City Comparison: Toronto vs New York

Capstone Project - The Battle of the Neighborhoods (Week 2)
Applied Data Science Capstone by IBM/Coursera

Introduction: Business Problem

Background:

New York City (NYC), often called simply New York, is the most populous city in the United States. With an estimated 2019 population of 8,336,817 distributed over about 302.6 square miles (784 km2), New York City is also the most densely populated major city in the United States. New York City has been described as the cultural, financial, and media capital of the world, significantly influencing commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports.

Toronto is the capital city of the Canadian province of Ontario. With a recorded population of 2,731,571 in 2016 distributed over about 243.3 square miles (630.2 km2) it is the most populous city in Canada and the fourth most populous city in North America. Toronto is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.

Introduction: Business Problem

Problem:

- Both cities are the main cities from each country and also the most populous cities in North America. So since they are important cities for each country, we want to find the differences between them comparing the neighborhoods and analyzing based on the venues of each cities.

Interest:

- The stakeholders will be international brands, who want to expand to North America, so that with the analysis, differences and characteristics that each city contributes presented in this project to choose the best option based on their requirements, brand category, etc.

Data

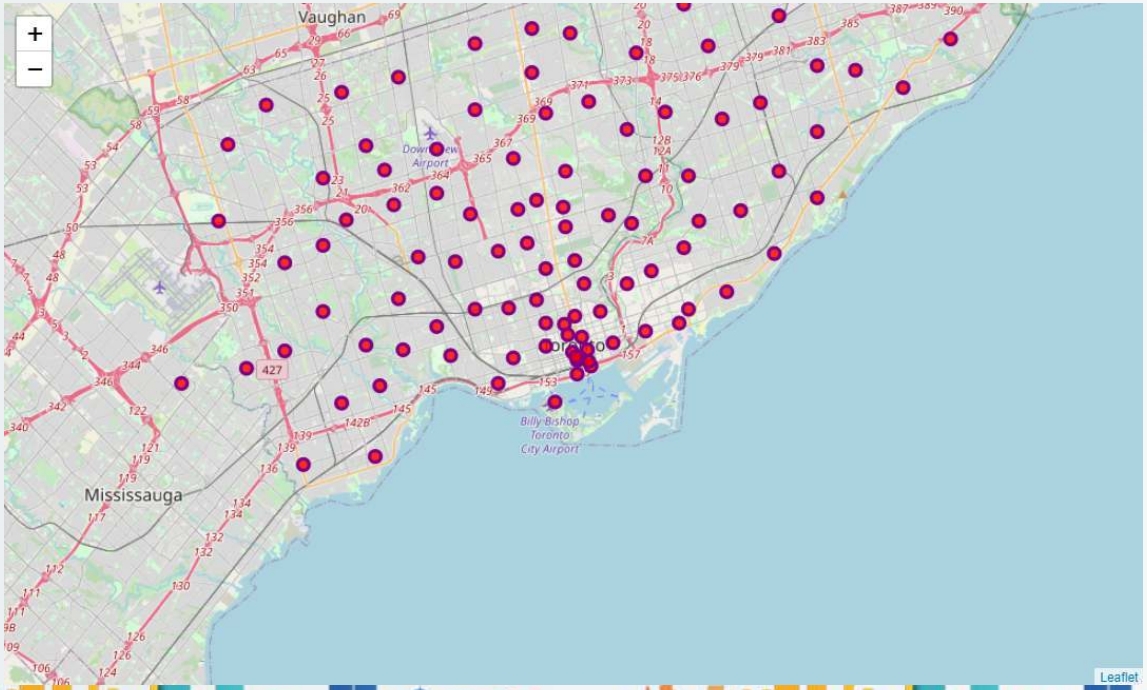
- Based on the description of our problem, first of all, we will need the Neighborhood/Borough data from each City and for that we gonna use a JSON file used in one of the course laboratories of New York City, which it contains Neighborhood/Borough with their respetive geospatial coordinates and in the case of Toronto City of Toronto, we gonna use web scrapping to get the Neighborhood/Borough with their postal codes additionally with a dataset that contains the geospatial coordinates of each postal code and finally join these two parts to obtain the desired dataframe.

- Secondly, we gonna use the Foursquare API to explore the venues from each Neighborhood of each Borough from each City, which will give us the following information from each venue: name, category, geospatial coordinates(Latitude, Longitude), city and venue ID, which It gonna help us to have the first analysis. Then, we will get detaliled information from some selected venues which will give us relevent information: verified status, rating, likes, tips and photos count with which we will finish our analysis. Finally, this will help us evaluate both cities and see the similarities / differences.

