IBM Data Science Professional Certificate Applied Capstone

Exploring to open a Portuguese Restaurant in Toronto

Mulhchar Naim

Business Problem

- Toronto
  - 3 Million inhabitants
  - 16th place worldwide with 272 restaurants per 100.000 inhabitants
- Portuguese community
  - More than 170.000 people just in Toronto
  - Portuguese cuisine as one of the richest in the world
- If na investor is looking to open a new Portuguese restaurant where would you recommend it?

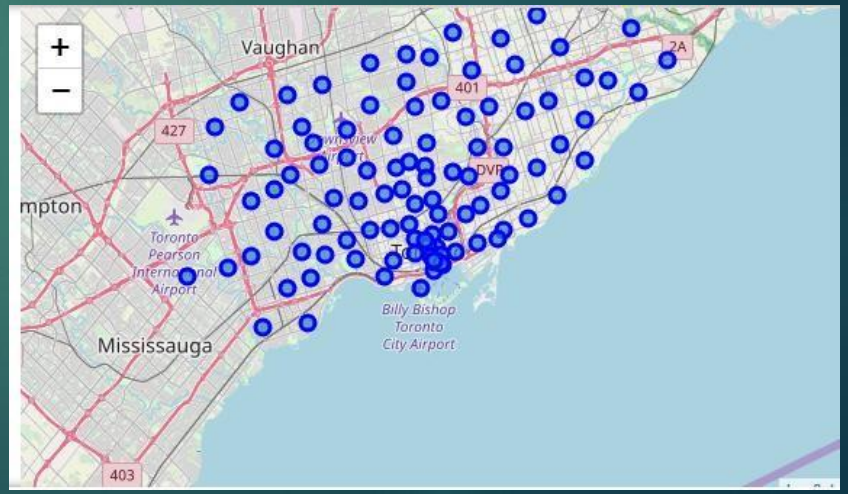
Data

- List of neighbourhoods in Toronto
  - Boroughs and postal codes (Wikipedia)
  - Geo coordinates
- Portuguese venues data
  - Using Foursquare API
    - Category Id for Portuguese 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

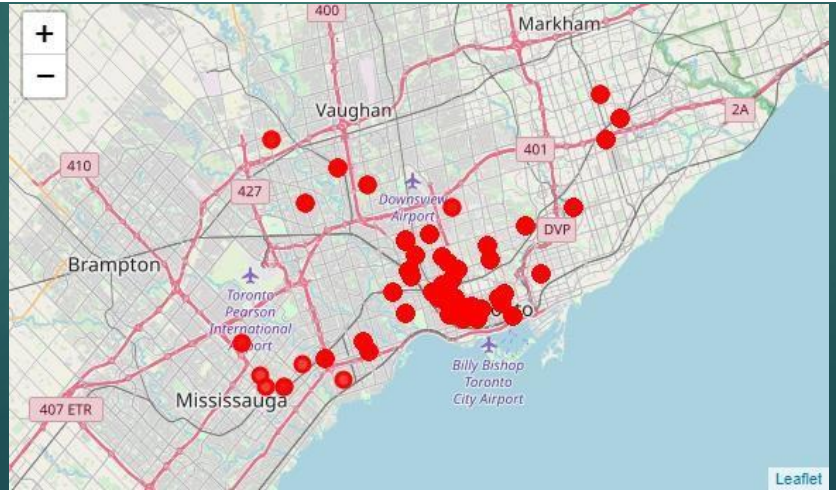


Methodology

Portuguese Restaurants Data Exploratory Analysis

- **Using Foursquare we retrieved all venues per neighbourhood**

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1	Cedarbrae	43.773136	-79.239476	Nando's Flame-Grilled Chicken	43.773113	-79.281166	Portuguese Restaurant



Methodology

Data wrangling

- **Create frequency based data frame per neighbourhood**

	Neighborhood	Asian Restaurant	BBQ Joint	Bakery	Bar	Beer Bar	Café	Dessert Shop	Gay Bar	Italian Restaurant	Medit R
0	Adelaide,King,Richmond	0.000000	0.0	0.290323	0.032258	0.032258	0.0	0.0	0.032258	0.032258	
1	Agincourt	0.333333	0.0	0.333333	0.000000	0.000000	0.0	0.0	0.000000	0.000000	
2	Agincourt North,L'Amoreaux East,Milliken,Steel...	0.333333	0.0	0.333333	0.000000	0.000000	0.0	0.0	0.000000	0.000000	
3	Albion Gardens,Beaumond Heights,Humbergate,Jam...	0.000000	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.000000	0.000000	
4	Alderwood,Long Branch	0.000000	0.0	0.250000	0.000000	0.000000	0.0	0.0	0.000000	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	Portuguese Restaurant	Bakery	Restaurant	Wine Bar	Sandwich Place	Mediterranean Restaurant	Italian Restaurant
1	Agincourt	Portuguese Restaurant	Bakery	Asian Restaurant	Wine Bar	Seafood Restaurant	Sandwich Place	Restaurant
2	Agincourt North,L'Amoreaux East,Milliken,Steel...	Portuguese Restaurant	Bakery	Asian Restaurant	Wine Bar	Seafood Restaurant	Sandwich Place	Restaurant
3	Albion Gardens,Beaumond Heights,Humbergate,Jam...	Portuguese Restaurant	Wine Bar	Seafood Restaurant	Sandwich Place	Restaurant	Pizza Place	Mediterranean Restaurant
4	Alderwood,Long Branch	Portuguese Restaurant	Restaurant	Bakery	Wine Bar	Seafood Restaurant	Sandwich Place	Pizza Place

