

SELECTING LOCATIONS FOR PIZZA OUTLETS

ABHIJIT DAS – APR2020



EXTRACT CANADA NEIGHBORHOOD DATA

- Scrape Wikipedia webpage to extract Toronto postal code (this was done in week 3 and 4)
- Postal codes saved in csv file
- Load postal codes from csv file. Data contains Postal Code, Borough, Name of Neighborhood,

2]:

	Postcode	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront

CANADA GEO DATA

- Load Canada geo location data from csv file (http://cocl.us/Geospatial_data)
- The location data contains postal code and their latitudes and longitudes

]:

	Postcode	Latitude	Longitude
0	M1B	43.806888	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

CANADA NEIGHBORHOOD AND GEO DATA COMBINED

- The two dataframes for Canada are combined to consolidate the neighborhood data

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge,Malvern	43.806886	-79.194353
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.180497
2	M1E	Scarborough	Guildwood,Morningside,West Hill	43.783573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

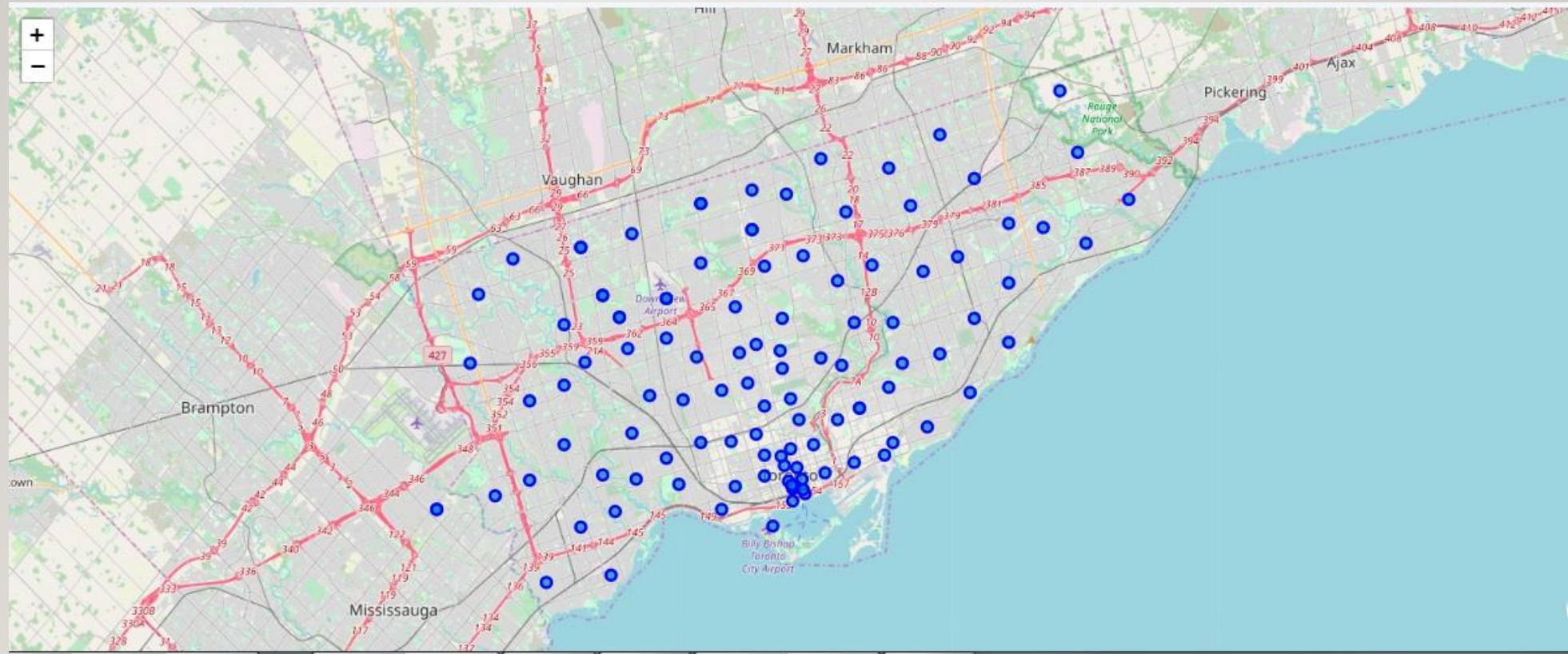
TORONTO NEIGHBORHOODS FROM CANADA DATA

- Separate the Toronto neighborhoods

	Postcode	Borough	Neighbourhood	Latitude	Longitude
43	M4E	East Toronto	The Beaches	43.676357	-79.293031
46	M4J	East Toronto	The Danforth East	43.685347	-79.338106
48	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
49	M4L	East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572
50	M4M	East Toronto	Studio District	43.659526	-79.340923

