



COURSERA CAPSTONE PROJECT

IBM DATA SCIENCE PROFESSIONAL CERTIFICATE

MOHAMMAD FAHAD KAZALBASH

TABLE OF CONTENTS

- Introduction: Business Problem
- Data
- Methodology
- Results and Discussion
- Conclusion

INTRODUCTION

- Background

- Canada aims to invite 1 million immigrants over the next three years.
- Toronto is the provincial capital of Ontario and the most populous city in Canada.
- The city continues to grow and attract immigrants.

- Business Problem

- Immigrating to a new country is a challenging task.
- Choosing where to live is crucial for any immigrant and requires informed decision making.
- This project will focus to resolve this challenge.
- This project aims to aid immigrants to identify the right neighborhood to settle in Toronto, based on their personal preferences.
- The project will explore different localities of Toronto and identify various venues of interest in each neighborhood.

DATA

Source #1

- Postal codes in **Canada through Wikipedia**
- Source** - https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Postcode ↕	Borough ↕	Neighbourhood
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Harbourfront
M6A	North York	Lawrence Heights
M6A	North York	Lawrence Manor
M7A	Queen's Park	Not assigned
M8A	Not assigned	Not assigned
M9A	Downtown Toronto	Queen's Park
M1B	Scarborough	Rouge

Source #2

- Geospatial data of Toronto Neighborhoods in csv format
- Source** - http://cocl.us/Geospatial_data

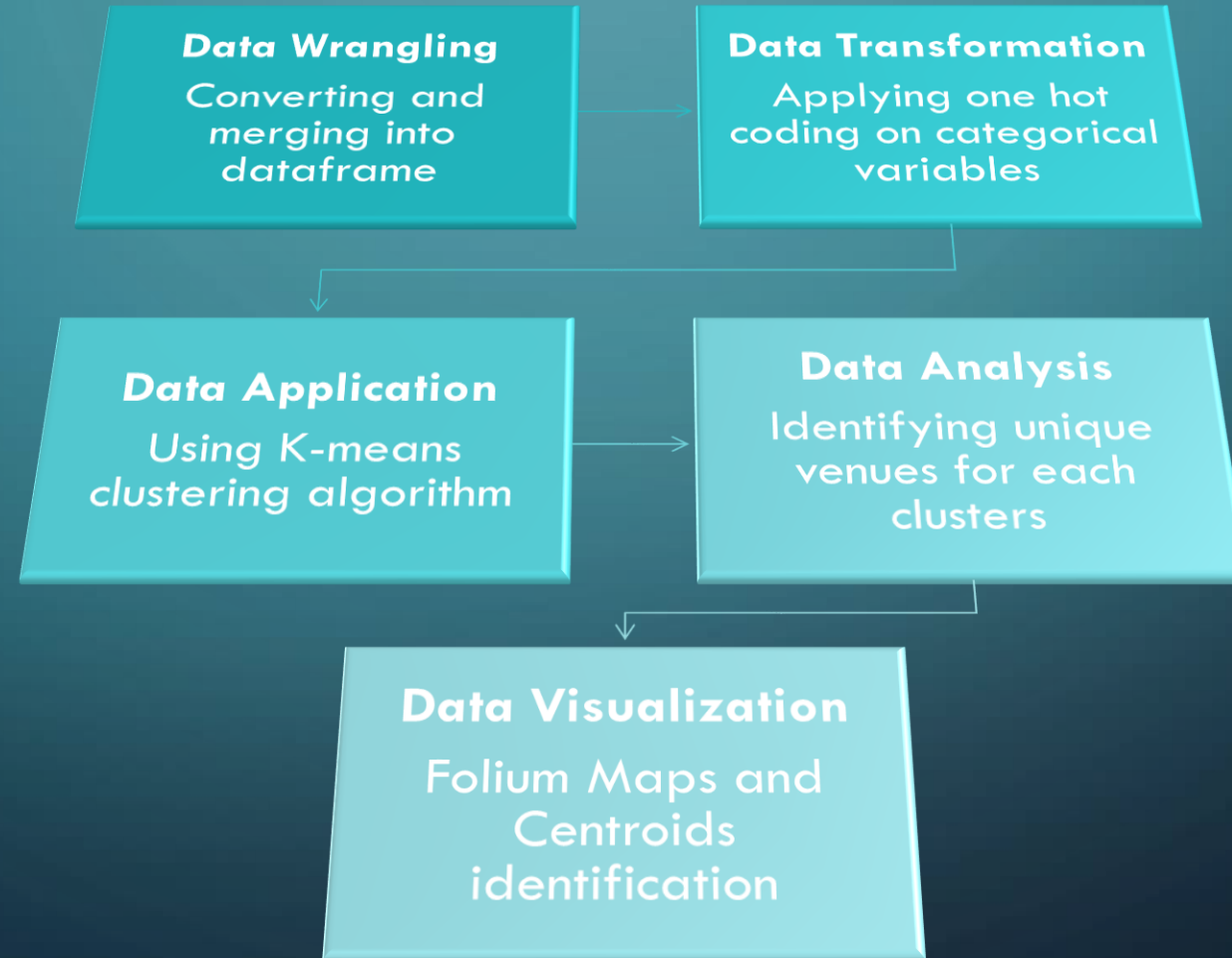
	Postcode	Borough	Neighbourhood	Latitude	Longitude	Cluster
0	M4E	East Toronto	The Beaches	43.676357	-79.293031	0
1	M4K	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188	0
2	M4L	East Toronto	The Beaches West,India Bazaar	43.668999	-79.315572	0
3	M4M	East Toronto	Studio District	43.659526	-79.340923	0
4	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	1