Neighborhood / Borough Data from each City

Toronto City

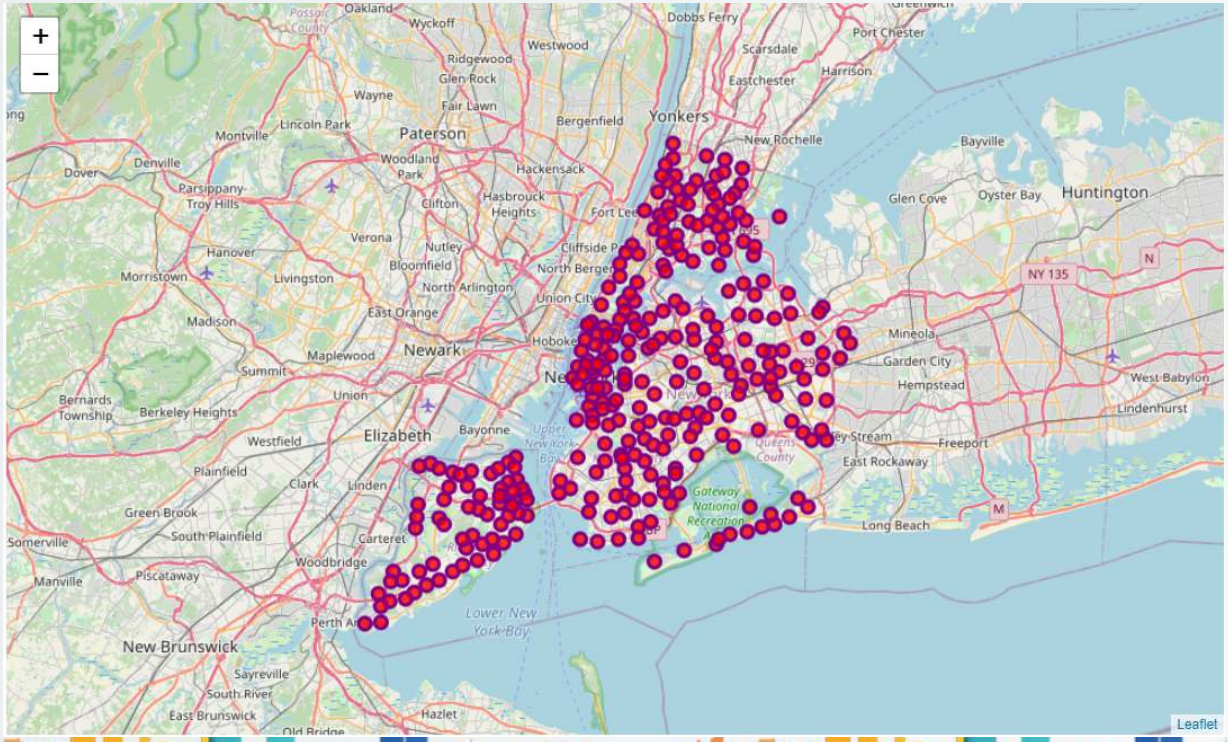
Borough	Neighborhood	Latitude	Longitude
Downtown Toronto	Berczy Park	43.644771	-79.373306
York	Caledonia-Fairbanks	43.689026	-79.453512
Scarborough	Woburn	43.770992	-79.216917
East York	Leaside	43.709060	-79.363452
Downtown Toronto	Central Bay Street	43.657952	-79.387383
Downtown Toronto	Christie	43.669542	-79.422564
Scarborough	Cedarbrae	43.773136	-79.239476
North York	Hillcrest Village	43.803762	-79.363452
North York	Bathurst Manor, Wilson Heights, Downsview North	43.754328	-79.442259
East York	Thorncliffe Park	43.705369	-79.349372



Neighborhood / Borough Data from each City

New York City

	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



Exploring venues from each Neighborhood

Toronto City - Venues Exploration

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	City	Venue ID
0	Parkwoods	43.753259	-79.329656	Allwyn's Bakery	43.759840	-79.324719	Caribbean Restaurant	Toronto	4b8991cbf964a520814232e3
1	Parkwoods	43.753259	-79.329656	Donalda Golf & Country Club	43.752816	-79.342741	Golf Course	Don Mills	4bd4846a6798ef3bd0c5618d
2	Parkwoods	43.753259	-79.329656	Tim Hortons	43.760668	-79.326368	Café	Toronto	57e286f2498e43d84d92d34a
3	Parkwoods	43.753259	-79.329656	Galleria Supermarket	43.753520	-79.349518	Supermarket	Toronto	4ccec87654f0b1f7f32824ca
4	Parkwoods	43.753259	-79.329656	Graydon Hall Manor	43.763923	-79.342961	Event Space	Toronto	4b8ec91af964a520053733e3

```
(3372, 9)
```

