Comparison of neighborhoods of New York and Toronto

Tomasz Tomaszewski

June, 2021

Table of contents

- Executive Summary

- Introduction

- Methodology

- Results
  - Collection and exploration of data.
  - Clustering of neighbourhoods.
  - Visualisation of data.

- Discussion
  - Findings & Implications

- Conclusion

- Appendix

EXECUTIVE SUMMARY

AIM OF ANALYSIS:

1) Exploration of New York and Toronto .

2) Comparison of neighbourhoods of both metropolies.

3) Finding most suitable neighbourhood.

KEY FINDINGS:

1) Diversity of venues in both cities.

2) Similarities and differences between metropolies.

3) Most frequent common venues:

 - Coffe shops and caffee's.

- Restaurants.

 -Parks.

5) Similarities and differences between within the cities.

INTRODUCTION

- Project will retrieve data and compare two major North American metropolies of New york and Toronto.

- The key questions to answer are:
  What are similarities and differences between the cities?
  What will be the best location to choose?
  Will this exploration facilitate decision making?

- Report is aimed to give insights, and serves as starting point before deciding on a location of choice.

- Report is intended for use by: business executives, students, tourists, job seekers etc.

METHODOLOGY
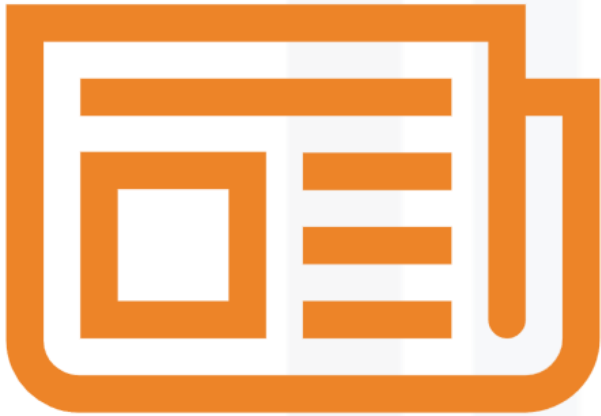
DATA COLLECTION

1) New York dataset in the form of JSON file.

Collected from URL:

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json

2) Toronto data collected in two steps:

- Webscrapping with Beautiful Soup library from URL:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

- Pandas "read_csv" method for extracting geospatial coordinates from following CSV file:

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs_v1/Geospatial_Coordinates.csv

3) Venues data downloaded with help of Foursquare API.

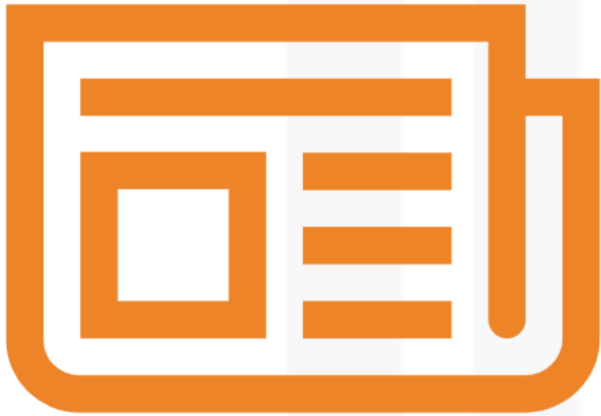4) All datasets stored as Pandas libraries.

METHODOLOGY
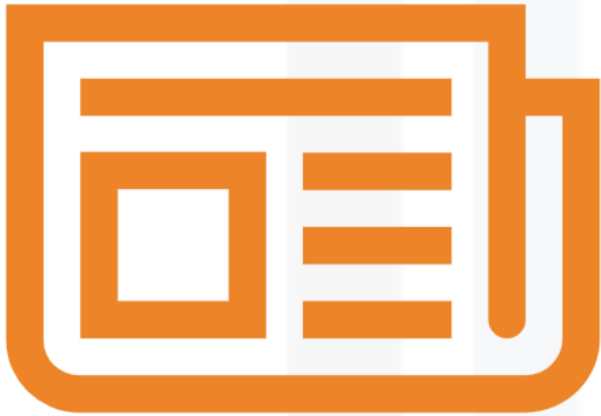
DATA EXPLORATION

Exploration of data set stored as Pandas libraries.

Visualising data with folium map.

DATA WRANGLING

Identification and removal of duplicate values in the dataset.

Identification of missing values in the dataset.

Normalization of data in the dataset.

METHODOLOGY

EXPLORATORY DATA ANALYSIS

Finding out the distribution of data.

Comparison of datasets..

Distribution of data on folium maps.

DATA PROCESSING AND VISUALISATION

Clustering of neighbourhoods.

Distribution of data on folium maps.

Creation of bar charts of main venues.

Comparison of data.