(40, 5)

TORONTO NEIGHBORHOODS



TORONTO VENUE DATA FROM FOURSQUARE

- Using Foursquare API, location data is extracted for Toronto neighborhoods. Location data for single locations is as shown below.
- Each venue has a “venue category”

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Domino's Pizza	43.679058	-79.297382	Pizza Place
1	The Beaches	43.676357	-79.293031	Fearless Meats	43.680337	-79.290289	Burger Joint
2	The Beaches	43.676357	-79.293031	Seaspray Restaurant	43.678888	-79.298167	Asian Restaurant
3	The Danforth East	43.685347	-79.338106	105 Eaton Bakery	43.684719	-79.340729	Bakery
4	The Danforth East	43.685347	-79.338106	Command Centre	43.687667	-79.339490	Greek Restaurant

VENUE DATA ONE-HOT ENCODED BY CATEGORY TYPES

- This data contains one row for every venue, data needs to be cleaned and combined
- The “category column” associated with the venue has a “one” all other columns have a zero. Hence there are multiple rows for venues in a neighborhood

	Longitude	American Restaurant	Arepa Restaurant	Asian Restaurant	BBQ Joint	Bagel Shop	Bakery	Belgian Restaurant	Bistro	Brazilian Restaurant	...	Sushi Restaurant	Taiwanese Restaurant	Tapas Restaurant	Thai Restaurant	Theme Restaurant	Vegetarian / Vegan Restaurant	Vietnamese Restaurant	Wings Joint	Neighborhood	Latitude
0	-79.293031	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	The Beaches	43.676357
1	-79.293031	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	The Beaches	43.676357
2	-79.293031	0	0	1	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	The Beaches	43.676357
3	-79.338106	0	0	0	0	0	1	0	0	0	...	0	0	0	0	0	0	0	0	The Danforth East	43.685347
4	-79.338106	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	The Danforth East	43.685347

5 rows x 73 columns

NEIGHBORHOOD VENUE DATAFRAME

- The one-hot encoded venue data is combined to create one row for every neighborhood. The columns contain the mean of the number of times a venue category appears in that neighborhood, this provides the weight of each venue category in a neighborhood

	Neighborhood	Longitude	American Restaurant	Arepa Restaurant	Asian Restaurant	BBQ Joint	Bagel Shop	Bakery	Belgian Restaurant	Bistro	...	Steakhouse	Sushi Restaurant	Taiwanese Restaurant	Tapas Restaurant	Thai Restaurant	Theme Restaurant	Vegetarian / Vegan Restaurant	Vietnamese Restaurant	Wings Joint	Lat
0	Berczy Park	-79.373306	0.000000	0.00	0.000000	0.040000	0.040000	0.080000	0.04	0.04	...	0.040000	0.040000	0.00	0.0	0.040000	0.00	0.040000	0.000000	0.00	43.64
1	Brockton, Parkdale Village, Exhibition Place	-79.428191	0.000000	0.00	0.000000	0.000000	0.000000	0.133333	0.00	0.00	...	0.000000	0.000000	0.00	0.0	0.000000	0.00	0.000000	0.066667	0.00	43.63
2	Business Reply Mail Processing Centre 969 Eastern	-79.321558	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.00	...	0.000000	0.000000	0.00	0.0	0.000000	0.00	0.000000	0.000000	0.00	43.65
3	CN Tower, King and Spadina, Railway Lands, Harbourfront	-79.394420	0.500000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.00	...	0.000000	0.000000	0.00	0.5	0.000000	0.00	0.000000	0.000000	0.00	43.62
4	Central Bay Street	-79.387383	0.040000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.00	...	0.040000	0.080000	0.00	0.0	0.040000	0.00	0.040000	0.000000	0.00	43.65
5	Christie	-79.422564	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.00	...	0.000000	0.000000	0.00	0.0	0.000000	0.00	0.000000	0.000000	0.00	43.65
6	Church and Wellesley	-79.383160	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.00	...	0.040000	0.080000	0.00	0.0	0.040000	0.04	0.000000	0.040000	0.04	43.65

SUMMARIZE VENUE CATEGORIES INTO 8 CATEGORIES

- There are venue categories which are similar, like “Noodle house” and “Chinese”, these essentially belong to one type “Asian”. The weights of “Noodle house” and “Chinese” is combined into “Asian”.
- The venue categories are grouped into 8 categories to create a new dataframe, the weights of the venue categories are combined into 8 summarized categories