Exploring venues from each Neighborhood

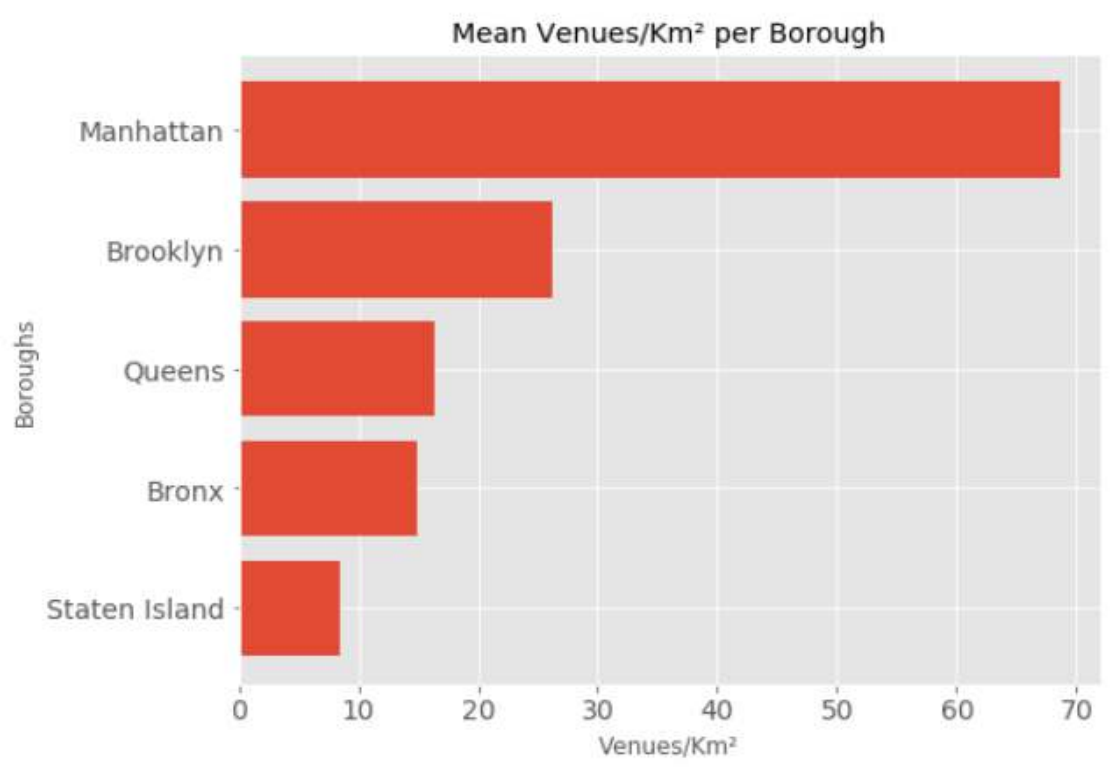
New York City - Venues Exploration

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	City	Venue ID
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop	Bronx	4c537892fd2ea593cb077a28
1	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop	Bronx	4c783cef3badb1f7e4244b54
2	Wakefield	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy	Bronx	5d5f5044d0ae1c0008f043c3
3	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy	Bronx	4d6af9426107f04dedeb297a
4	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop	Bronx	4c25c212f1272d7f836385c5

```
(4688, 9)
```

