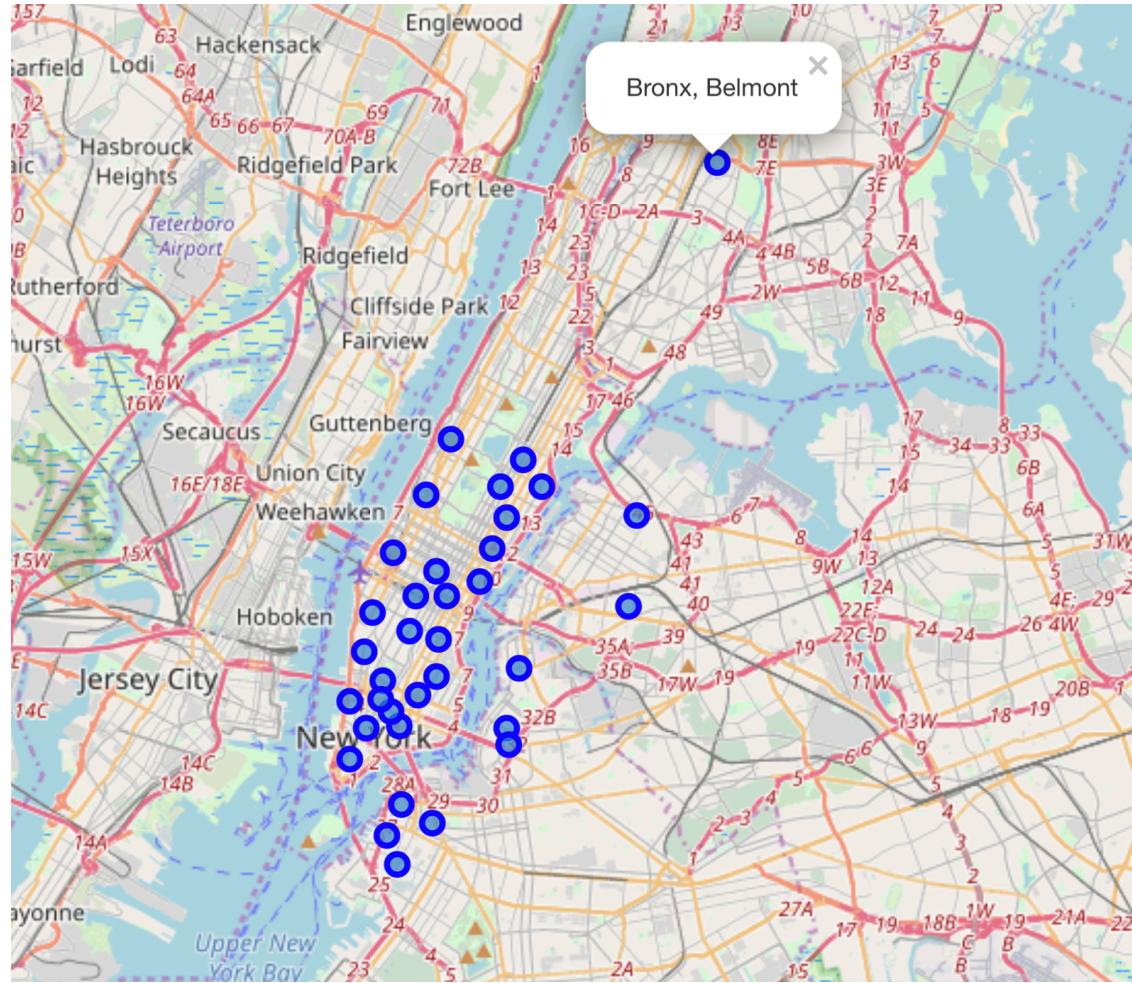
Table of the number of Venues Categories in New York City

Venues Categroy	Counts
Pizza Place	434
Italian Restaurant	321
Coffee Shop	281
Deli / Bodega	262
Bakery	229
Bar	218
Chinese Restaurant	214
Sandwich Place	185
Mexican Restaurant	175
American Restaurant	175

