THE BATTLE OF NEIGHBORHOODS

IBM Data Science Professional Certificate: Capstone Project

Authored by Hritik Panchasara on March 31, 2020

PROBLEM DEFINITION

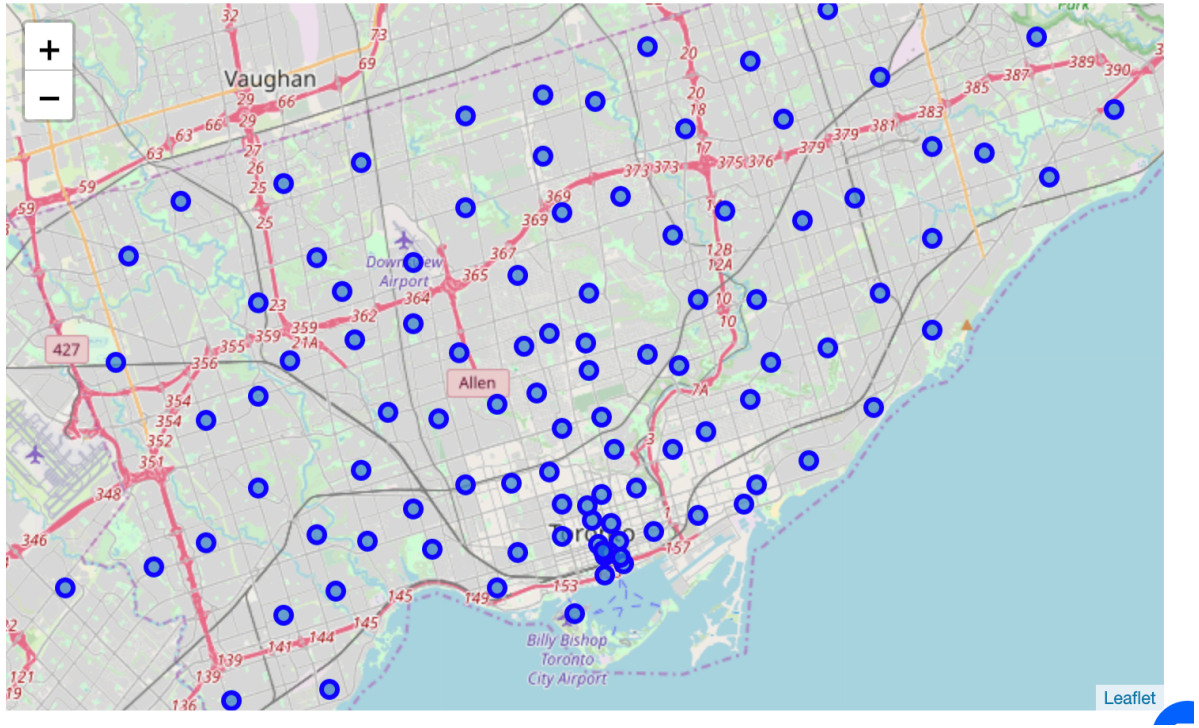
- Comparing the most common venue categories in the neighborhoods and boroughs of Toronto for any prospective stakeholder looking to invest
- Information about venue categories and their frequencies across the city of Toronto will help optimize and find the ideal neighborhood to catalyze an investment project
- Advantages of each neighborhood will be clearly drawn out so that the best possible final venue category can be chosen depending on the needs of the stakeholder

DATA SOURCING

- The following sources will be used to acquire data which will then be used to analyze the city of Toronto:
  - A Wikipedia page consisting of neighborhood data in Toronto. Web scraping will allow us to access and format this data using the pandas library to create dataframes. (Link: https://en.Wikipedia.org/wiki/List_of_postal_codes_of_Canada:M)
  - For geographical coordinates of each postal code, a CSV file will be read. (Link: https://cocl.us/Geospatial_data)
  - Leverage neighborhood data using FourSquare API to get venue details against the information stored in dataframes