Source #3

- Foursquare Database for location information such as venues, tips and photos
- Source** - <http://foursquare.com>

	PostCode	Borough	Neighborhood	BoroughLatitude	BoroughLongitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	M4E	East Toronto	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	M4E	East Toronto	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	M4E	East Toronto	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	M4E	East Toronto	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	M4E	East Toronto	The Beaches	43.676357	-79.293031	Dip 'n Sip	43.678897	-79.297745	Coffee Shop

METHODOLOGY



DATA IMPORTATION

- Postal address of Toronto's neighbourhoods and their respective geospatial data i.e. longitudes and latitudes were merged in to pandas data frame format
- To explore the data, Foursquare API was used to list venues of the neighbourhoods and segment them, limited to 500m radius. 277 unique categories of venues were identified.

	PostalCode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

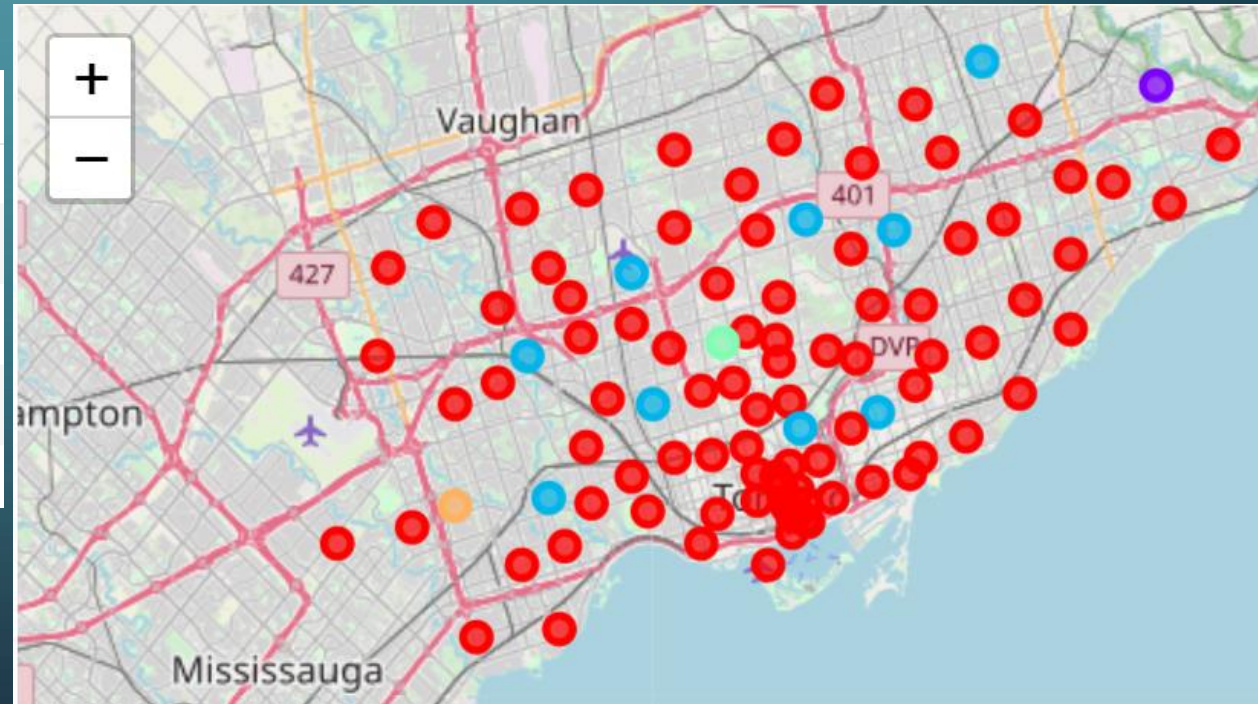
The geographical coordinate of Toronto are 43.653963, -79.387207.