RESULTS

- **Retrieving New York dataset from JSON file with use of JSON library.**
- Stored as Pandas dataframe.

```
neighborhoods_ny.head(10)
```

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585
5	Bronx	Kingsbridge	40.881687	-73.902818
6	Manhattan	Marble Hill	40.876551	-73.910660
7	Bronx	Woodlawn	40.898273	-73.867315
8	Bronx	Norwood	40.877224	-73.879391
9	Bronx	Williamsbridge	40.881039	-73.857446

RESULTS

Toronto dataset

- **Webscraping data with Python and Beautiful Soup.**

```
df.head() #Toronto dataframe review
```

	PostalCode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Queen's Park	Ontario Provincial Government

- Collecting geospatial coordinates from CSV file.

```
df_toronto_coord = pd.read_csv(geo_coordinates)
df_toronto_coord.head()
```

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-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

RESULTS

- Venues data were downloaded with help of Foursquare API, which is popular source of venues and location data.

Foursquare data for New York and Toronto

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
1	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop
4	Marble Hill	40.876551	-73.91066	Astral Fitness & Wellness Center	40.876705	-73.906372	Gym

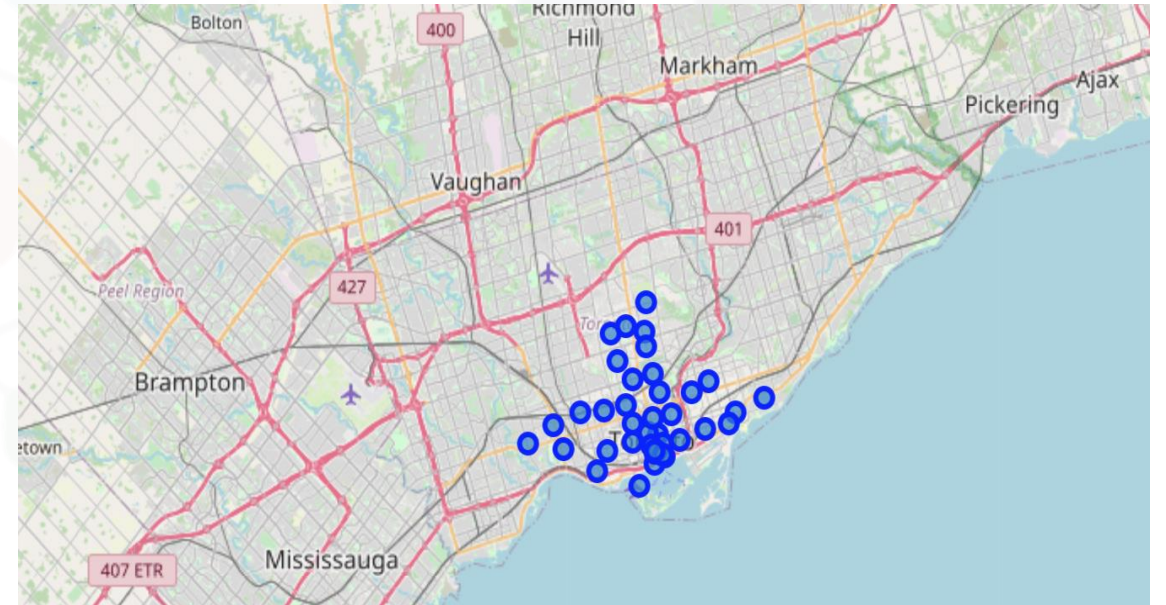|	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Regent Park, Harbourfront	43.65426	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant
4	Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa

IBM Developer

SKILLS NETWORK

RESULTS

- **Visualising extracted neighboirhoods data on map with Folium library.**

New York
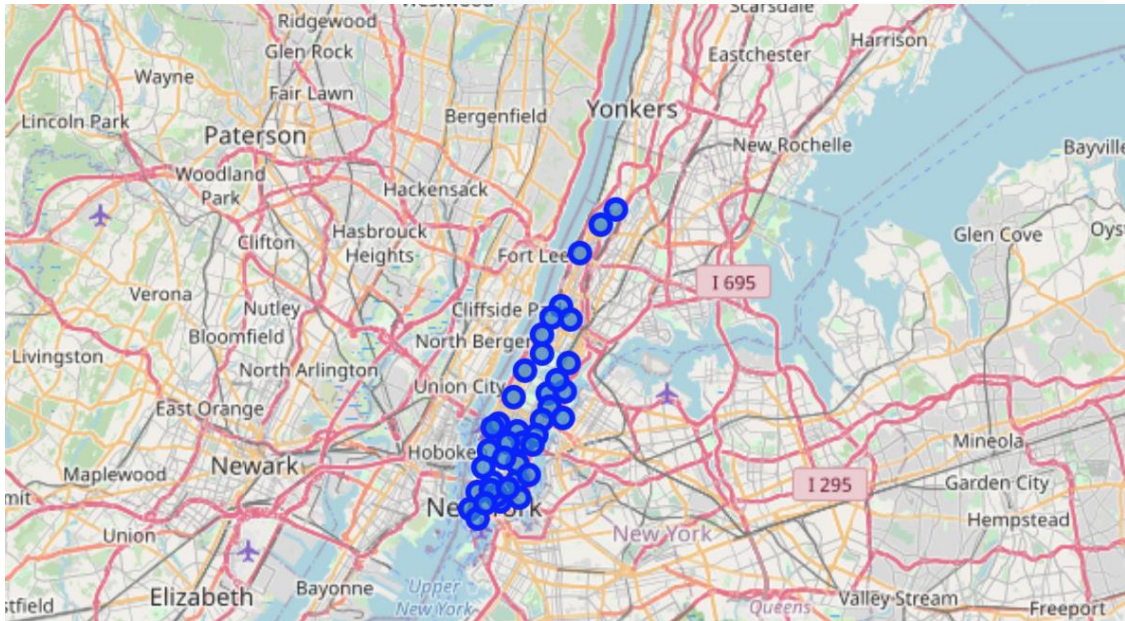
Toronto

RESULTS

- **Clustering of data into 5 clusters with K means method.**

```
#Cluster 5
```

```
mht_clust_5=manhattan_merged2.loc[manhattan_merged2['Cluster Labels'] == 4, manhattan_merged2.columns[[1] + list(range(5, man
hattan_merged2.shape[1]))]]
mht_clust_5.head()
```

|	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
22	Little Italy	Bakery	Italian Restaurant	Café	Ice Cream Shop	Hotel	Salon / Barbershop	Thai Restaurant	Chinese Restaurant	Mediterranean Restaurant	Sandwich Place

```
trt_clust_5=toronto_merged2.loc[toronto_merged2['Cluster Labels'] == 4, toronto_merged2.columns[[1] + list(range(6, toronto_m
erged2.shape[1]))]]
trt_clust_5
```

File display

|	Borough	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
35	Downtown Toronto	Coffee Shop	Café	Chinese Restaurant	Italian Restaurant	Pub	Pizza Place	Pharmacy	Bakery	Restaurant	American Restaurant

RESULTS

- **Presenting clusters on Folium maps. Part 1.**
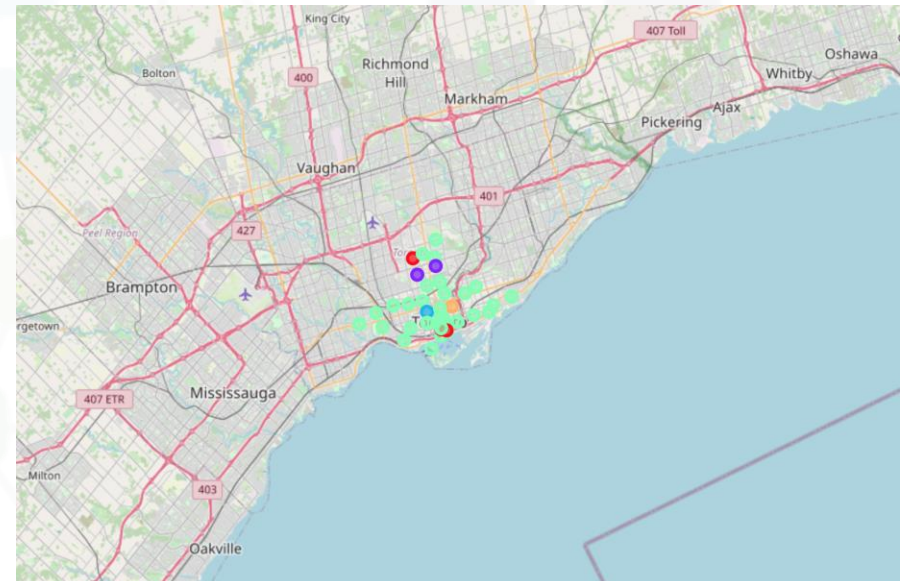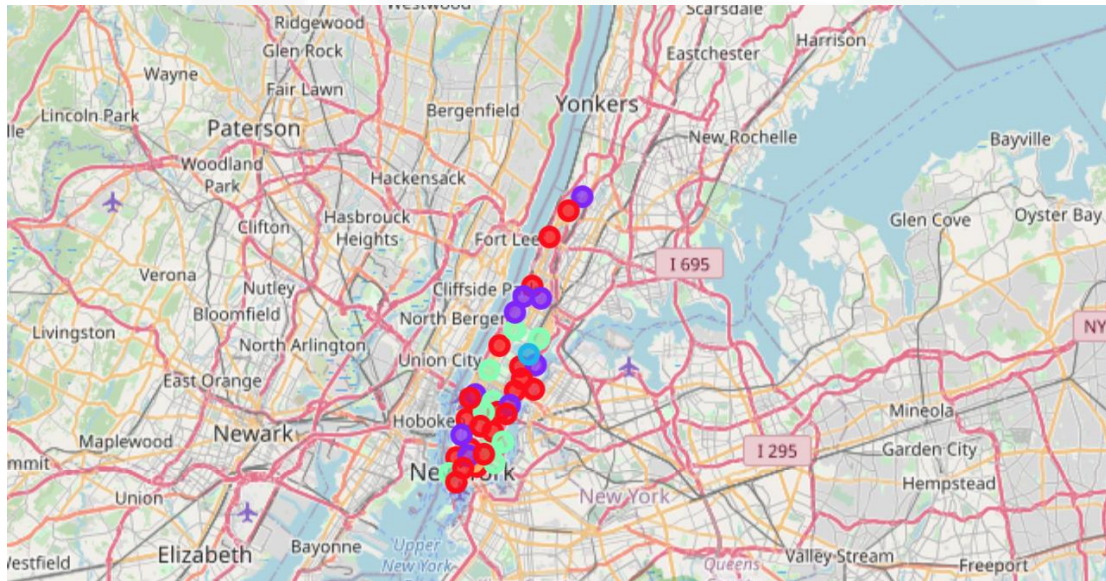
RESULTS

