The Battle of Neighbourhoods

APPLIED DATA SCIENCE CAPSTONE PROJECT

BALTEJ TOOR

The Business Need

- Despite the opportunities and resources afforded to new businesses, there are difficulties that face entrepreneurs looking to grow in Toronto's neighbourhoods
- It is challenging to allocate resources to the gathering of information on competitors and BIA initiatives
- Businesses need the ability to evaluate locations on the basis of **competition density** and enhancement via **Business Improvement Areas (BIAs)**

Data Requirements

- **Foursquare API**
  - Foursquare venue exploration is used to retrieve nearby venue categories/types as an indication of business competition
  - API call utilizes [latitude, longitude] coordinates with a predefined RADIUS
  - Merged with geospatial coordinate data of neighbourhoods via http://cocl.us/Geospatial_data_ - geospatial_coordinates.csv, web scraping of https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

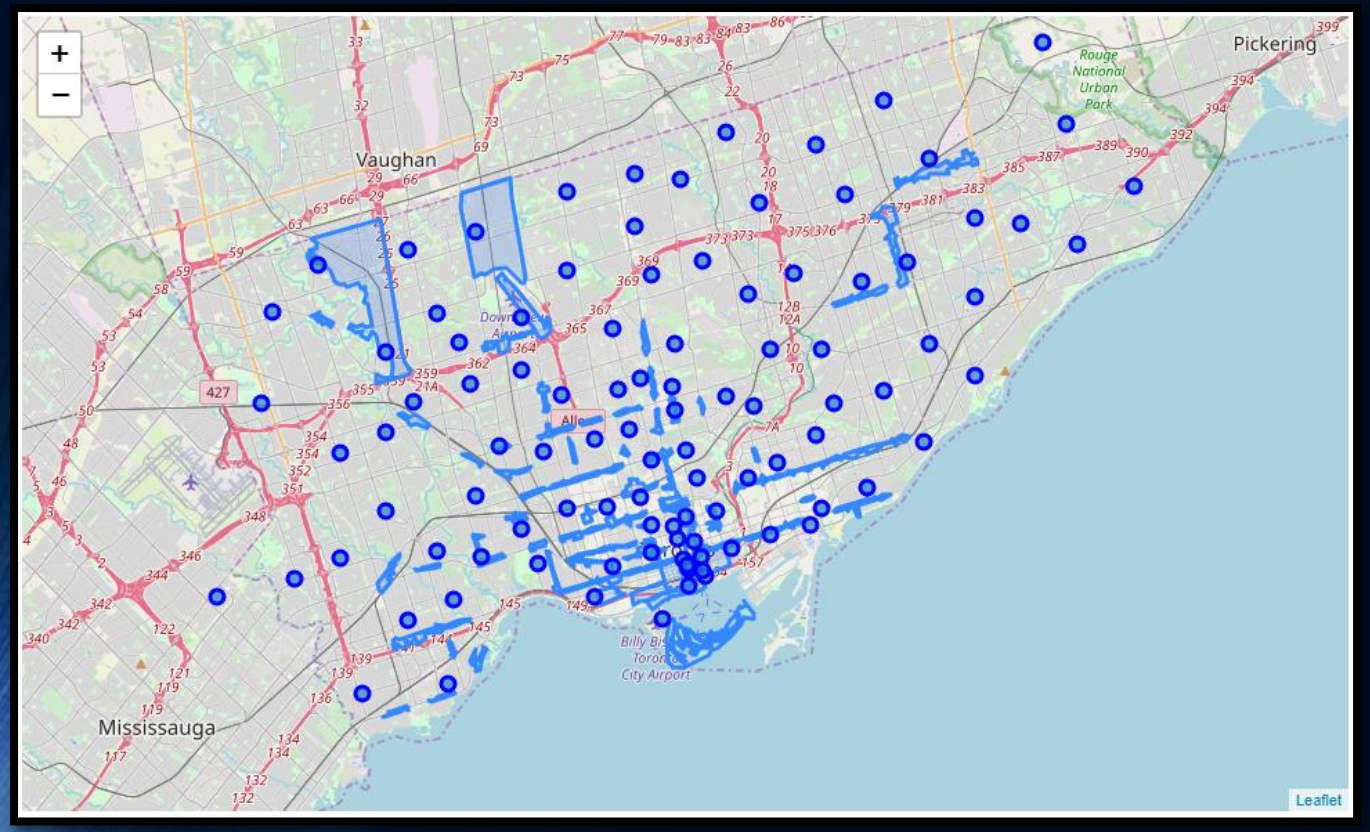
- **Toronto BIA Data**
  - Incorporates BIA location data which provides [latitude, longitude] geometries as well as centralized coordinates (used for BIA availability determination)