]:

	Neighborhood	Latitude	Longitude	PizzaOutlet	North American	Comfort Food	Asian Cuisine	South American Cuisine	Middle Eastern Cuisine	European Cuisine	Healthy Choices
0	Berczy Park	43.644771	-79.373306	0.040000	0.320000	0.120000	0.200000	0.000000	0.000000	0.280000	0.040000
1	Brockton, Parkdale Village, Exhibition Place	43.636847	-79.428191	0.066667	0.466667	0.000000	0.133333	0.066667	0.000000	0.133333	0.133333
2	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	0.200000	0.200000	0.400000	0.000000	0.200000	0.000000	0.000000	0.000000
3	CN Tower, King and Spadina, Railway Lands, Harbou...	43.628947	-79.394420	0.000000	0.500000	0.000000	0.000000	0.000000	0.000000	0.500000	0.000000
4	Central Bay Street	43.657952	-79.387383	0.160000	0.240000	0.040000	0.360000	0.000000	0.080000	0.040000	0.080000
5	Christie	43.669542	-79.422564	0.142857	0.857143	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	Church and Wellesley	43.665860	-79.383160	0.080000	0.280000	0.080000	0.320000	0.080000	0.040000	0.120000	0.000000
7	Commerce Court, Victoria Hotel	43.648199	-79.379817	0.000000	0.520000	0.120000	0.160000	0.000000	0.000000	0.080000	0.120000
8	Davisville	43.704324	-79.388790	0.160000	0.320000	0.000000	0.320000	0.000000	0.000000	0.080000	0.120000
9	Davisville North	43.712751	-79.390197	0.250000	0.000000	0.000000	0.250000	0.000000	0.000000	0.250000	0.250000
10	Dufferin, Dovercourt Village	43.669005	-79.442259	0.250000	0.375000	0.000000	0.000000	0.125000	0.125000	0.125000	0.000000
11	Enclave of M5E	43.646435	-79.374846	0.000000	0.360000	0.200000	0.160000	0.000000	0.000000	0.200000	0.080000
12	First Canadian Place, Underground city	43.648429	-79.382280	0.040000	0.440000	0.120000	0.160000	0.000000	0.000000	0.120000	0.120000

CREATE TRAINING DATA

- The locations having the “PizzaOutlet” as popular category, I have taken that as rating higher than 0.1 are marked as recommended. The others are marked not-recommended.
- A subset of this is selected as the training data, following is the training data
- A classification model algorithm “K-Nearest Neighbor” is used to build the model
- The model is then used to predict which locations would be favorable for pizza locations

]:

	Neighborhood	Latitude	Longitude	PizzaOutlet	North American	Comfort Food	Asian Cuisine	South American Cuisine	Middle Eastern Cuisine	European Cuisine	Healthy Choices	Recommended
0	Berczy Park	43.644771	-79.373306	0.040000	0.320000	0.12	0.200000	0.000000	0.00	0.280000	0.040000	0
1	Brockton, Parkdale Village, Exhibition Place	43.636847	-79.428191	0.066667	0.466667	0.00	0.133333	0.066667	0.00	0.133333	0.133333	0
2	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	0.200000	0.200000	0.40	0.000000	0.200000	0.00	0.000000	0.000000	1
3	CN Tower, King and Spadina, Railway Lands, Harbourfront	43.628947	-79.394420	0.000000	0.500000	0.00	0.000000	0.000000	0.00	0.500000	0.000000	0
4	Central Bay Street	43.657952	-79.387383	0.160000	0.240000	0.04	0.360000	0.000000	0.08	0.040000	0.080000	1
5	Christie	43.669542	-79.422564	0.142857	0.857143	0.00	0.000000	0.000000	0.00	0.000000	0.000000	1
6	Church and Wellesley	43.665960	-79.383160	0.080000	0.280000	0.08	0.320000	0.080000	0.04	0.120000	0.000000	0
7	Commerce Court, Victoria Hotel	43.648199	-79.379817	0.000000	0.520000	0.12	0.160000	0.000000	0.00	0.080000	0.120000	0