DATA WRANGLING

- Venues were grouped by the neighbourhoods and the 10 most common venues were identified for each of the neighbourhoods
- Using unsupervised k-means algorithm, 5 clusters of neighbourhoods were established that represents different venues/facilities of interest around that locality

	Neighbourhood	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
0	Adelaide, King, Richmond	Coffee Shop	Café	Bar	Steakhouse	Sushi Restaurant	Thai Restaurant	Salad Place	Restaurant	Burger Joint	Bakery
1	Agincourt	Lounge	Breakfast Spot	Latin American Restaurant	Skating Rink	Electronics Store	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Eastern European Restaurant
2	Agincourt North, L'Amoreaux East, Milliken, St...	Park	Playground	Coffee Shop	Eastern European Restaurant	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant
3	Albion Gardens, Beaumond Heights, Humbergate, ...	Grocery Store	Fast Food Restaurant	Sandwich Place	Pizza Place	Beer Store	Pharmacy	Fried Chicken Joint	Event Space	Ethiopian Restaurant	Empanada Restaurant
4	Alderwood, Long Branch	Pizza Place	Dance Studio	Gym	Skating Rink	Pharmacy	Coffee Shop	Sandwich Place	Pub	Athletics & Sports	Doner Restaurant



CLUSTER 1

- The biggest cluster with 99 neighbourhoods
- Centroid - Adelaide, King, Richmond Neighbourhood
- Common features - coffee shops, bars and Cafés
- Most of the venues in this cluster are dining experiences.

	Neighbourhood	Cluster Labels	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
1	Highland Creek, Rouge Hill, Port Union	0	History Museum	Bar	Yoga Studio	Electronics Store	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant
2	Guildwood, Morningside, West Hill	0	Medical Center	Breakfast Spot	Pizza Place	Rental Car Location	Electronics Store	Intersection	Mexican Restaurant	Drugstore	Dog Run	Doner Restaurant
3	Woburn	0	Coffee Shop	Korean Restaurant	Yoga Studio	Electronics Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Eastern European Restaurant
4	Cedarbrae	0	Bank	Hakka Restaurant	Caribbean Restaurant	Gas Station	Athletics & Sports	Thai Restaurant	Fried Chicken Joint	Bakery	Donut Shop	Discount Store
5	Scarborough Village	0	Playground	Convenience Store	Yoga Studio	Electronics Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Eastern European Restaurant

ADELAIDE, KING, RICHMOND:

	venue	freq
0	Coffee Shop	0.08
1	Café	0.05
2	Bar	0.04
3	Steakhouse	0.04
4	Burger Joint	0.03
5	Bakery	0.03
6	Sushi Restaurant	0.03
7	Thai Restaurant	0.03
8	Salad Place	0.03
9	Restaurant	0.03
10	Asian Restaurant	0.03
11	Gastropub	0.02
12	Concert Hall	0.02
13	Pizza Place	0.02
14	Cosmetics Shop	0.02
15	Vegetarian / Vegan Restaurant	0.02
16	American Restaurant	0.02
17	Hotel	0.02
18	Gym	0.02
19	Plaza	0.01

CLUSTER 2

- This cluster is an outlier
- Centroid - Rouge, Malvern
- Common venues - fast food restaurant, Yoga studio and Eastern European.

Neighbourhood	Cluster Labels	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
0 Rouge, Malvern	1	Fast Food Restaurant	Yoga Studio	Eastern European Restaurant	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Electronics Store

ROUGE, MALVERN:

	venue	freq
0	Fast Food Restaurant	1.0
1	Accessories Store	0.0
2	Metro Station	0.0
3	Molecular Gastronomy Restaurant	0.0
4	Modern European Restaurant	0.0
5	Mobile Phone Shop	0.0
6	Miscellaneous Shop	0.0
7	Middle Eastern Restaurant	0.0
8	Mexican Restaurant	0.0
9	Men's Store	0.0
10	Luggage Store	0.0

CLUSTER 3

- 2nd biggest cluster with 9 localities
- Centroid – Rosedale
- Common venues - parks, trails and playgrounds.