- **Presenting clusters on Folium maps. Part 2.**
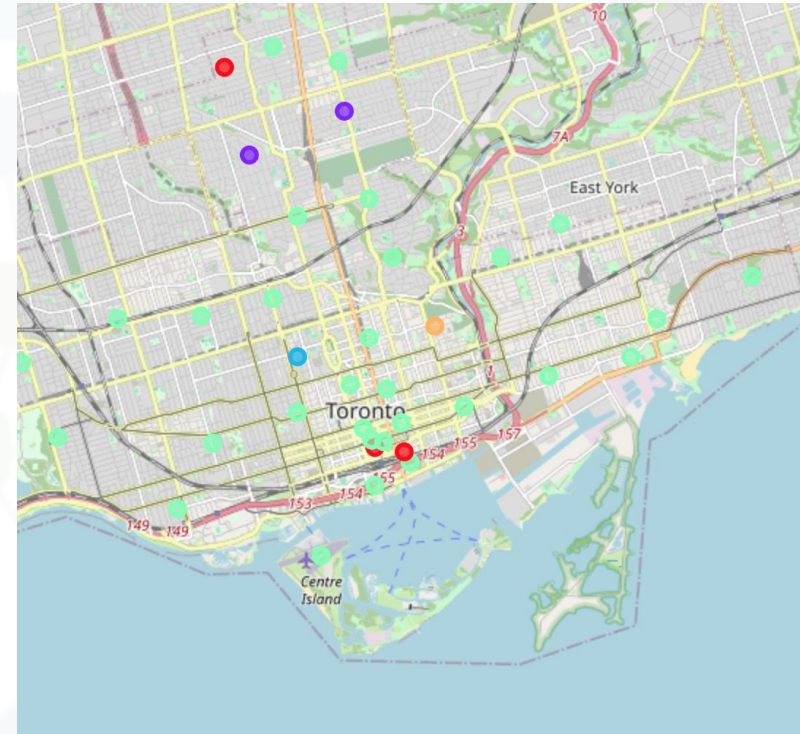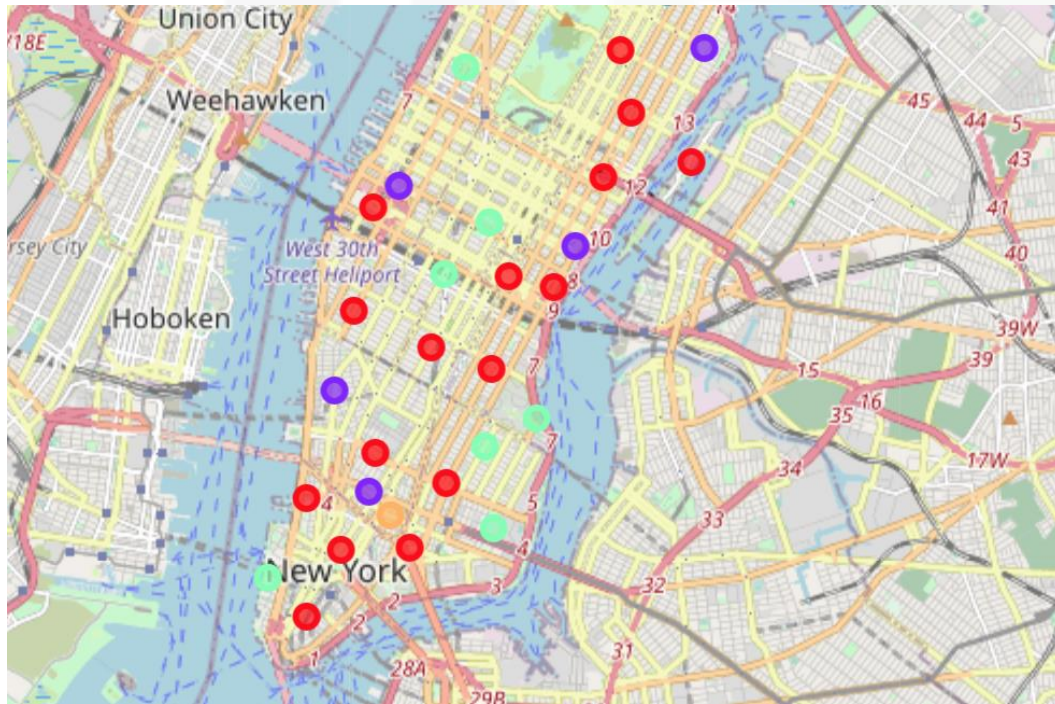
RESULTS