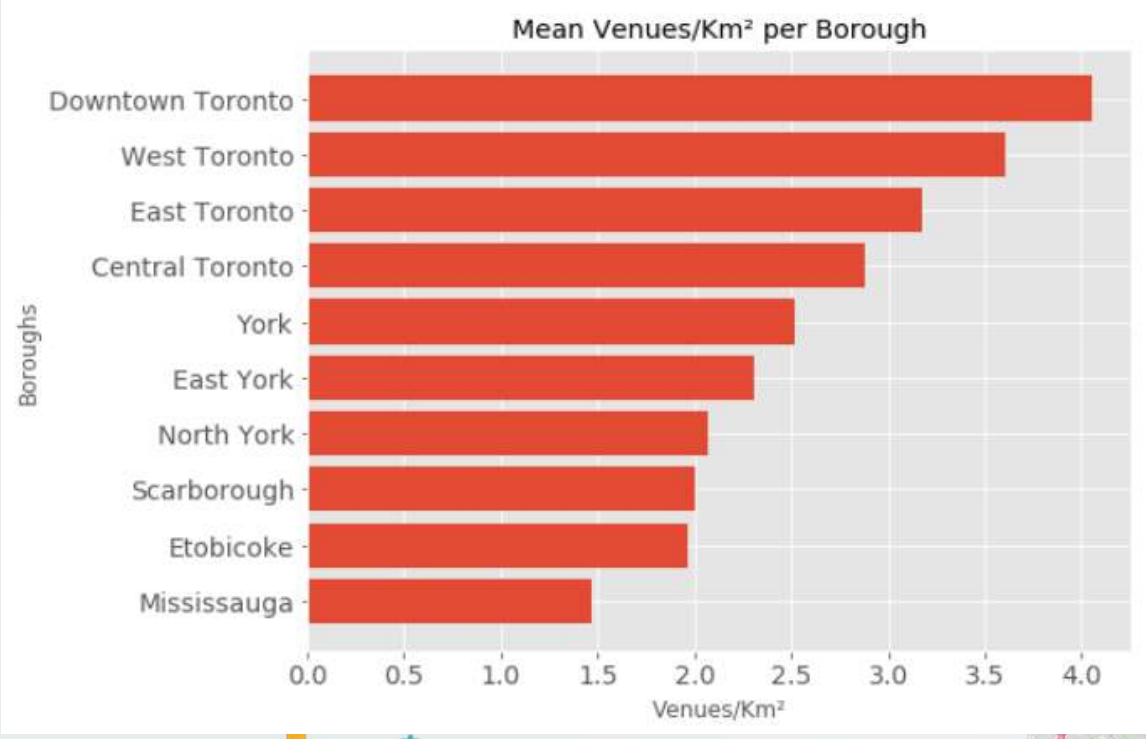
Venues Density Calculation

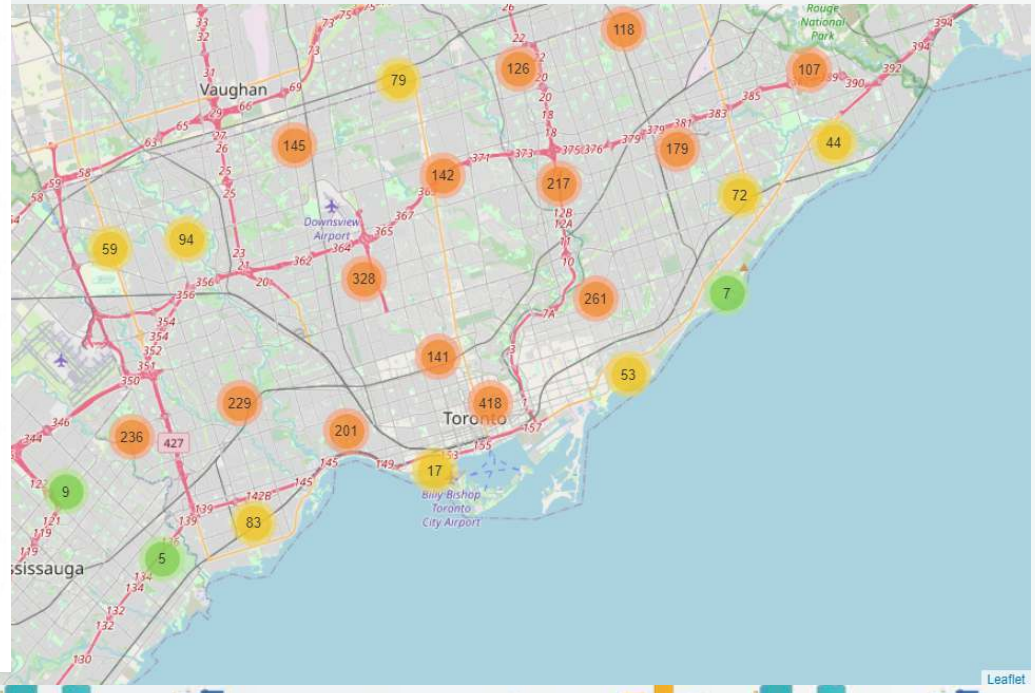
New York City