	Neighbourhood	Cluster Labels	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
14	Agincourt North, L'Amoreaux East, Milliken, St...	2	Park	Playground	Coffee Shop	Eastern European Restaurant	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant
19	Silver Hills, York Mills	2	Park	Cafeteria	Yoga Studio	Eastern European Restaurant	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Electronics Store
23	Parkwoods	2	Park	Food & Drink Shop	Yoga Studio	Electronics Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Eastern European Restaurant
28	CFB Toronto, Downsview East	2	Park	Airport	Yoga Studio	Electronics Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Eastern European Restaurant
38	East Toronto	2	Park	Coffee Shop	Convenience Store	Yoga Studio	Electronics Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant
48	Rosedale	2	Park	Trail	Playground	Diner	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant
72	Caledonia-Fairbanks	2	Park	Fast Food Restaurant	Women's Store	Market	Eastern European Restaurant	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant
88	The Kingsway, Montgomery Road, Old Mill North	2	Park	Smoke Shop	River	Comfort Food Restaurant	Comic Shop	Falafel Restaurant	Event Space	Colombian Restaurant	Ethiopian Restaurant	Empanada Restaurant
95	Weston	2	Park	Yoga Studio	Electronics Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Eastern European Restaurant	Empanada Restaurant

ROSEDALE :

	venue	freq
0	Park	0.50
1	Trail	0.25
2	Playground	0.25
3	Accessories Store	0.00
4	Molecular Gastronomy Restaurant	0.00
5	Modern European Restaurant	0.00
6	Mobile Phone Shop	0.00
7	Miscellaneous Shop	0.00
8	Middle Eastern Restaurant	0.00
9	Mexican Restaurant	0.00
10	Men's Store	0.00

CLUSTER 4

- Cluster is another outlier
- Centroid – Roselawn
- Common venues - Garden Cemetery

	Neighbourhood	Cluster Labels	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
61	Roselawn	3	Garden	Yoga Studio	Eastern European Restaurant	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Electronics Store

```
ROSELAWN:
      venue  freq
0      Garden  1.0
1  Accessories Store  0.0
2    Metro Station  0.0
3  Modern European Restaurant  0.0
4    Mobile Phone Shop  0.0
5  Miscellaneous Shop  0.0
6  Middle Eastern Restaurant  0.0
7    Mexican Restaurant  0.0
8      Men's Store  0.0
9  Monument / Landmark  0.0
10 Mediterranean Restaurant  0.0
```

CLUSTER 5

- This is the last cluster which is also an outlier
- Centroid – Cloverdale, Islington, Grove, Deane Park
- Common venue - Golf course.

	Neighbourhood	Cluster Labels	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
91	Cloverdale, Islington, Martin Grove, Princess ...	4	Golf Course	Yoga Studio	Eastern European Restaurant	Discount Store	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant	Electronics Store

```
CLOVERDALE, ISLINGTON, MARTIN GROVE, PRINCESS GARDENS, WEST DEANE PARK:
venue freq
0      Golf Course  1.0
1      Accessories Store  0.0
2      Motel        0.0
3      Molecular Gastronomy Restaurant  0.0
4      Modern European Restaurant  0.0
5      Mobile Phone Shop  0.0
6      Miscellaneous Shop  0.0
7      Middle Eastern Restaurant  0.0
8      Mexican Restaurant  0.0
9      Metro Station  0.0
10     Men's Store  0.0
```

CONCLUSION

- The project has highlighted 5 clusters with unique attributes
- If a person is looking for a neighborhood with diverse and easily accessible dining options then Cluster 1 would be recommended.
- For a family or outdoors loving person, the neighborhoods in Cluster 3 are more suitable, due to the common venues such as parks, playgrounds and trails which are ideal for kids and people who love outdoor activities.
- For people who are looking for something specific like golf, then cluster 5 is the best fit for their lifestyle.
- Whereas, Cluster 4 with the single neighbourhood namely Roselawn is a cemetery, which tends to be relatively cheaper to live nearby
- This study can be further analyzed and refined by including other factors such as crime rate, cost and school districts