Determining Neighbourhood Venue Frequency

- The top ***5*** frequent venue categories per neighbourhood were selected
- ***n = 5*** corresponds to the number of venue types to incorporate into competition density calculations
- Ensures that the majority of relevant venue categories are used for analysis

	Neighbourhood	1st Most Common Venue	# of 1st Venue	2nd Most Common Venue	# of 2nd Venue	3rd Most Common Venue	# of 3rd Venue	4th Most Common Venue	# of 4th Venue	5th Most Common Venue	# of 5th Venue
0	Adelaide, King, Richmond	Coffee Shop	8	Café	6	Steakhouse	4	Thai Restaurant	3	Cosmetics Shop	3
1	Agincourt	Supermarket	1	Skating Rink	1	Mediterranean Restaurant	1	Seafood Restaurant	1	Shanghai Restaurant	1
2	Agincourt North, L'Amoreaux East, Milliken, St...	Chinese Restaurant	3	Fast Food Restaurant	2	Pizza Place	2	Coffee Shop	1	Malay Restaurant	1
3	Albion Gardens, Beaumond Heights, Humbergate, ...	Grocery Store	3	Pizza Place	2	Hardware Store	1	Auto Garage	1	Coffee Shop	1
4	Alderwood, Long Branch	Convenience Store	2	Pizza Place	2	Pool	1	Donut Shop	1	Sandwich Place	1
5	Bathurst Manor, Downsview North, Wilson Heights	Coffee Shop	2	Pizza Place	2	Gas Station	1	Community Center	1	Sandwich Place	1
6	Bayview Village	Japanese Restaurant	2	Bank	2	Café	1	Shopping Mall	1	Grocery Store	1
7	Bedford Park, Lawrence Manor East	Coffee Shop	3	Italian Restaurant	3	Sandwich Place	2	Fast Food Restaurant	2	Breakfast Spot	1
8	Berczy Park	Coffee Shop	9	Café	5	Beer Bar	4	Restaurant	4	Hotel	4
9	Birch Cliff, Cliffside West	Construction & Landscaping	1	Thai Restaurant	1	College Stadium	1	Diner	1	Café	1

Toronto Neighbourhood & BIA Location Visualization



- Provides a sense of the orientation of the neighbourhoods relative to BIAs
  - Differences in the relationship between *BIA availability* and *competition density* from region to region

BIA Availability Evaluation

- Requires calculating the distance of each BIA to each neighbourhood
- Utilizes a geodesic distance function – shortest distance between two coordinate points on a model of the Earth
- *BIA availability* involves checking whether each BIA is within the RADIUS of a neighbourhood, updating a count 'Available BIAs'

Competition Density Evaluation

- Leverages the neighbourhood venue frequency data to generate additional features for analysis
- *Competition density* is the sum of related business venue categories per neighbourhood
- Sample target category was a *coffee-related* business, 'Coffee Shops' and 'Cafés' categories from Foursquare were incorporated

