IBM Data Science Professional Certificate

Opening a INDIAN Restaurant in Toronto

G SAI VARUN KUMAR

Business Problem

Toronto

- 3 Million inhabitants
- 16th place worldwide with 272 restaurants per 100.000 inhabitants

Indian community

- More than 640,000 people just in Toronto
- Indian cuisine as one of the richest in the world

If an investor is looking to open a new Indian restaurant where would you recommend it?

Data

List of neighbourhoods in Toronto

- Boroughs and postal codes (Wikipedia)
- Geo coordinates

Indian venues data

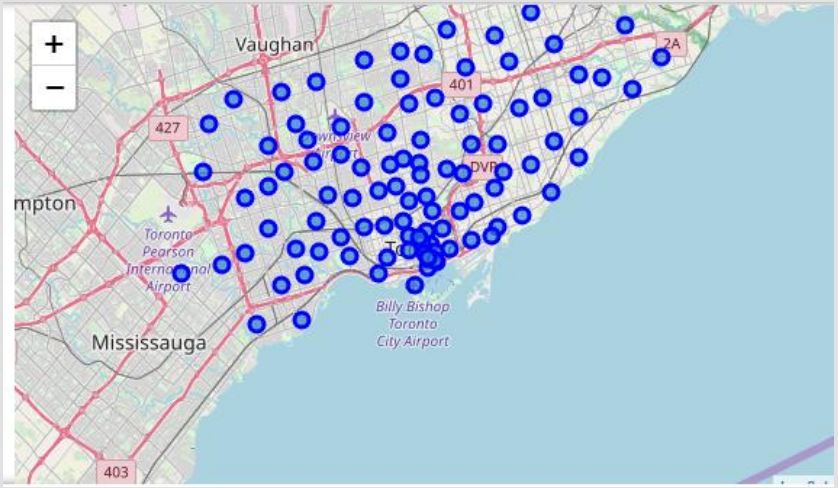
- Using Foursquare API
  - Category Id for Indian Restaurant

Methodology

Toronto Neighbourhoods Data Exploratory Analysis

- Using Wikipedia data and Coursera CSV

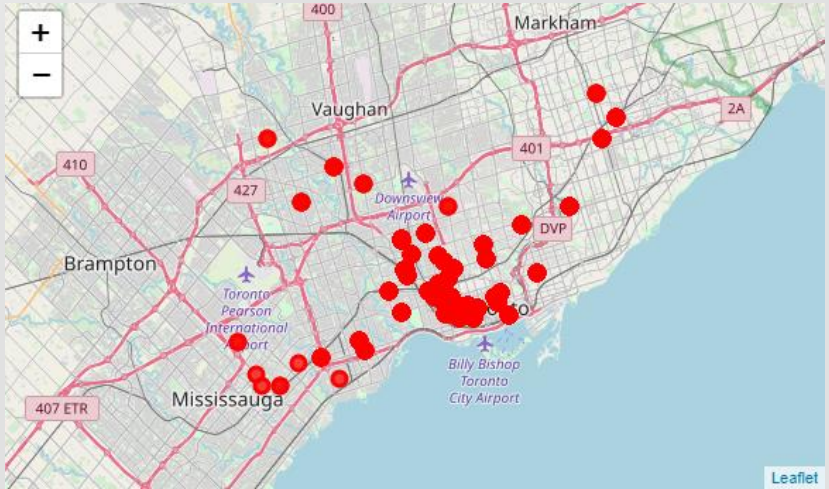
	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H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



Methodology

Indian Restaurants Data Exploratory Analysis

- Using Foursquare we retrieved all venues per neighbourhood



Methodology

Data wrangling

- Create frequency based data frame per neighbourhood

	Neighborhood	American Restaurant	Bar	Breakfast Spot	Café	Chinese Restaurant	Coffee Shop	Deli / Bodega	Diner	Event Space
0	Adelaide,King,Richmond	0.0	0.02381	0.011905	0.02381	0.000000	0.011905	0.011905	0.02381	0.011905
1	Agincourt	0.0	0.00000	0.000000	0.00000	0.083333	0.000000	0.000000	0.00000	0.000000
2	Agincourt North,L'Amoreaux East,Milliken,Steel...	0.0	0.00000	0.000000	0.00000	0.058824	0.000000	0.000000	0.00000	0.000000
3	Albion Gardens,Beaumond Heights,Humbergate,Jam...	0.0	0.00000	0.000000	0.00000	0.000000	0.000000	0.000000	0.00000	0.000000
4	Alderwood,Long Branch	0.0	0.00000	0.000000	0.00000	0.000000	0.000000	0.000000	0.00000	0.000000