- **Top 15 venues for New York.**

Venue Category	
Coffee Shop	144
Italian Restaurant	133
Café	77
Bakery	77
Pizza Place	77
American Restaurant	77
Park	73
Hotel	69
Bar	66
Mexican Restaurant	58
Cocktail Bar	53
Gym / Fitness Center	52
Gym	51
Wine Shop	49
French Restaurant	45

Most frequent venues of Manhattan by number of venues
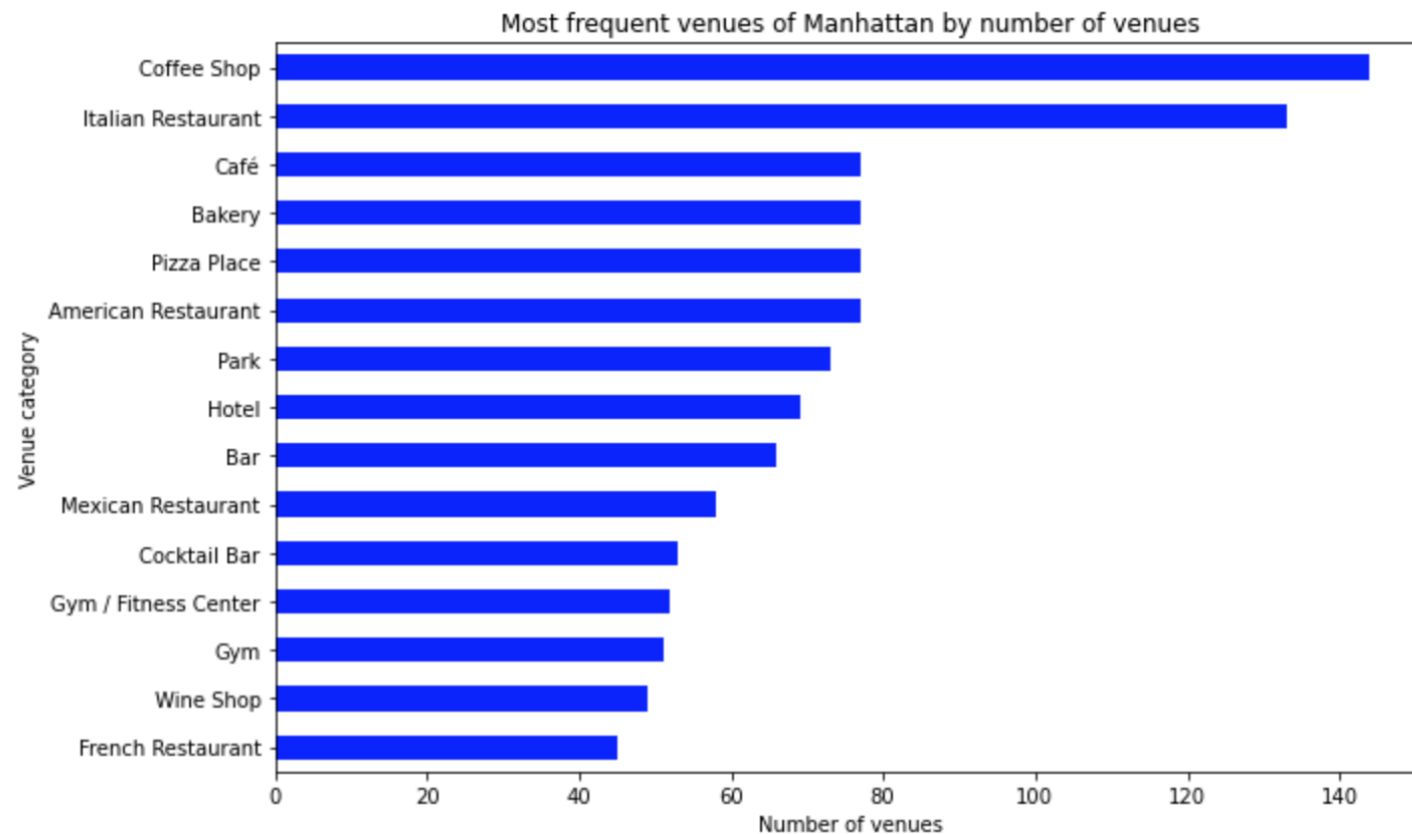
RESULTS