Venues Density Calculation

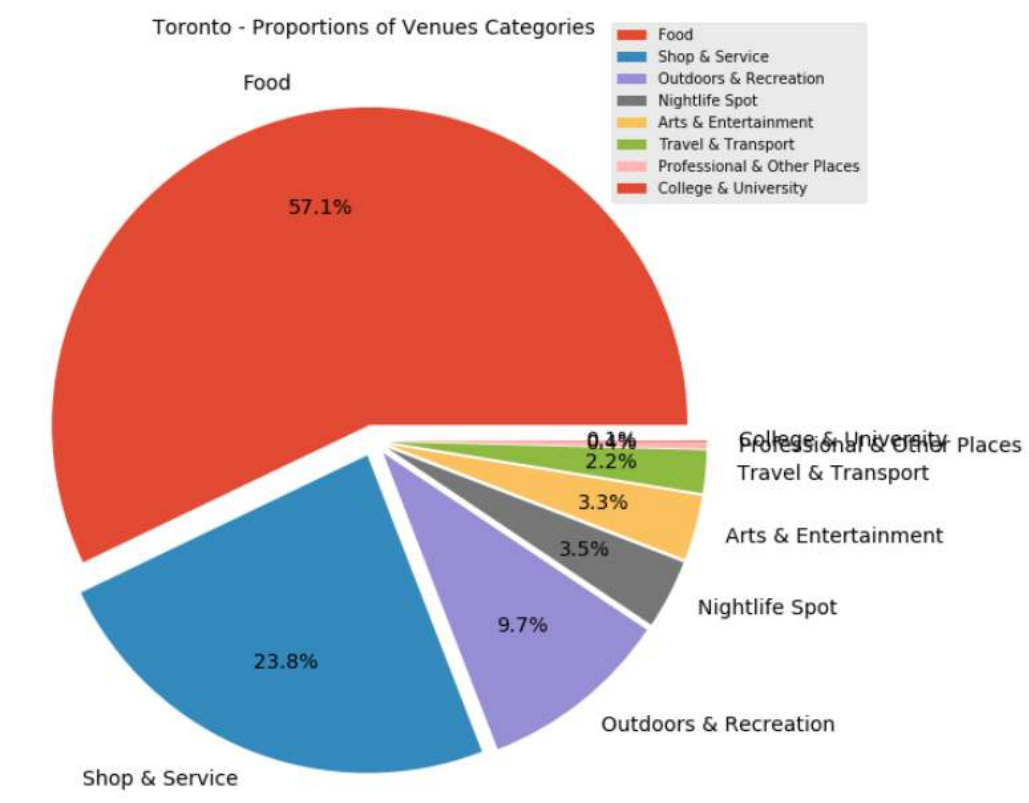
Toronto City

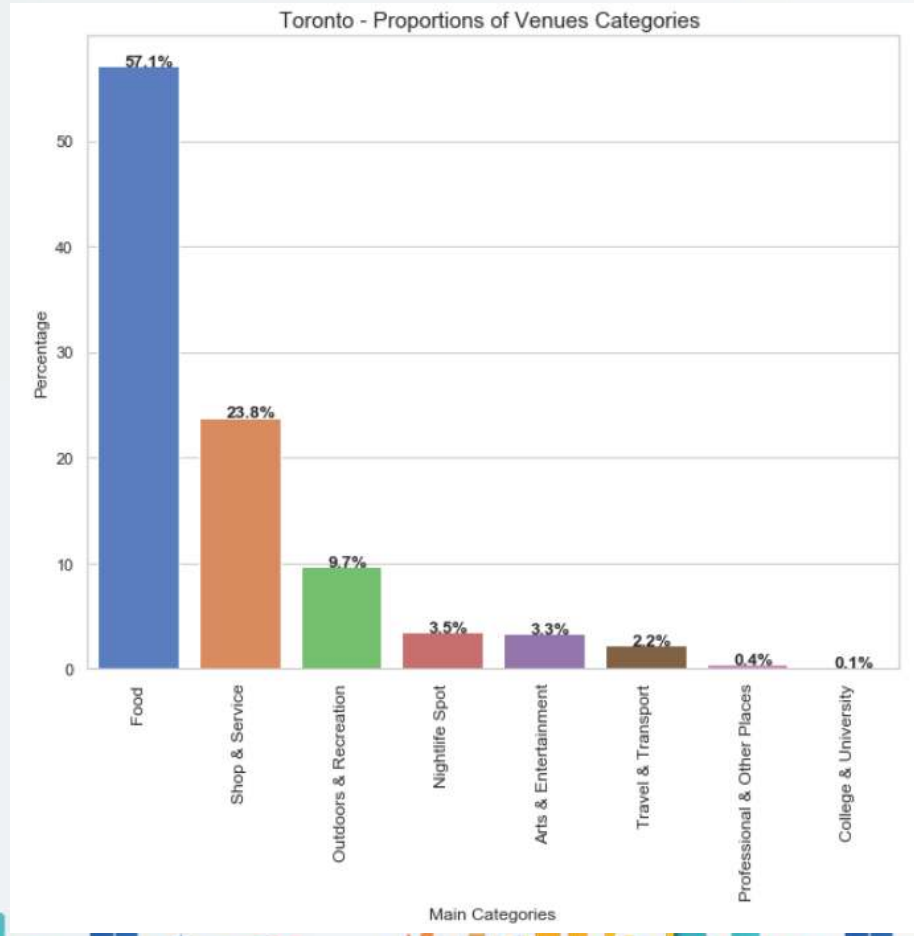




Category Diversity Analysis

