

# Applied Data Science Capstone Coursera

The Battle of Neighbourhoods  
Selecting location for a new health cafe in North York,  
Canada

Srishti Arora

21-09-2020

# Problem Statement

If someone is looking to open a health café in North York Canada,  
where would you recommend that they open it?

# Introduction

- Assumption - new cafe is a health cafe serving high protein, low calorie drinks and snacks
- Important aspects to consider
  - Who is the target audience
  - Which neighbourhood would provide the maximum accessibility to that audience
  - Where will the competition be the least
- Not aware of owner's budget and the property prices
- Need to provide multiple options and flexibility to select a neighbourhood

# Data Description and Collection

- a) Scrape list of postal codes of Canada from Wikipedia
- b) Read the Geospatial Coordinates csv to get latitude and longitude details
- c) Merge the datasets obtained in a) and b)
- d) Fetch top 100 venues that are within a radius of 500 meters to all the neighbourhoods in North York using Foursquare API

# Methodology

- Pick out the most and least favourable locations for setting up a new health café in North York, Canada
- By web scrapping Wikipedia and merging it with the Geospatial Coordinates csv

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
3	M3B	North York	Don Mills	43.745906	-79.352188
4	M6B	North York	Glencairn	43.709577	-79.44507

# Methodology

- Exploiting the Foursquare API

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
2	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
3	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
4	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop

# Methodology - Which neighbourhoods to prefer?

- Where can he find customers who will prefer health over money or taste?
- Someone who is visiting a gym, fitness centre or sports related venue is most probably health conscious
- Would like to know which neighbourhoods have more such venues
- Although not the sole criterion to classify someone as health vigilant

# Methodology - Which neighbourhoods to avoid?

- Not just market demand but also competition is important
- Availability of multiple alternatives to eat at will reduce the footfall at the café irrespective of type of eateries in the vicinity
- Find out the neighbourhoods which already have a lot of eating joints
- Recommend against selecting those neighbourhoods

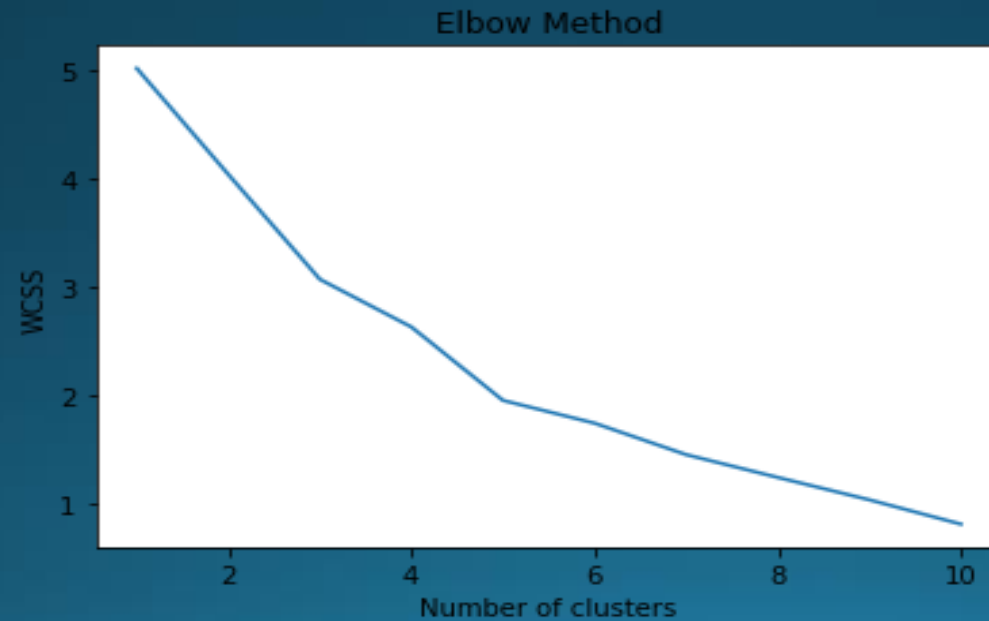


# Methodology - Alternate neighbourhoods

- Ignoring the property prices and operational costs in any particular neighbourhood
- Recommendations might not fit into the entrepreneur's budget
- Cluster the neighbourhoods based on their similarity with respect to presence of particular venue category and frequency of occurrence of each category

# Methodology - Clustering neighbourhoods

- standard elbow method to determine the number of clusters
- k-means clustering



# Results

## a. Neighbourhoods to be preferred

- Don Mills
- Downsview
- Fairview, Henry Farm, Oriole
- Hillcrest Village
- Humberlea, Emery

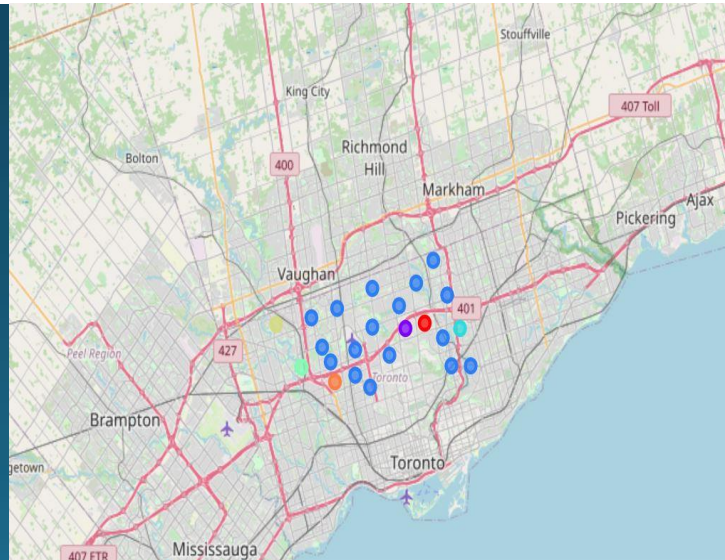
## b. Neighbourhoods to be avoided

- Fairview, Henry Farm, Oriole
- Willowdale, Willowdale East
- Bedford Park, Lawrence Manor East
- Don Mills
- Bathurst Manor, Wilson Heights, Downsview North

Neighbourhood Recommendations -  
**Downsview, Hillcrest, Humberlea and Emery**  
(set a. minus set b.)

# Clustering Results

	Postal Code	Borough	Neighborhoods	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	M3A	North York	Parkwoods	43.753259	-79.329656	3	Food & Drink Shop	Park	Women's Store	Diner	Coffee Shop
1	M4A	North York	Victoria Village	43.725882	-79.315572	2	Coffee Shop	Pizza Place	Hockey Arena	Portuguese Restaurant	Intersection
2	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	2	Clothing Store	Women's Store	Vietnamese Restaurant	Boutique	Coffee Shop
3	M3B	North York	Don Mills	43.745906	-79.352188	2	Gym	Beer Store	Coffee Shop	Japanese Restaurant	Caribbean Restaurant
4	M3C	North York	Don Mills	43.725900	-79.340923	2	Gym	Beer Store	Coffee Shop	Japanese Restaurant	Caribbean Restaurant



# Discussion

- Certain limitations to this study
- Could not consider all the important factors like:
  - Property prices
  - Population density
  - Operational costs
- Critical to be aware of the shortcomings

# Conclusion

- 80% of the work is done if
  - Understand the problem statement well
  - Know how to play with data, merge different data sources, clean and format data as per the requirements
- Data speaks a lot and beautiful insights can be generated
- Not only important to do great technical analysis, ability to write comprehensible reports is crucial