Histogram of the number of venues of each neighborhood in New York City



EDA



EDA

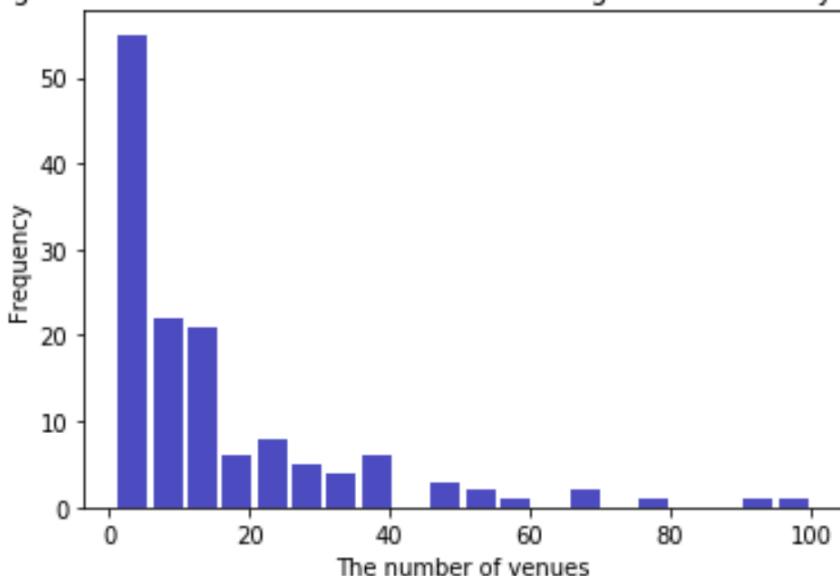
Table of the number of Venues in City of Toronto

Venues	Counts
Tim Hortons	62
Subway	43
Starbucks	35
Pizza Pizza	27
TD Canada Trust	25
Shoppers Drug Mart	22
Dollarama	14
The Beer Store	13
McDonald's	13
LCBO	12

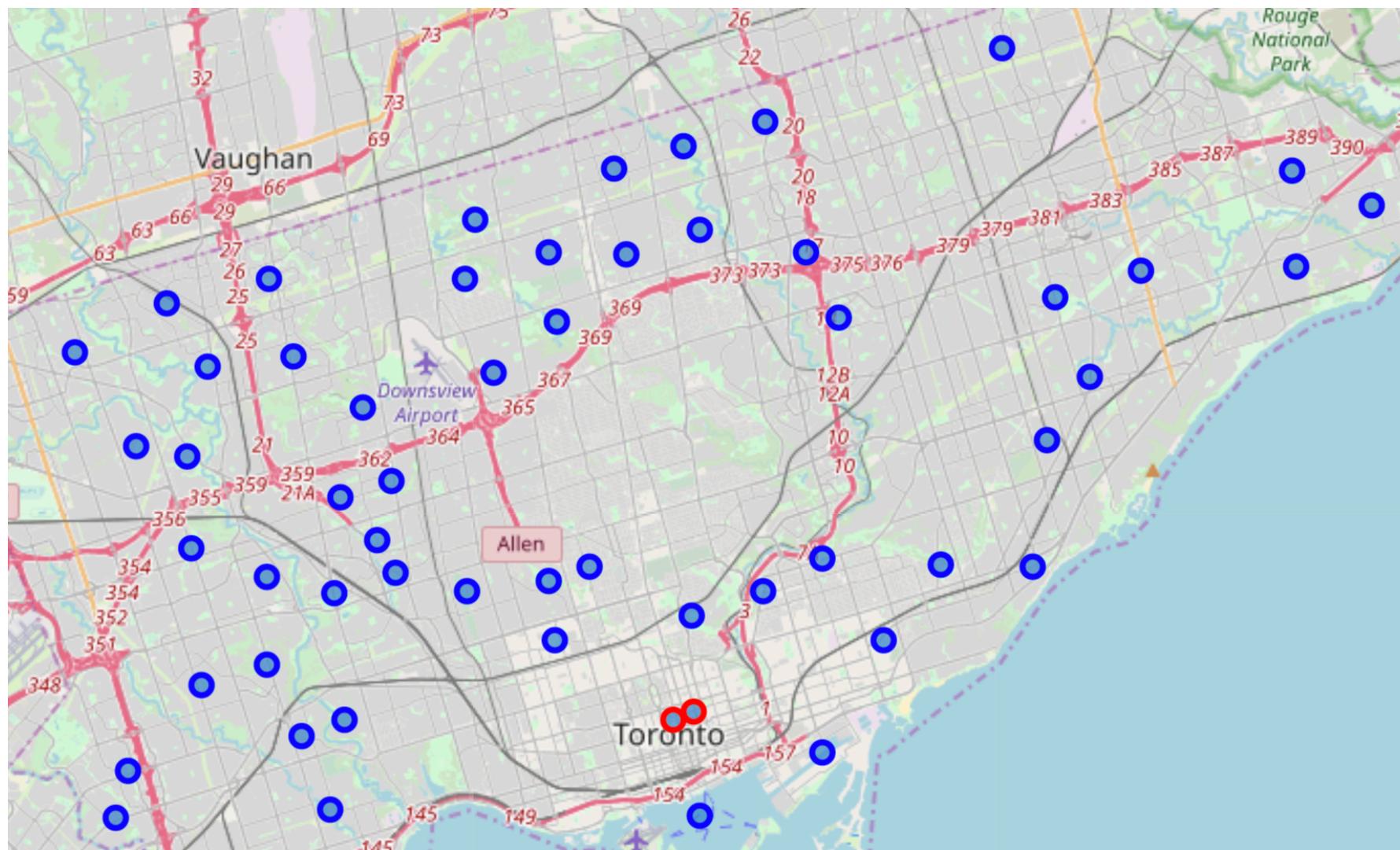
Table of the number of Venues Categories in City of Toronto

