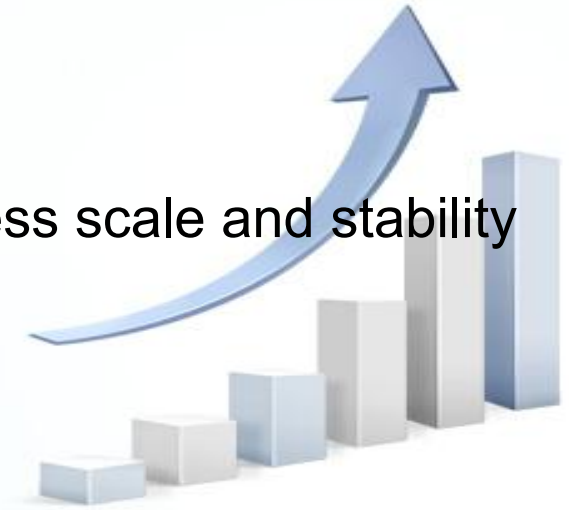
The Battle of Neighborhoods

Applied Data Science Capstone Project

YW @ 6/15/2019

Background

- Who is Food-In Pro?
  - Supply food ingredients and food additives
- New plan to open a small distribution warehouse in Toronto
  - What is the part of the business development for this project?
- Contribution/referene from sales and marketing strategies
  - Which type of clients will be targeted (client's categories)?
  - A rating system to identify the levels of clients based on business scale and stability



Problem Description

- Assignment to the data analyst:
  - Give a recommendation to the facility address searching group about the best area/neighborhood(s) to locate the wareshouse in the city of Toronto.

- Priority principles for recommendation:

  1) more amount of high rate clients (especially supermarkets and groceries)
  2) more amount of total clients

Data Acquisition and Cleaning

- Data of neighborhoods: Wikipedia (scraped by Beautifulsoup)
- Geographical coordinates from the course material
- Location data: querying Foursquare API
  - query keyboards: supermarket, restaurants and ice cream
- Data cleaning
  - Only the neighborhoods of “Toronto boroughs” are selected
  - Drop “not assigned” postcodes
  - Drop unreasonable client category (e.g. Pharmacy)

What the original dataframe look like...

Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Ryerson	43.657162	-79.378937	Metro	43.658404	-79.376748	Supermarket
Garden District	43.657162	-79.378937	Metro	43.658404	-79.376748	Supermarket
St. James Town	43.651494	-79.375418	Metro	43.649027	-79.373313	Supermarket
Berczy Park	43.644771	-79.373306	Loblaws	43.644462	-79.369486	Supermarket
Berczy Park	43.644771	-79.373306	Metro	43.649027	-79.373313	Supermarket
Central Bay Street	43.657952	-79.387383	The Market by Longo's Elizabeth	43.655357	-79.385115	Supermarket
Central Bay Street	43.657952	-79.387383	Metro	43.660569	-79.383768	Supermarket
Christie	43.669542	-79.422564	Fiesta Farms	43.668471	-79.420485	Supermarket
Christie	43.669542	-79.422564	Loblaws	43.671807	-79.421102	Supermarket
Adelaide	43.650571	-79.384568	Rabba Marché	43.649216	-79.386908	Supermarket

Features:

- 2802 venues
- coordinates
- venue category

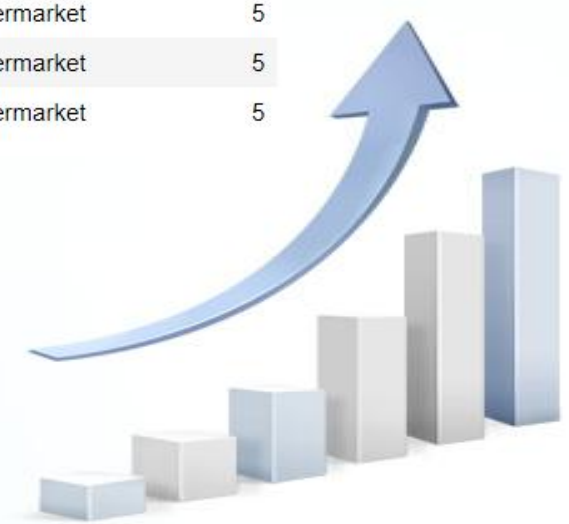
Methodology

- Examine the number of venues in each neighborhood
- Examine the rates of venues in each neighborhood
  - any variance against the number of venues in ranking?
- Examine the distribution of supermarket/grocery
- Examine the common venues by clustering
- Project is achieved by Python 3 in Jupyter Notebook