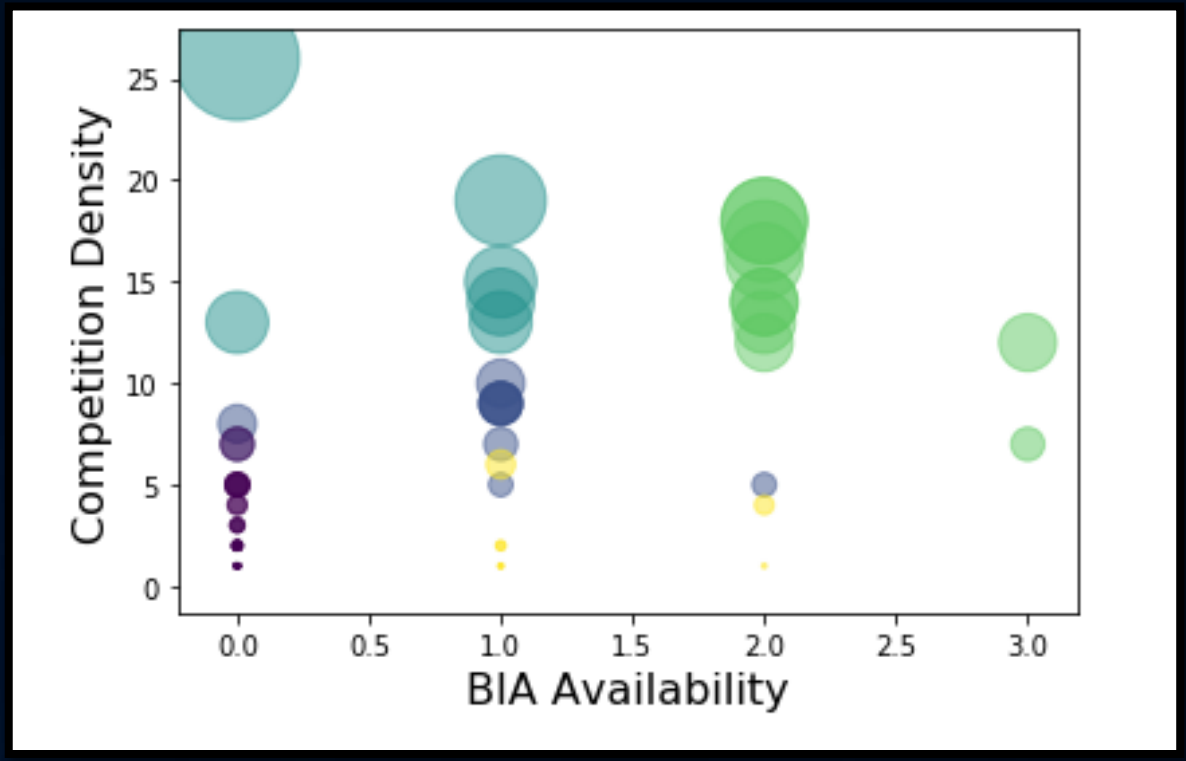
	Neighbourhood	Available BIAs	Coffee Shops	Cafés
0	Adelaide, King, Richmond	2	8	6
1	Agincourt	1	0	0
2	Agincourt North, L'Amoreaux East, Milliken, St...	0	1	0
3	Albion Gardens, Beaumond Heights, Humbergate, ...	0	1	0
4	Alderwood, Long Branch	0	0	0
5	Bathurst Manor, Downsview North, Wilson Heights	0	2	0
6	Bayview Village	0	0	1
7	Bedford Park, Lawrence Manor East	0	3	0
8	Berczy Park	2	9	5
9	Birch Cliff, Cliffside West	0	0	1

Machine Learning Analysis – *K-Means Clustering*

- *K-means* clustering was performed to discover insights from the relevant data in a "*neighbourhood segmentation"* problem
- ***k = 5*** was chosen based on observational assessment of the degree of difference in the resulting clusters (avoiding 1-member clusters)
- To avoid any issues with local minima, the algorithm was run multiple times with different centroids for the best result
- Cluster summarizing factors included a third measure to be used as a cluster comparison mechanism:

BIA/Comp ratio = 'Available BIAs'/'Competition Density'

- A cluster scatter plot visualization shows a simplistic cluster orientation of the relationship between *BIA Availability* and *Competition Density* (marker radius proportional to competition density)



Insights from Cluster Characteristics

- Cluster *mean BIA/Comp ratio* was used as a comparison metric for the 5 resulting clusters – *higher* mean BIA/Comp ratio → generally *better* the neighbourhood for a new coffee-related business (based on the metrics tracked)

CLUSTER 0 (results truncated for size)

```
Mean BIA/Comp Ratio for Cluster 0 is 0.0
Std Deviation of BIA/Comp Ratio for Cluster 0 is 0.0
```

	Neighbourhood	Available BIAs	Competition Density	BIA/Comp Ratio
2	Agincourt North, L'Amoreaux East, Milliken, St...	0	1	0.0
3	Albion Gardens, Beaumond Heights, Humbergate, ...	0	1	0.0
4	Alderwood, Long Branch	0	0	0.0
5	Bathurst Manor, Downsview North, Wilson Heights	0	2	0.0
6	Bayview Village	0	1	0.0
7	Bedford Park, Lawrence Manor East	0	3	0.0
9	Birch Cliff, Cliffside West	0	1	0.0
10	Bloordale Gardens, Eringate, Markland Wood, Ol...	0	1	0.0
12	Business Reply Mail Processing Centre 969 Eastern	0	0	0.0
14	CN Tower, Bathurst Quay, Island airport, Harbo...	0	0	0.0

CLUSTER 1

```
Mean BIA/Comp Ratio for Cluster 1 is 0.12873015873015875
Std Deviation of BIA/Comp Ratio for Cluster 1 is 0.11258396454191189
```