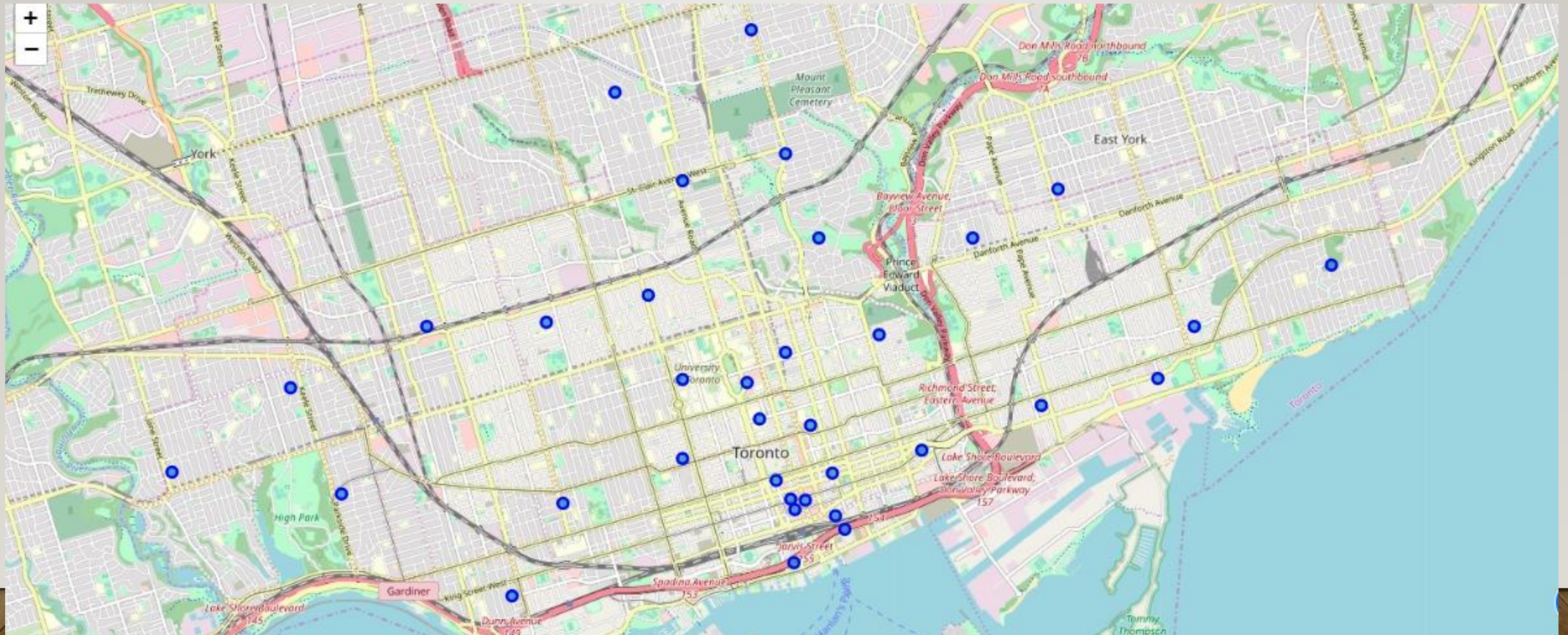
PREDICTED RESULTS

- The model is used on the entire dataset to predict which locations are favorable for pizza locations. The prediction is the “Recommendation” field as “0” for “Not recommended” and “1” for recommended.
- The predicted “recommendation” is merged into the original dataset for plotting a map

	Neighborhood	Latitude	Longitude	PizzaOutlet	North American	Comfort Food	Asian Cuisine	South American Cuisine	Middle Eastern Cuisine	European Cuisine	Healthy Choices	Recommended	Borough
0	Berczy Park	43.644771	-79.373306	0.040000	0.320000	0.12	0.200000	0.000000	0.00	0.280000	0.040000	0	
1	Brookton, Parkdale Village, Exhibition Place	43.636847	-79.428191	0.066667	0.466667	0.00	0.133333	0.066667	0.00	0.133333	0.133333	0	
2	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	0.200000	0.200000	0.40	0.000000	0.200000	0.00	0.000000	0.000000	1	East Toronto
3	CN Tower, King and Spadina, Railway Lands, Harbourfront	43.628947	-79.394420	0.000000	0.500000	0.00	0.000000	0.000000	0.00	0.500000	0.000000	0	Downtown Toronto
4	Central Bay Street	43.657952	-79.387383	0.160000	0.240000	0.04	0.360000	0.000000	0.08	0.040000	0.080000	1	

PLOT OF TORONTO LOCATIONS RECOMMENDED FOR PIZZA OUTLETS

- The map below shows the Toronto neighborhoods that are recommended for pizza outlets



THANK YOU

Thanks for reviewing my Capstone Project – Abhijit Das