Methodology

Data wrangling

- Create top venues data frame

	Neighborhood	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
0	Adelaide,King,Richmond	Indian Restaurant	Vegetarian / Vegan Restaurant	Café	Diner	Italian Restaurant	Bar	Pizza Place
1	Agincourt	Indian Restaurant	Chinese Restaurant	Grocery Store	Vegetarian / Vegan Restaurant	Hotel	Bar	Breakfast Spot
2	Agincourt North,L'Amoreaux East,Milliken,Steel...	Indian Restaurant	Indian Chinese Restaurant	Chinese Restaurant	Hotel	Bar	Breakfast Spot	Café
3	Albion Gardens,Beaumond Heights,Humbergate,Jam...	Indian Restaurant	Vegetarian / Vegan Restaurant	Hotel	Bar	Breakfast Spot	Café	Chinese Restaurant
4	Alderwood,Long Branch	Indian Restaurant	Vegetarian / Vegan Restaurant	Hotel	Bar	Breakfast Spot	Café	Chinese Restaurant

Methodology

K-Means

- To perform clustering on the neighbourhoods based on the venue frequency

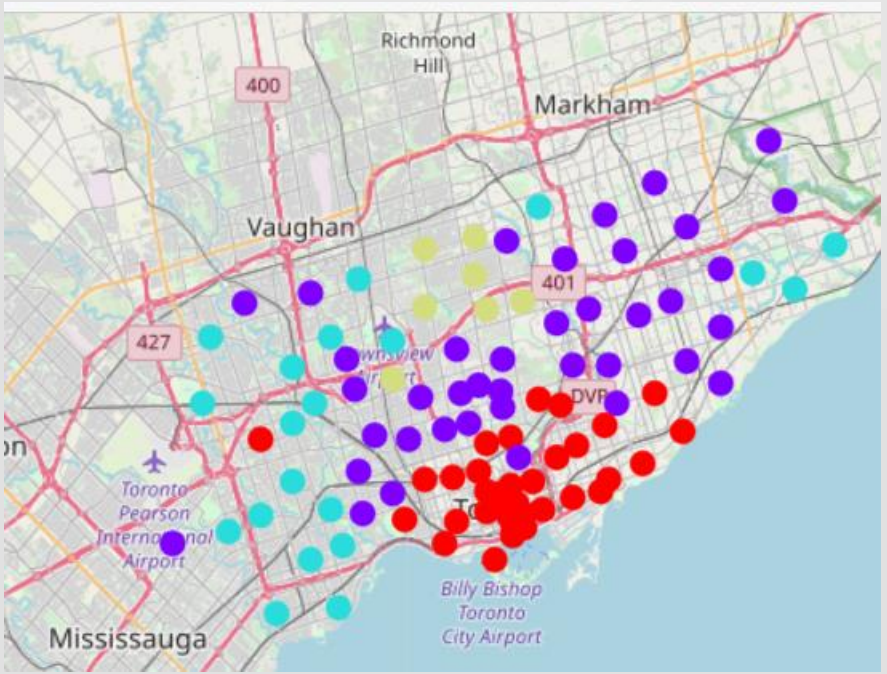
DBSCAN

- To perform venue clustering based on concentration

Results

K-Means

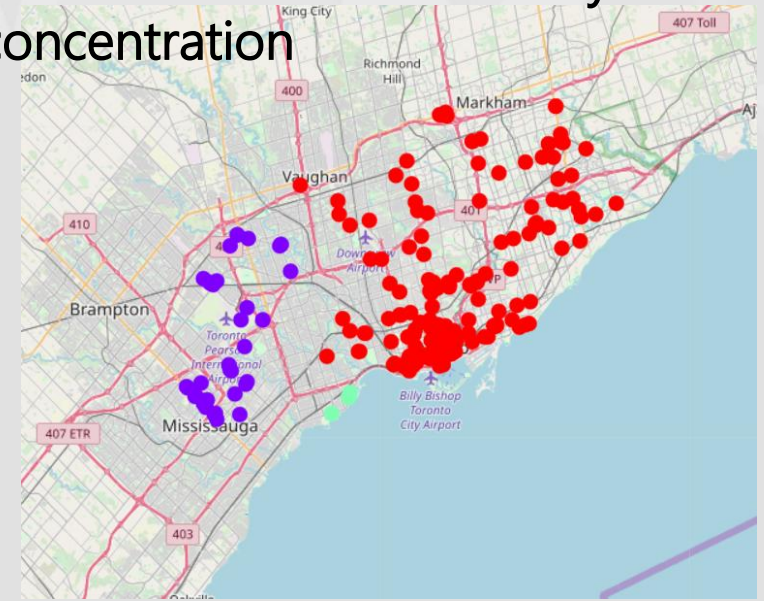
- With 4 clusters
  - Purple (high concentration)
  - Cyan (low concentration)
  - Red (medium concentration)
  - Green (low concentration)



Results

DBSCAN

- Per venue location and applied to neighbourhoods shows density areas and outliers where there is little concentration



Discussion and Conclusion

Recommendation

- Avoid Central area with high density clusters on both K-means and DBSCAN
- Outliers areas in DBSCAN are a good bet
- Red and Green clusters in K-Means have less competition

Conclusion

- Good insight to avoid high competition areas
- More data would be useful (e.g. population density, transports, etc.)