- **Top 15 venues for Toronto.**

Venue Category	
Coffee Shop	150
Café	84
Restaurant	50
Italian Restaurant	41
Hotel	39
Park	37
Japanese Restaurant	32
Bakery	31
Pizza Place	27
Gym	24
Seafood Restaurant	24
Clothing Store	24
Bar	22
Sushi Restaurant	22
Sandwich Place	20

Most frequent venues of Toronto by number of venues
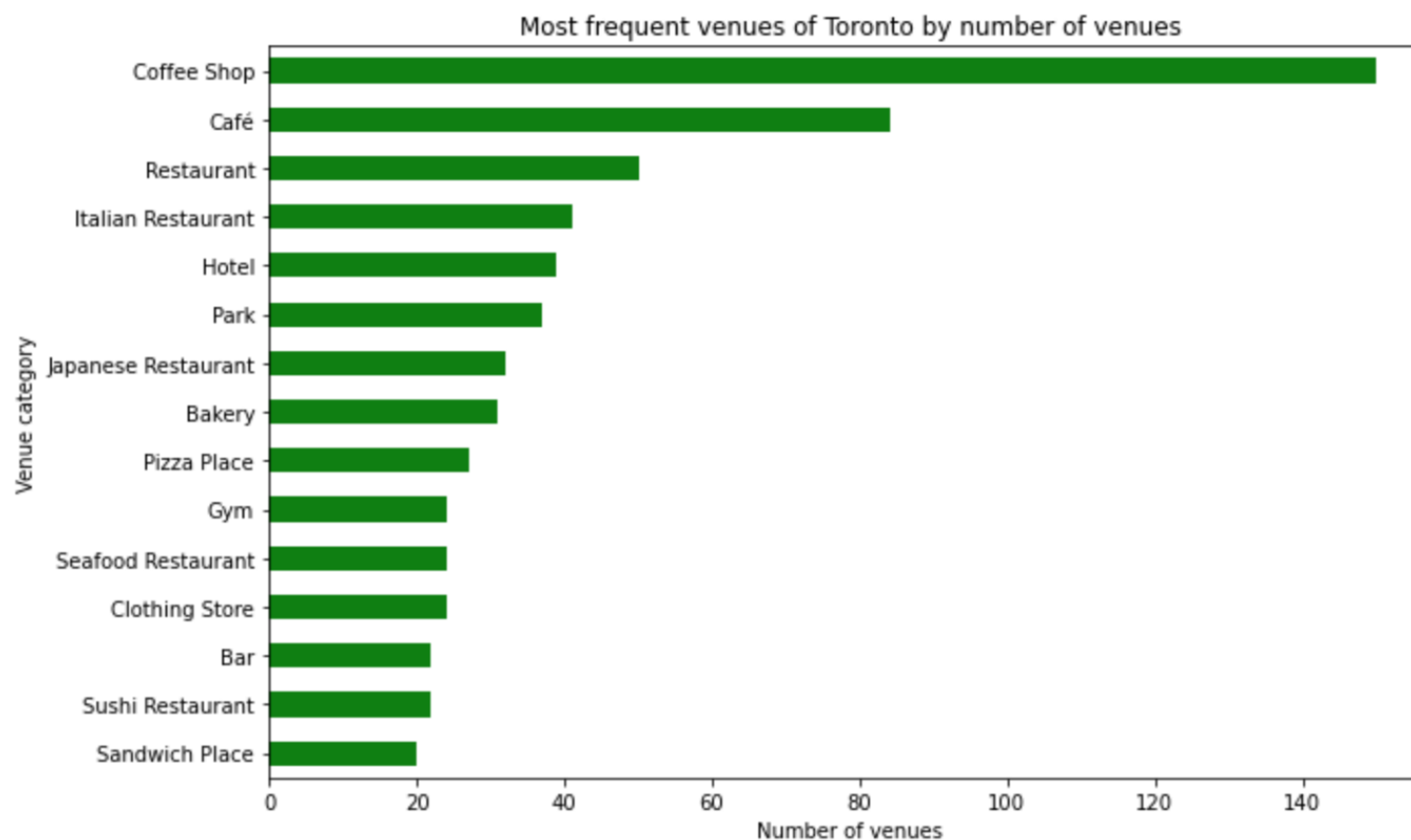
RESULTS