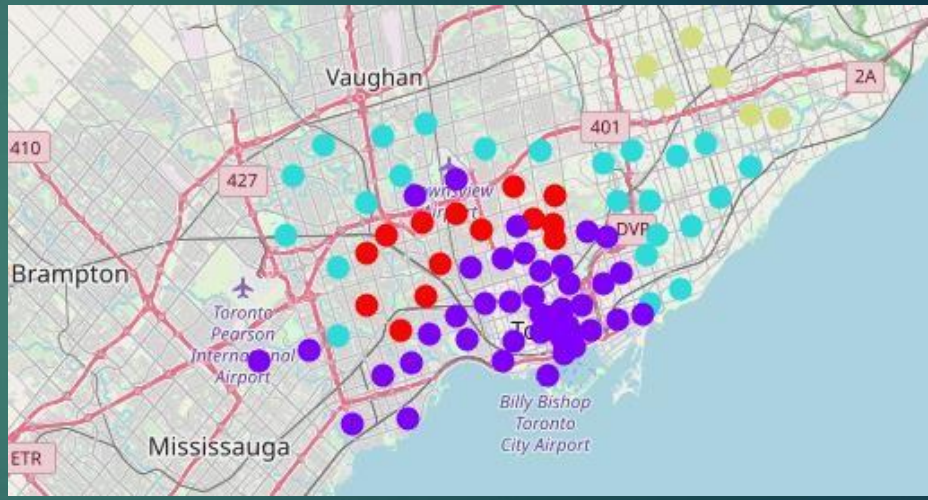
Methodology

- **K-Means**
  - **To perform clustering on the neighbourhoods based on the venue frequency**
- **DBSCAN**
  - **To perform venue clustering based on concentration**

Results

K-Means

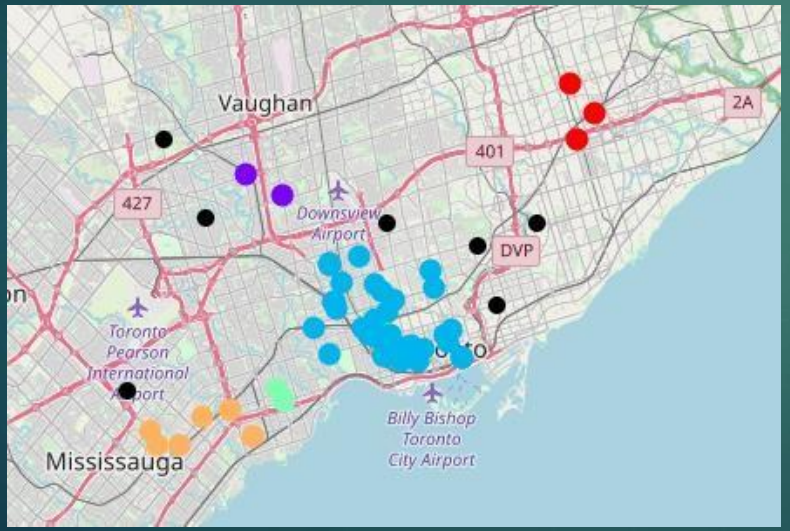
- With 4 clusters
  - Purple (high concentration)
  - Cyan (medium concentration)
  - Red (low concentration)
  - Green (low concentration)

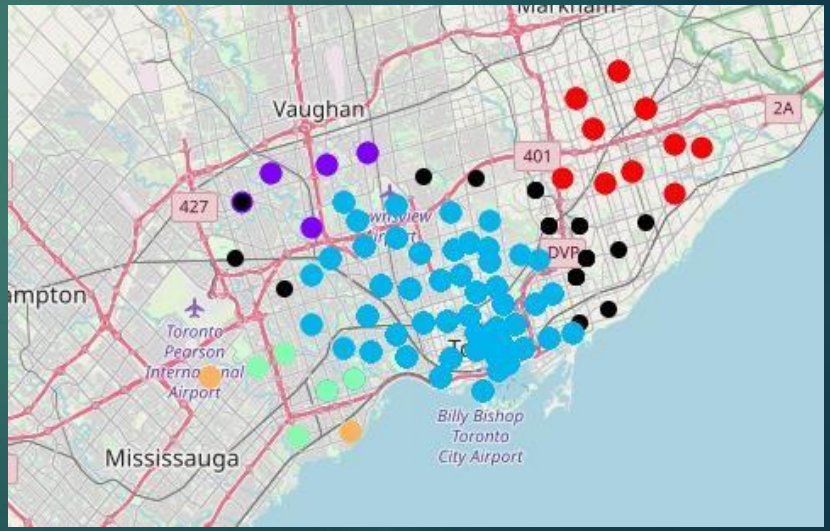


Results

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.)