Venues Category	Counts
Coffee Shop	151
Park	89
Cafe	80
Pizza Place	76
Sandwich Place	63
Italian Restaurant	58
Fast Food Restaurant	45
Grocery Store	42
Bakery	41
Bar	41

Histogram of the number of venues of each neighborhood in City of Toronto

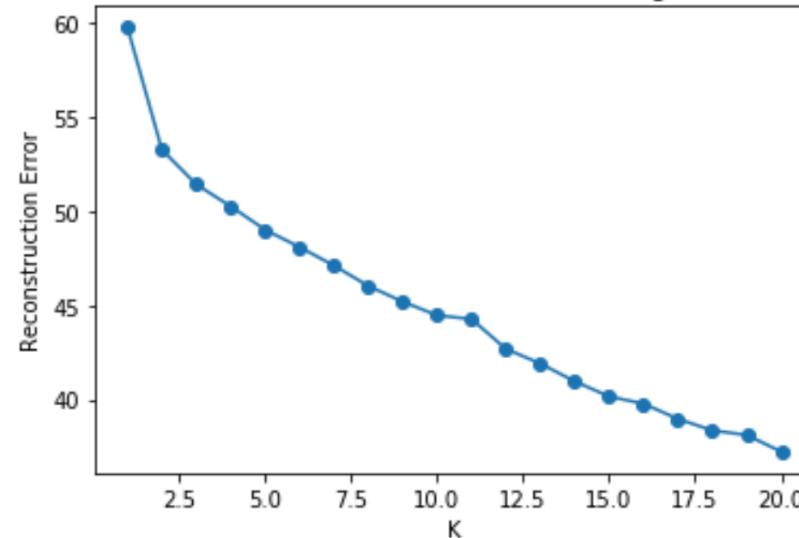


EDA



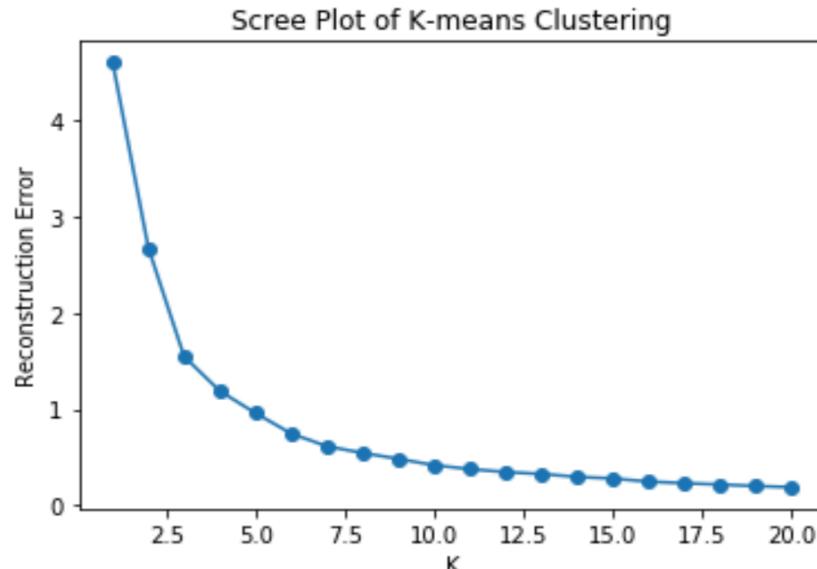
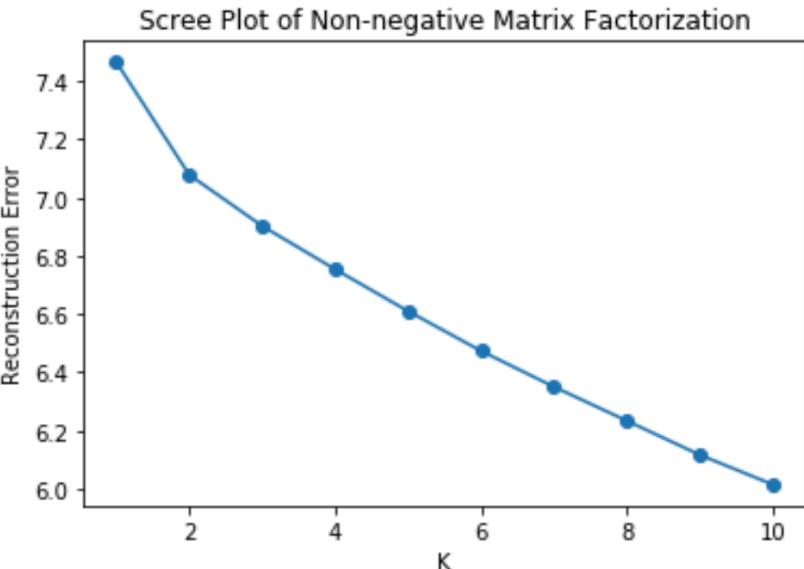
K-means

Scree Plot of K-means Clustering



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
Italian Restaurant	Italian Restaurant	Pizza Place	Pizza Place	Pizza Place	Pizza Place	Exhibit	Eye Doctor	Eye Doctor	Factory
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
Park	Park	Zoo Exhibit	Zoo Exhibit	Event Space	Exhibit	Eye Doctor	Factory	Falafel Restaurant	Farm
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
Bus Stop	Coffee Shop	Pizza Place	Zoo Exhibit	Financial or Legal Service	Event Space	Exhibit	Factory	Factory	Falafel Restaurant

NMF + K-means



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
Park	Park	Coffee Shop	Zoo Exhibit	Pizza Place	Exhibit	Eye Doctor	Factory	Falafel Restaurant	Farm
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
Italian Restaurant	Italian Restaurant	Pizza Place	Pizza Place	Pizza Place	Exhibit	Exhibit	Eye Doctor	Factory	Falafel Restaurant
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
Pizza Place	Pizza Place	Chinese Restaurant	Deli / Bodega	Pharmacy	Bank	Donut Shop	Fast Food Restaurant	Middle Eastern Restaurant	Falafel Restaurant

Conclusion

- Two neighborhoods in City of Toronto are missing
- Non-negative matrix factorization introduces noise
- Model selection of clustering methods