	Neighbourhood	Available BIAs	Competition Density	BIA/Comp Ratio
11	Brockton, Exhibition Place, Parkdale Village	1	9	0.111111
15	Cabbagetown, St. James Town	1	9	0.111111
21	Christie	1	5	0.200000
28	Davisville	1	9	0.111111
30	Deer Park, Forest Hill SE, Rathnelly, South Hi...	1	9	0.111111
41	East Toronto	0	8	0.000000
52	High Park, The Junction South	1	7	0.142857
65	Little Portugal, Trinity	2	5	0.400000
69	Northwest	0	7	0.000000
78	Ryerson, Garden District	1	10	0.100000

CLUSTER 2

```
Mean BIA/Comp Ratio for Cluster 2 is 0.04460831566094723
Std Deviation of BIA/Comp Ratio for Cluster 2 is 0.03548046552664549
```

	Neighbourhood	Available BIAs	Competition Density	BIA/Comp Ratio
19	Central Bay Street	1	13	0.076923
27	Commerce Court, Victoria Hotel	1	19	0.052632
50	Harbourfront	1	15	0.066667
51	Harbourfront East, Toronto Islands, Union Station	1	14	0.071429
74	Rosedale	0	26	0.000000
79	Scarborough Village	0	13	0.000000

CLUSTER 3

Mean BIA/Comp Ratio for Cluster 3 is 0.17496678158442866
Std Deviation of BIA/Comp Ratio for Cluster 3 is 0.09801340874469185

	Neighbourhood	Available BIAs	Competition Density	BIA/Comp Ratio
0	Adelaide, King, Richmond	2	14	0.142857
8	Berczy Park	2	14	0.142857
20	Chinatown, Grange Park, Kensington Market	3	7	0.428571
22	Church and Wellesley	3	12	0.250000
32	Design Exchange, Toronto Dominion Centre	2	18	0.111111
44	First Canadian Place, Underground city	2	17	0.117647
49	Harbord, University of Toronto	2	13	0.153846
82	Stn A PO Boxes 25 The Esplanade	2	16	0.125000
83	Studio District	2	18	0.111111
84	The Annex, North Midtown, Yorkville	2	12	0.166667