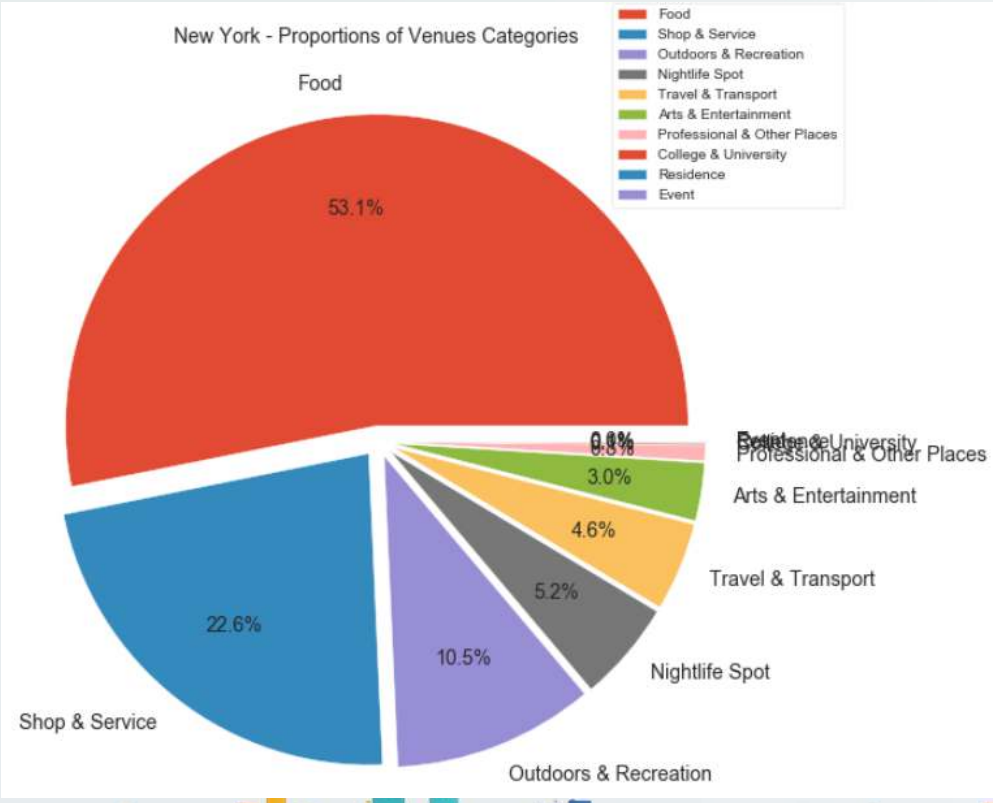
Toronto City Categories

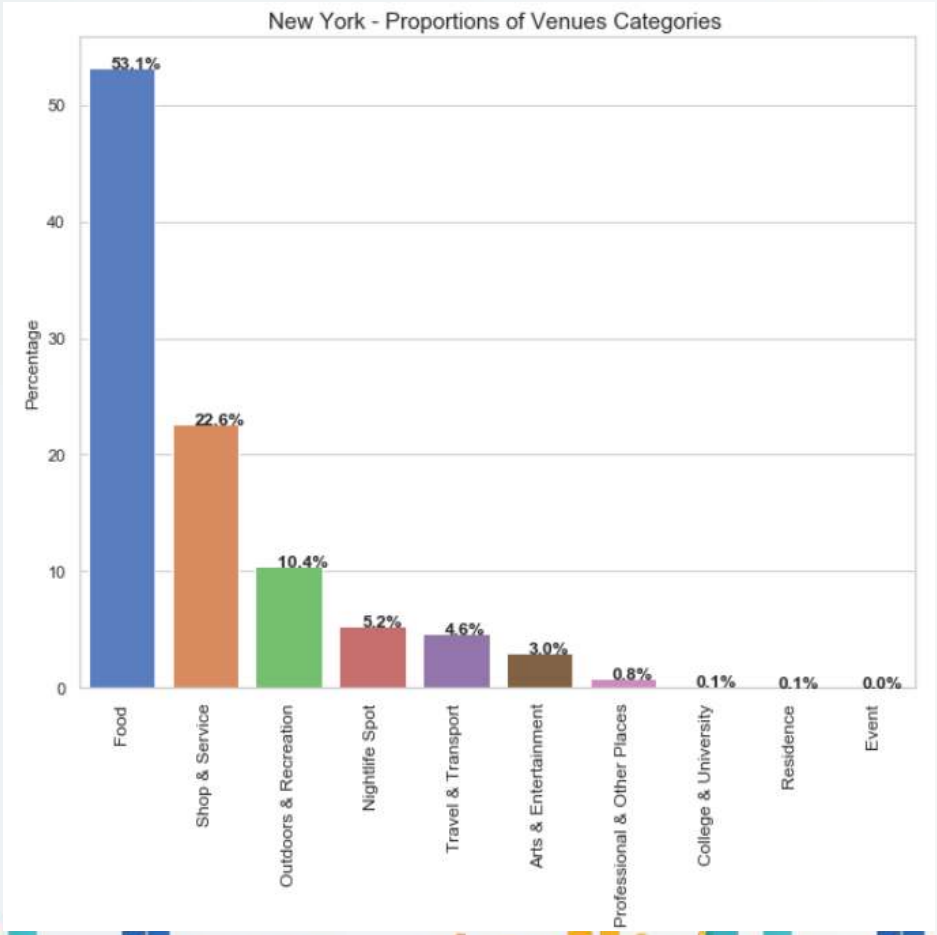




Category Diversity Analysis