- **New York venues in number of neighbourhoods.**

	Number of neighbourhoods
Coffee Shop	39
Italian Restaurant	33
Pizza Place	33
Bakery	32
Café	32
American Restaurant	30
Bar	28
Park	28
Gym	27
Mexican Restaurant	27
French Restaurant	26
Wine Shop	25
Seafood Restaurant	25
Burger Joint	25
Gym / Fitness Center	24

Most frequent venues of Manhattan by number of neighborhoods
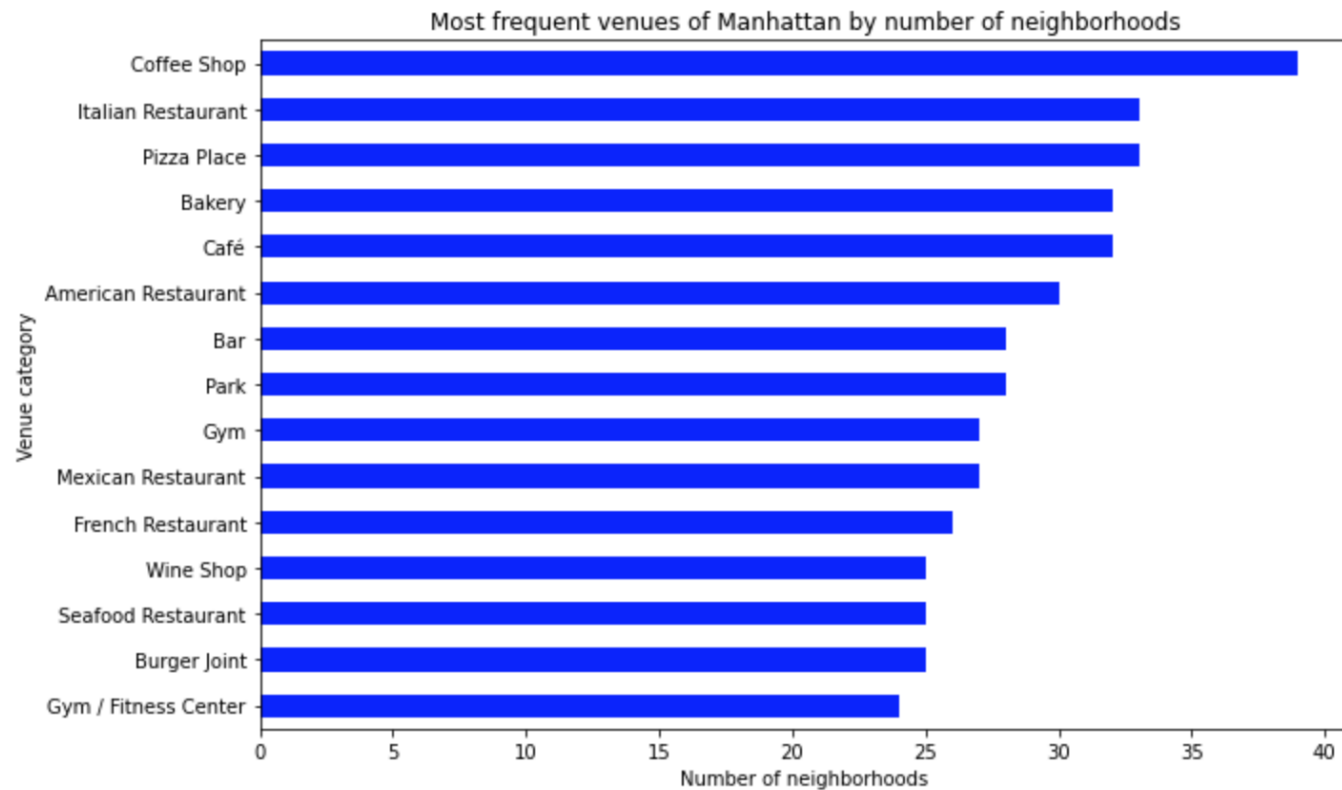
RESULTS

- **Toronto venues by number of neighbourhoods.**

	Number of neighbourhoods
Park	27
Coffee Shop	26
Café	26
Restaurant	25
Italian Restaurant	22
Bakery	19
Pizza Place	18
Bar	15
Japanese Restaurant	15
Pub	14
Sandwich Place	14
Thai Restaurant	13
Diner	13
Bookstore	13
Sushi Restaurant	13