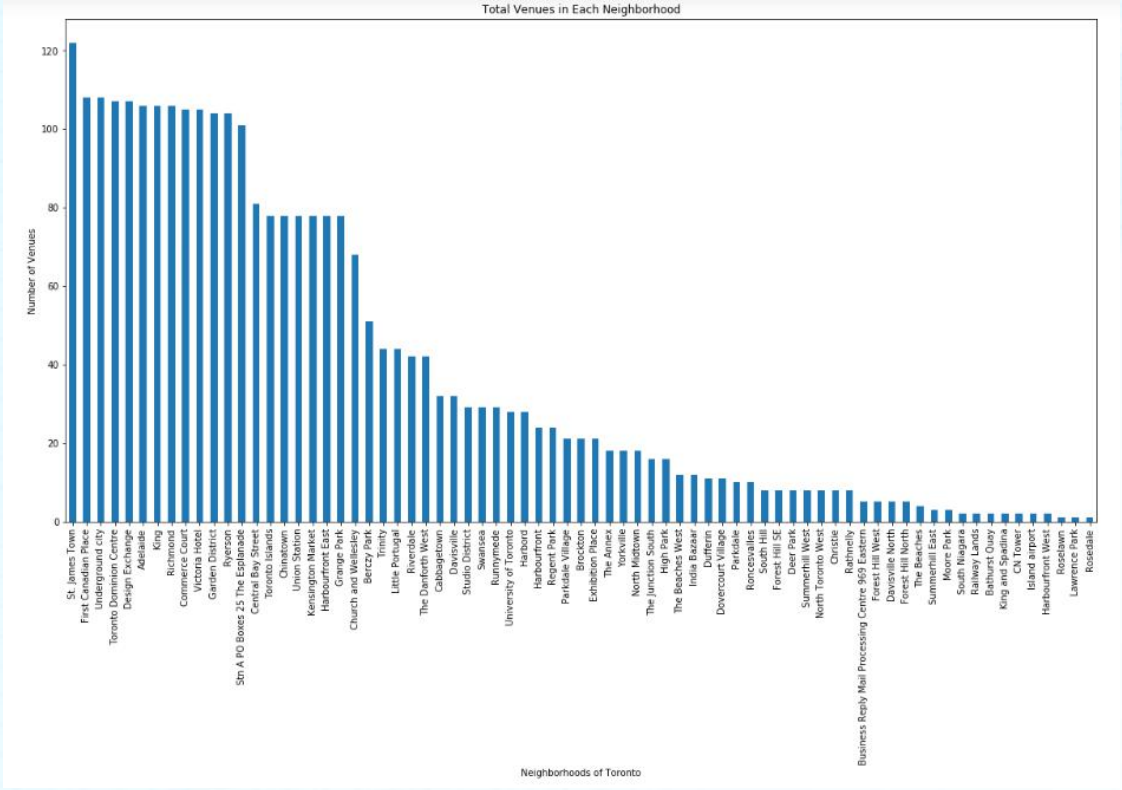
  https://github.com/Nuercom/Forlearn/blob/master/Final.ipynb

Add rates according to venue category...

Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Category Rate
Ryerson	43.657162	-79.378937	Metro	43.658404	-79.376748	Supermarket	5
Garden District	43.657162	-79.378937	Metro	43.658404	-79.376748	Supermarket	5
St. James Town	43.651494	-79.375418	Metro	43.649027	-79.373313	Supermarket	5
Berczy Park	43.644771	-79.373306	Loblaws	43.644462	-79.369486	Supermarket	5
Berczy Park	43.644771	-79.373306	Metro	43.649027	-79.373313	Supermarket	5
Central Bay Street	43.657952	-79.387383	The Market by Longo's Elizabeth	43.655357	-79.385115	Supermarket	5
Central Bay Street	43.657952	-79.387383	Metro	43.660569	-79.383768	Supermarket	5
Christie	43.669542	-79.422564	Fiesta Farms	43.668471	-79.420485	Supermarket	5
Christie	43.669542	-79.422564	Loblaws	43.671807	-79.421102	Supermarket	5
Adelaide	43.650571	-79.384568	Rabba Marché	43.649216	-79.386908	Supermarket	5

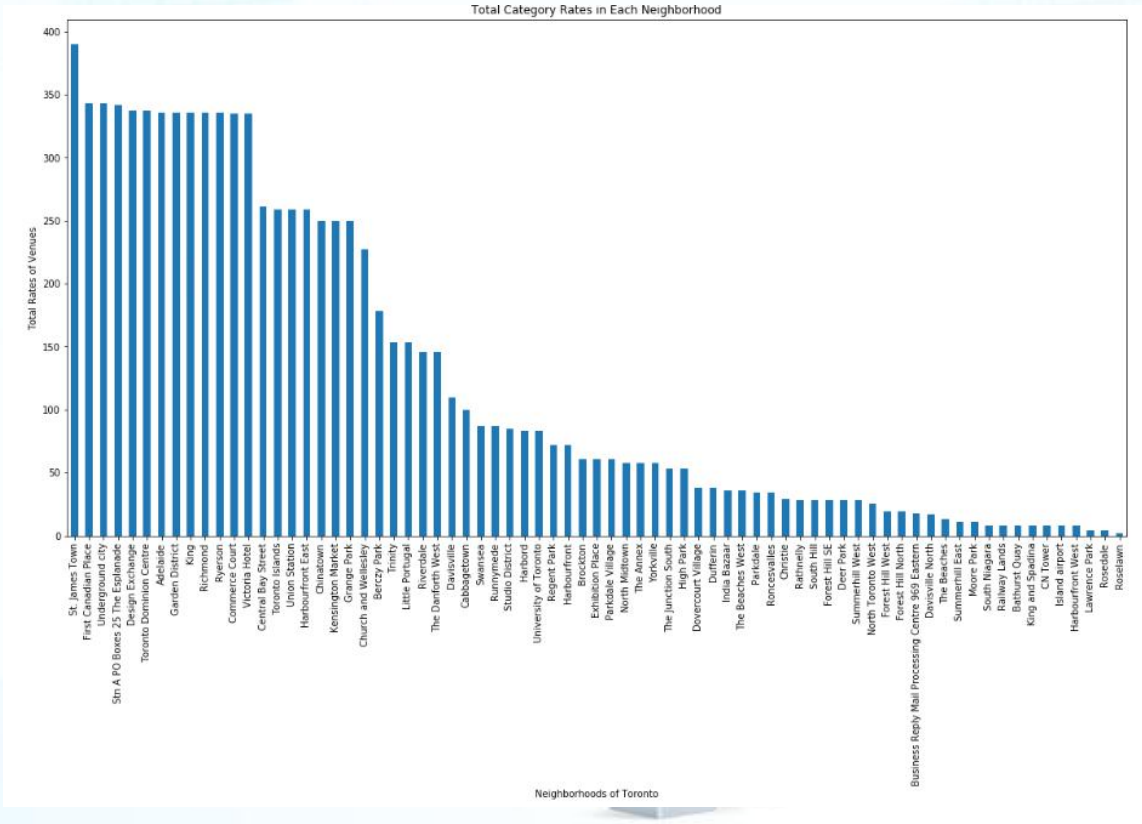


Comparison of venues ranked by numbers and rates

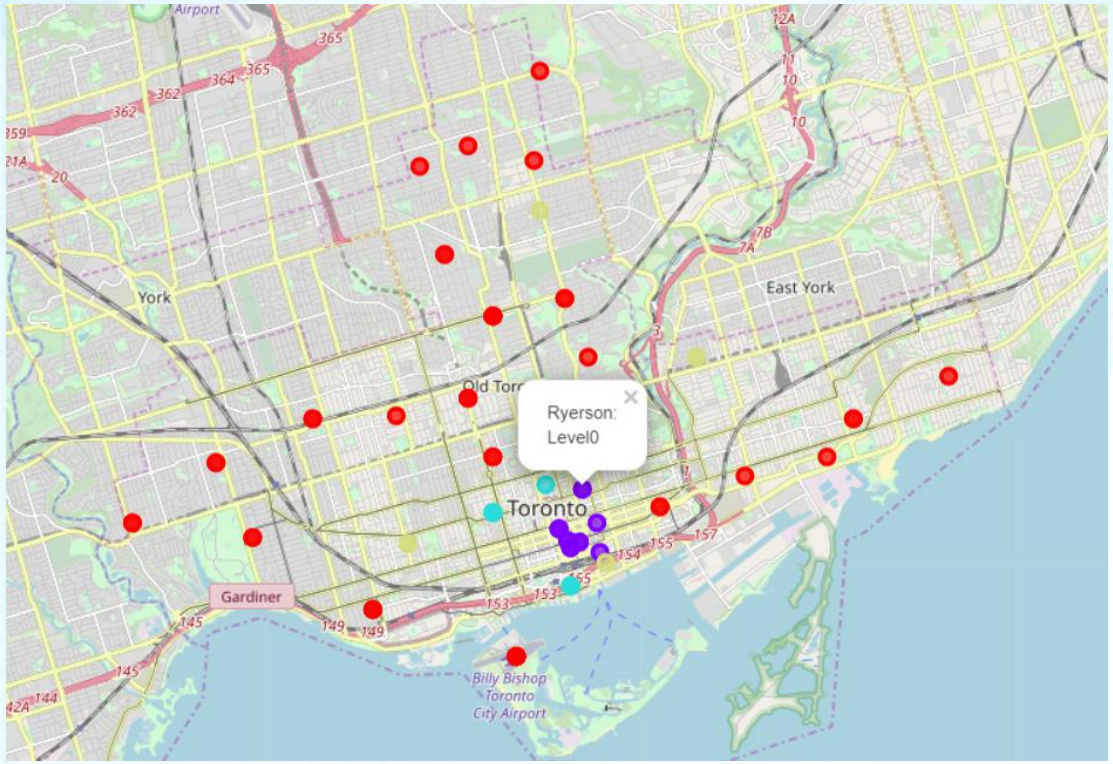


Number of Venues

Rate of Venues



Map of grouped neighborhood by category rates



What the distribution is like?

Neighborhoods which have 2 or more supermarkets/groceries

Neighborhood	Supermarket/grocery
Kensington Market	5
Chinatown	5
Grange Park	5
Dufferin	2
Christie	2
Dovercourt Village	2
Stn A PO Boxes 25 The Esplanade	2
St. James Town	2
Berczy Park	2
Central Bay Street	2