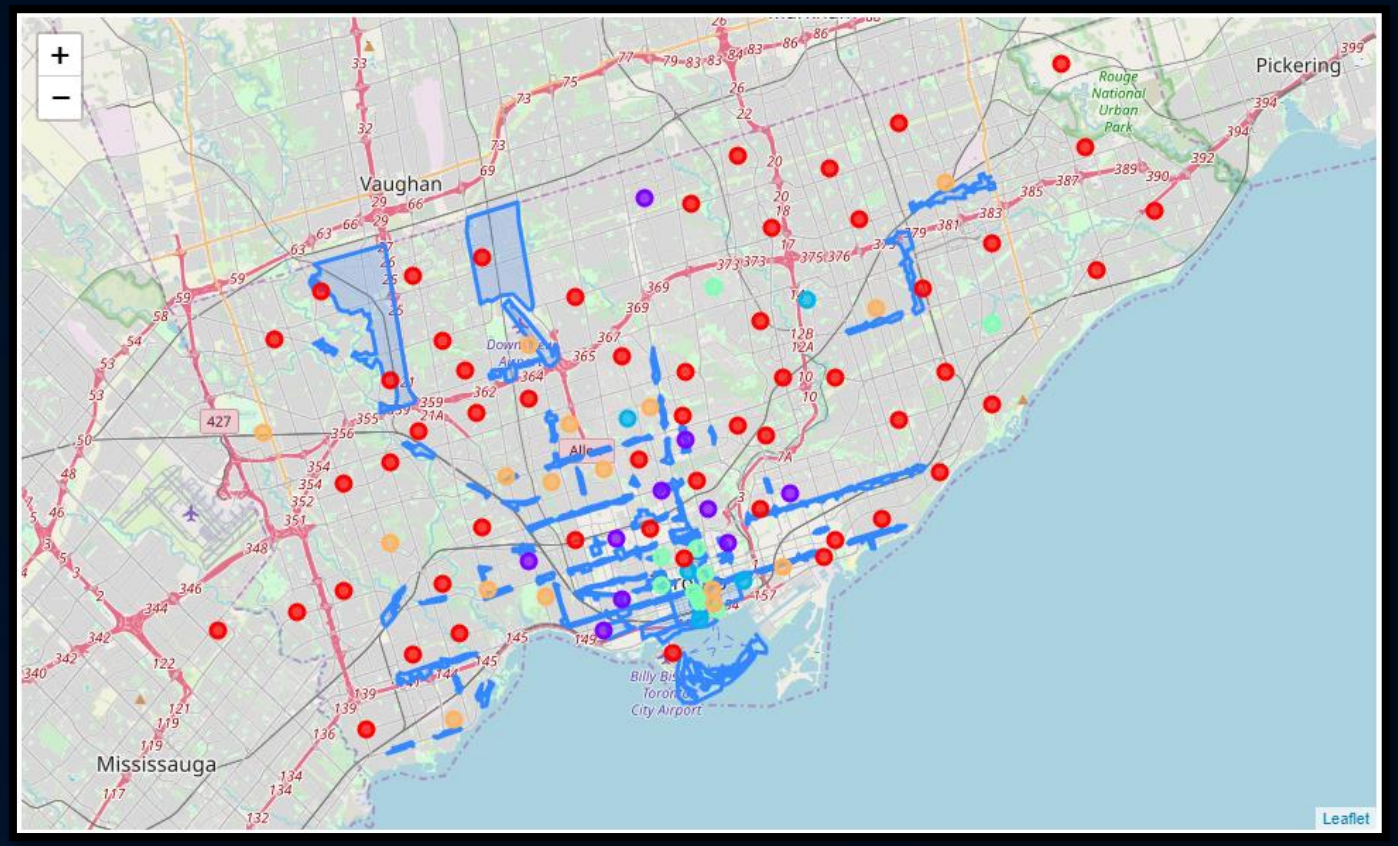
CLUSTER 4

Mean BIA/Comp Ratio for Cluster 4 is 1.0370370370370372
Std Deviation of BIA/Comp Ratio for Cluster 4 is 0.5098877088847604

	Neighbourhood	Available BIAs	Competition Density	BIA/Comp Ratio
1	Agincourt	1	0	1.000000
13	CFB Toronto, Downsview East	1	2	0.500000
16	Caledonia-Fairbanks	1	0	1.000000
31	Del Ray, Keelesdale, Mount Dennis, Silverthorn	1	0	1.000000
47	Glencairn	1	1	1.000000
55	Humber Bay Shores, Mimico South, New Toronto	1	0	1.000000
58	Humewood-Cedarvale	1	0	1.000000
66	Maryvale, Wexford	1	0	1.000000
68	North Toronto West	1	1	1.000000
70	Northwood Park, York University	1	0	1.000000
71	Parkdale, Roncesvalles	2	1	2.000000
73	Queen's Park	1	0	1.000000
77	Runnymede, Swansea	2	0	2.000000
81	St. James Town	2	0	2.000000
85	The Beaches	1	6	0.166667
86	The Beaches West, India Bazaar	1	2	0.500000
87	The Danforth West, Riverdale	2	4	0.500000
93	Weston	1	1	1.000000

- The individual cluster breakdown shows that ***cluster 4*** has the highest mean BIA/Comp ratio with the cluster rankings as follows:
    1. **Cluster 4** – Mean BIA/Comp ratio – 1.0370370370370372
    2. **Cluster 3** – Mean BIA/Comp ratio – 0.17496678158442866
    3. **Cluster 1** – Mean BIA/Comp ratio – 0.12873015873015875
    4. **Cluster 2** – Mean BIA/Comp ratio – 0.04460831566094723
    5. **Cluster 0** – Mean BIA/Comp ratio – 0.0
- Cluster 4 represents neighbourhoods with access to BIAs with minimal to no competition, a suitable balance of BIA availability and competition density

Toronto Neighbourhood Clustering Visualization



Cluster 0, Cluster 1, Cluster 2, Cluster 3, Cluster 4

- Further interpretation shows circumstantial viability for cluster 0 and cluster 3 neighbourhoods as well (each cluster fills particular business goals)

CLUSTER 4

- Provides best relative all-around environment to start a new coffee-related business
- Occupies "sweet spot" relative to denser parts of the city (increased competition density areas)

CLUSTER 0

- If BIA availability is less significant, these locations are more desirable due to minimal competition
- Similar competition density to cluster 4
- Neighbourhoods more widely dispersed, outskirts of the city

CLUSTER 3

- More suitable in cases where BIA availability is more important (regardless of competition density)
- Greater variety of BIA resources per location
- Locations predominantly in the city's center

Conclusion

- This project provides entrepreneurs and new business owners the ability to evaluate Toronto neighbourhoods on the basis of competition and BIA initiatives available
- Clustering analysis of neighbourhoods produced distinct partitions that can be compared in regards to competition density, BIA availability and geographical orientation (e.g. cluster 4 to cluster 0 and cluster 3 similarities and differences)
- Neighbourhood segmentation implements a resource that can be adjusted for individual business needs depending on emphasis placed on maximizing BIA opportunities vs minimizing the area's competition