DATA CLEANSING

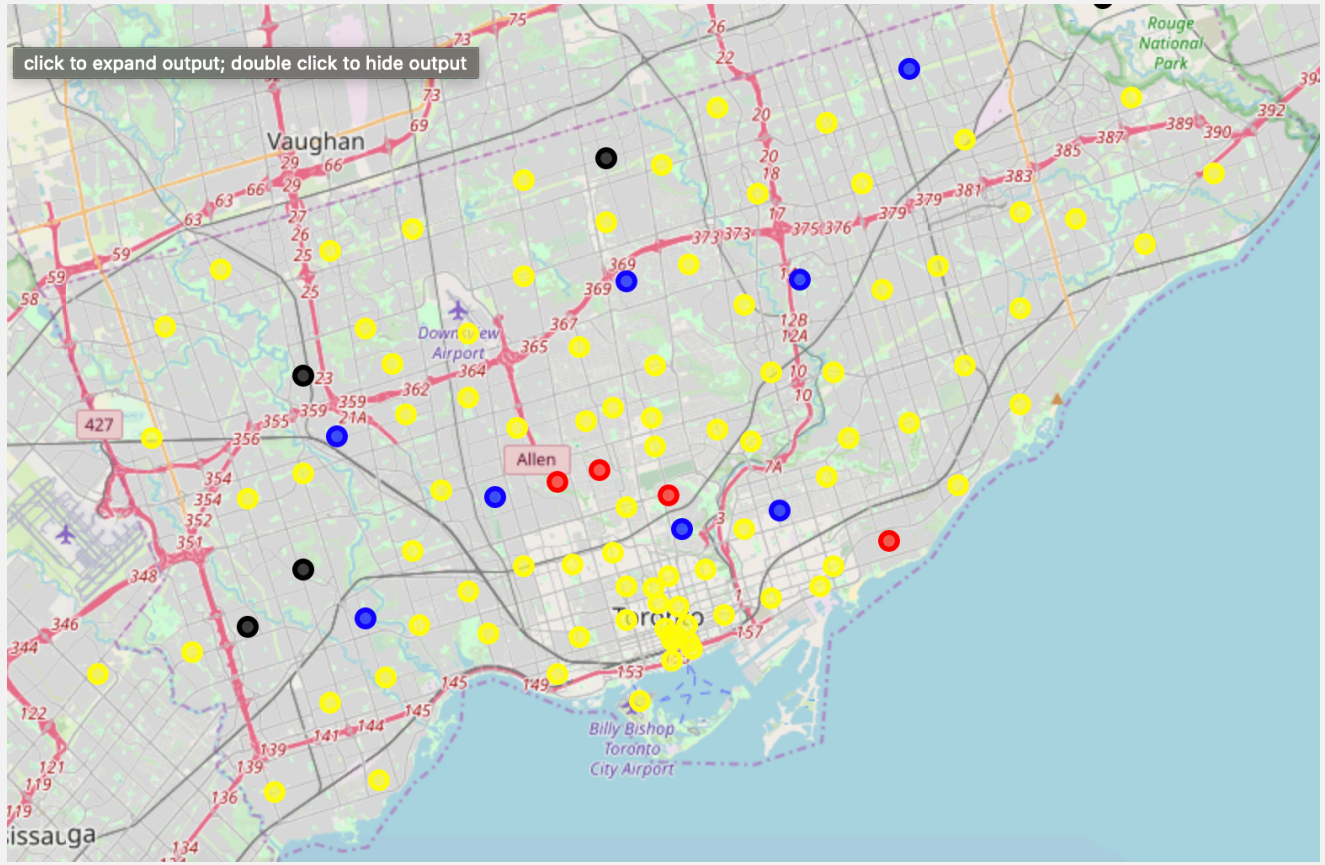
- Data downloaded from multiple sources were all combined into a single table
- When the scraped data was was first accessed, there were tuples that did not have an assigned borough or neighborhood. We drop cells with a borough value of "NaN" or "Not Assigned"
- One tuple with a primary value of a postcode can consist of multiple neighborhoods. These rows can be combined and the neighborhood values can be separated with commas using 'groupby'



INITIAL ANALYSIS

The dataset shows 10 boroughs and 103 neighborhoods.