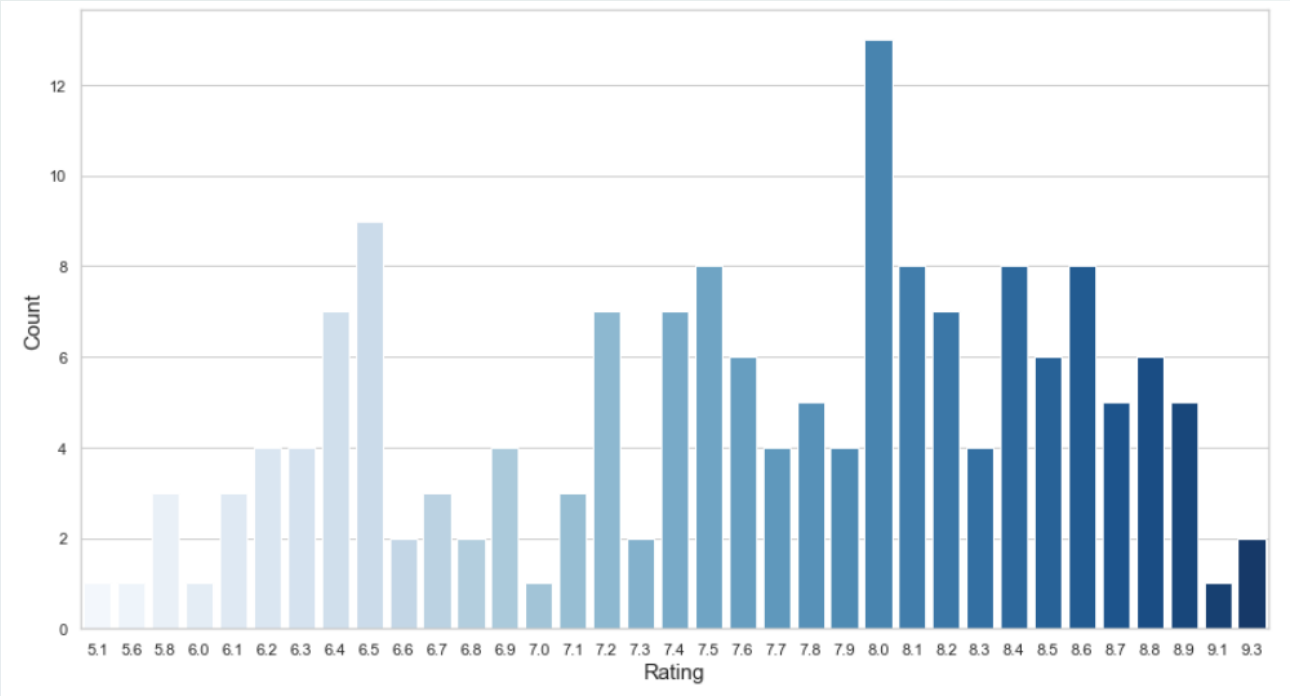
New York City Categories





Detailed Venues Analysis (Rating, Likes, Tips, Photos)