What are their rates
from previous results

Cluster neighborhoods for common venues

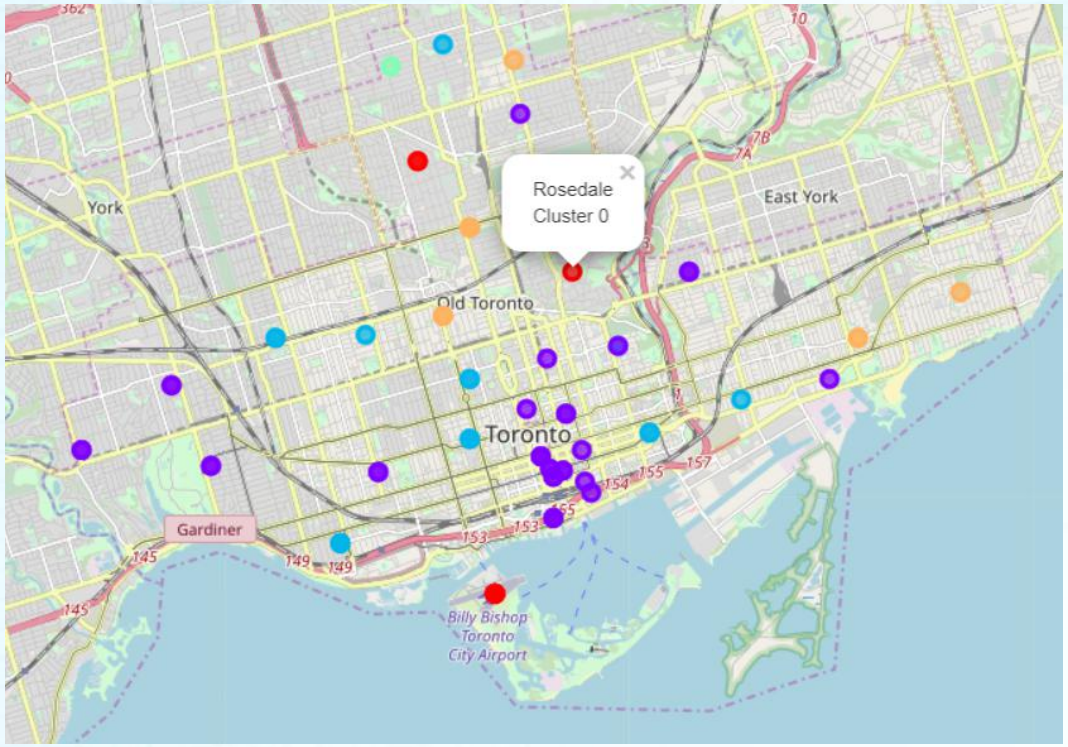
- Refer the notebook for clustering
  - k-mean clustering where k=5

- II, III and IV category occupies the top 2 common venues

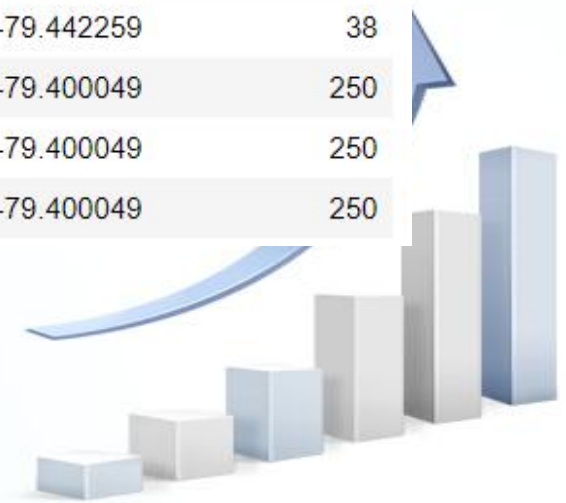
- “Supermarket” neighborhood could not be idenified.

  However,.....

Observation of neighborhoods from non-bottom venues of supermarket/grocery in those clusters



Neighbourhood	Latitude	Longitude	Category_Rate
Berczy Park	43.644771	-79.373306	178
The Danforth West	43.679557	-79.352188	146
Riverdale	43.679557	-79.352188	146
Parkdale	43.648960	-79.456325	34
Roncesvalles	43.648960	-79.456325	34
Christie	43.669542	-79.422564	29
Dovercourt Village	43.669005	-79.442259	38
Dufferin	43.669005	-79.442259	38
Chinatown	43.653206	-79.400049	250
Grange Park	43.653206	-79.400049	250
Kensington Market	43.653206	-79.400049	250



Recommendation

- Comprehensively review the results against the principles.
- Chinatown, Grange Park and Kensington Market win the principle 1), and have the medium level against principle 2). They are the recommended neighborhoods therefore.
- Since they are adjacent, the area is recommended.

Conclusion

&
Question?