- Different neighborhoods have similar venues that have the highest frequency in said neighborhoods.
- A clustering model that utilizes the K-means algorithm can be used to cluster similar types of venues. We have taken K=3 clusters.
- These clusters are finally visualized. Same colors represent same cluster types. After visualizing the folium-generated map, some observations were made:



CLUSTER 1

	Borough	Cluster Labels BON	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	Scarborough	0.0	1.0	Park	Playground	Doner Restaurant	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Dog Run	Donut Shop
23	North York	0.0	1.0	Park	Bank	Convenience Store	Drugstore	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant
25	North York	0.0	1.0	Park	Food & Drink Shop	Convenience Store	Women's Store	Donut Shop	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant
40	East York	0.0	1.0	Park	Coffee Shop	Convenience Store	Empanada Restaurant	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Dessert Shop	Drugstore
50	Downtown Toronto	0.0	1.0	Park	Playground	Trail	Empanada Restaurant	Electronics Store	Eastern European Restaurant	Ethiopian Restaurant	Dumpling Restaurant	Drugstore	Department Store
74	York	0.0	1.0	Park	Women's Store	Market	Doner Restaurant	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Dog Run	Drugstore
90	Etobicoke	0.0	1.0	Smoke Shop	River	Park	Women's Store	Dog Run	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Donut Shop
98	York	0.0	1.0	Convenience Store	Park	Drugstore	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Women's Store

CLUSTER 2

	Borough	Cluster Labels BON	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	Scarborough	0.0	1.0	Park	Playground	Doner Restaurant	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Dog Run	Donut Shop
23	North York	0.0	1.0	Park	Bank	Convenience Store	Drugstore	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Dumpling Restaurant
25	North York	0.0	1.0	Park	Food & Drink Shop	Convenience Store	Women's Store	Donut Shop	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant
40	East York	0.0	1.0	Park	Coffee Shop	Convenience Store	Empanada Restaurant	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Dessert Shop	Drugstore
50	Downtown Toronto	0.0	1.0	Park	Playground	Trail	Empanada Restaurant	Electronics Store	Eastern European Restaurant	Ethiopian Restaurant	Dumpling Restaurant	Drugstore	Department Store
74	York	0.0	1.0	Park	Women's Store	Market	Doner Restaurant	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Dog Run	Drugstore
90	Etobicoke	0.0	1.0	Smoke Shop	River	Park	Women's Store	Dog Run	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Donut Shop
98	York	0.0	1.0	Convenience Store	Park	Drugstore	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Women's Store

CLUSTER 3

	Borough	Cluster Labels BON	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	Scarborough	1.0	2.0	Fast Food Restaurant	Dessert Shop	Event Space	Ethiopian Restaurant	Empanada Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Donut Shop
1	Scarborough	1.0	2.0	History Museum	Bar	Drugstore	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Women's Store	Dim Sum Restaurant
2	Scarborough	1.0	2.0	Electronics Store	Rental Car Location	Mexican Restaurant	Bank	Intersection	Moving Target	Medical Center	Spa	Breakfast Spot	Donut Shop
3	Scarborough	1.0	2.0	Coffee Shop	Korean Restaurant	Women's Store	Drugstore	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Donut Shop	Eastern European Restaurant
4	Scarborough	1.0	2.0	Fried Chicken Joint	Gas Station	Hakka Restaurant	Bank	Athletics & Sports	Caribbean Restaurant	Thai Restaurant	Bakery	Dim Sum Restaurant	Diner
5	Scarborough	1.0	2.0	Grocery Store	Playground	Donut Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Dog Run	Doner Restaurant	Drugstore

RESULTS

Borough	Most Common Venue Category
East Toronto	Cafes and Restaurants
Central Toronto	Gym and Park
York	Trails and Parks
Scarborough	Fast Food Restaurants
North York	Grocery Stores, Recreational Area/Parks
Etobicoke	Pizza Place
Downtown Toronto	Coffee Shop
East York	Restaurants

CONCLUSION

- Analysis proves that great number of different types of venue categories exist in Toronto. There is a correlation between clusters that similar types exist in the same borough
- There is additional scope to this project as more factors can be considered such as crime occurrences in certain boroughs, levels of noise/proximity to major roads, demographic information to target the right audience, real estate availability, socioeconomic groups residing in different areas and so on.