Toronto City

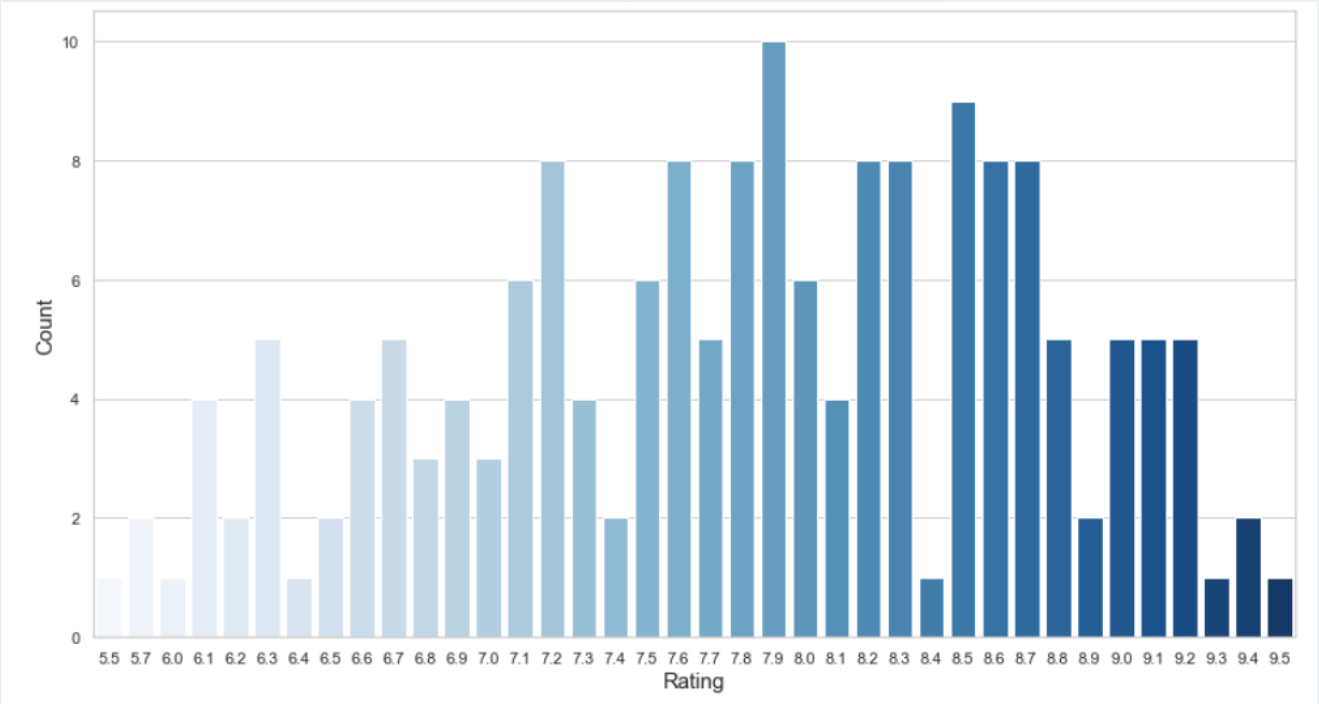



```
Toronto City  | DF_Dimensions:(164, 12) :

 Rating Average: 7.604878048780487
 Tip Count Average: 22.23170731707317
 Likes Count Average: 75.6829268292683
 Photos Count Average: 120.0609756097561
```

Detailed Venues Analysis (Rating, Likes, Tips, Photos)

New York City




```
New York City City  | DF_Dimensions:(172, 12) :

 Rating Average: 7.796511627906975
 Tip Count Average: 21.517441860465116
 Likes Count Average: 72.79651162790698
 Photos Count Average: 63.11046511627907
```

Conclusion

•New York City is the most populous city in venues than Toronto. New York has a higher density of venues per square kilometer, 23 Venues / Km², while the city of Toronto has a venues density value of 2.6 Venues / Km². In New York City you will find 10 times more venues than in Toronto.

•The city of New York presents a greater diversity of categories among its locations compared to the city of Toronto. And both cities present similar proportions of the largest categories and we could verify this with the grouping by clusters.

•The city of New York has a slightly higher average rating value for locations, while the city of Toronto has an average rating of approximately twice that of NY for Photos per location.