Most frequent venues of Toronto by number of neighborhoods
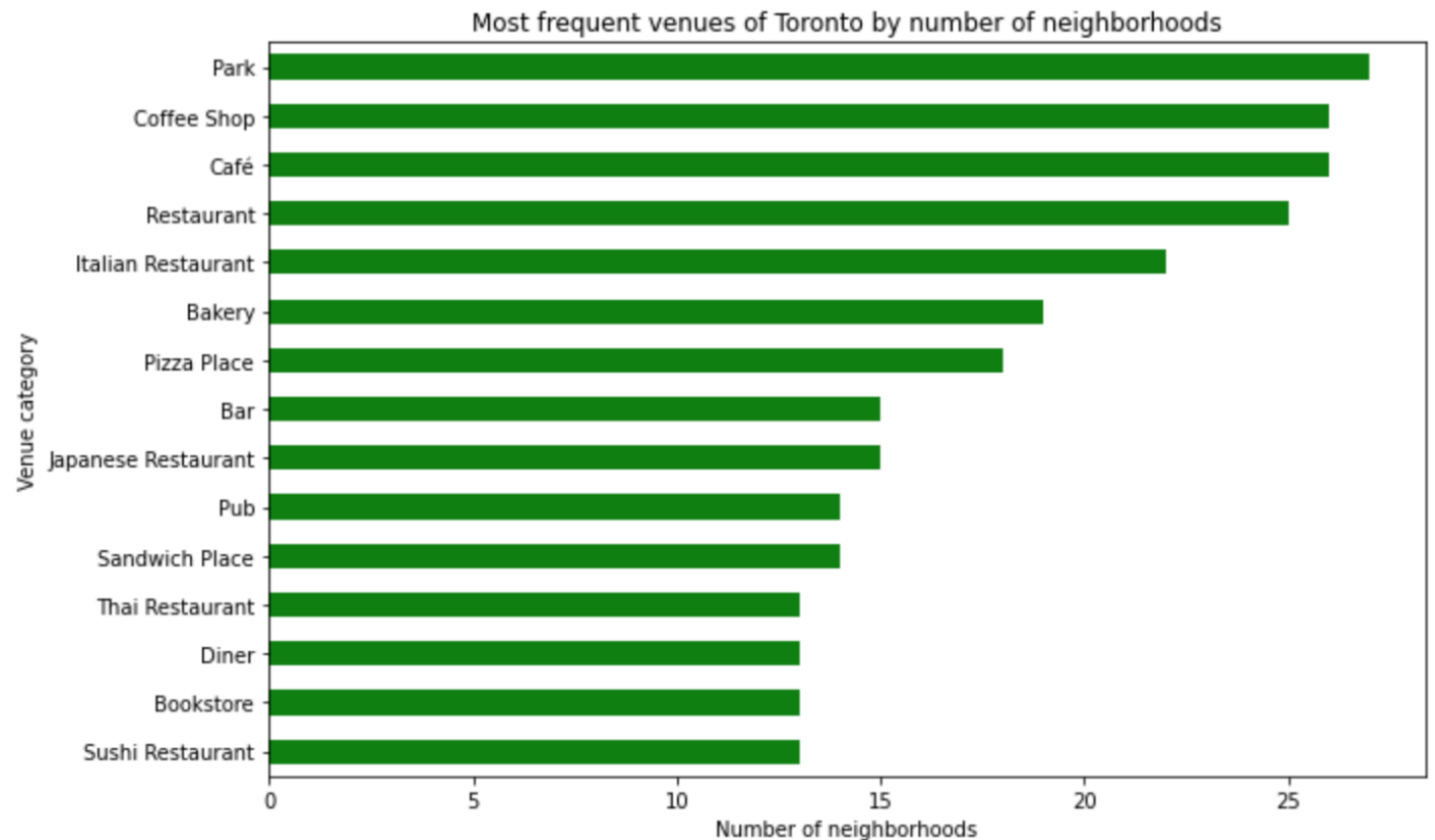
RESULTS - FINDINGS

Findings:

- Cofee shops, caffee's, restaurants and parks are very popular in both cities.

- Clusters of similar neighbourhoods can be identified based on similarities.

- Analysing of data for decision making:

1) choosing location to relocate and/or live,

2) finding best place to work based on neighbourhood facilities,

3) exploring the city based on most interesting venues.

IBM Developer

SKILLS NETWORK

DISCUSSION

- What can be other sources of data?

- Data driven decisions?

- Quality of data?

IBM **Dev**eloper

SKILLS NETWORK

OVERALL FINDINGS & IMPLICATIONS

Findings

- Diversity of venues in both NYC and Toronto.
- Similarities of two metropolies.
- Similarities of neighbourhoods within the city.
- Identification of clusters.

Implications

- Easier decision making for exploration of the city or moving to the city or within the city.
- Probably more data would be useful.
- Project can only serve as a first step in further exploration and analysis of data.

CONCLUSION

Both New York City and Toronto are varied and similar at the same time.

Both have a lot to offer for almost anyone.