Analyzing the neighborhoods for startup idea

Bhushan Hatwalne

April 10, 2019

1. Background Discussion

A startup or start-up is started by individual founders or entrepreneurs to search for a repeatable and scalable business model. More specifically, a startup is a newly emerged business venture that aims to develop a viable business model to meet a marketplace need or problem. Founders design startups to effectively develop and validate a scalable business model. Hence, the concepts of startups and entrepreneurship are similar. However, entrepreneurship refers all new businesses, including self-employment and businesses that never intend to grow big or become registered, while startups refer to new businesses that intend to grow beyond the solo founder, have employees, and intend to grow large. Startup culture has been popular especially in the last few years and almost 80% of all the startups that are launched turn out to be successful. So, in this project we are going to explore the neighborhoods of two most important cities in the world .i.e. New York City, USA and Toronto, Canada and predict the best neighborhood for beginning a startup.

2. Business Problem

Selecting the appropriate neighborhood strongly determines the success of a startup, so even though a person can start his business anywhere in both of the cities, however he might not be quite certain that his business will ultimately bear fruits. So, in this project we will be analyzing the datasets of the neighborhoods for both New York City and Toronto to find the most common and trending places among the various neighborhoods and depending on the most common places around, we can suggest the startup idea that is most likely to be successful in the

vicinity of that place. While predicting the startup idea, we will also be looking into the neighborhoods of both Toronto and NY and determine the similarity among the two places.

3. Datasets For the Problem

We will be using the following datasets for solving the business problem.

The Manhattan dataset

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

The Toronto Dataset

	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

In both the datasets, we are getting the neighborhoods of both the cities along with the latitudes and longitudes for each of them.

4. Methodology

After cleaning both the datasets, we ultimately get the latitudes, longitudes as well as the top 10 most common venues for each of the neighborhoods and then we apply the k-means classification algorithm in order to classify the venues into clusters.

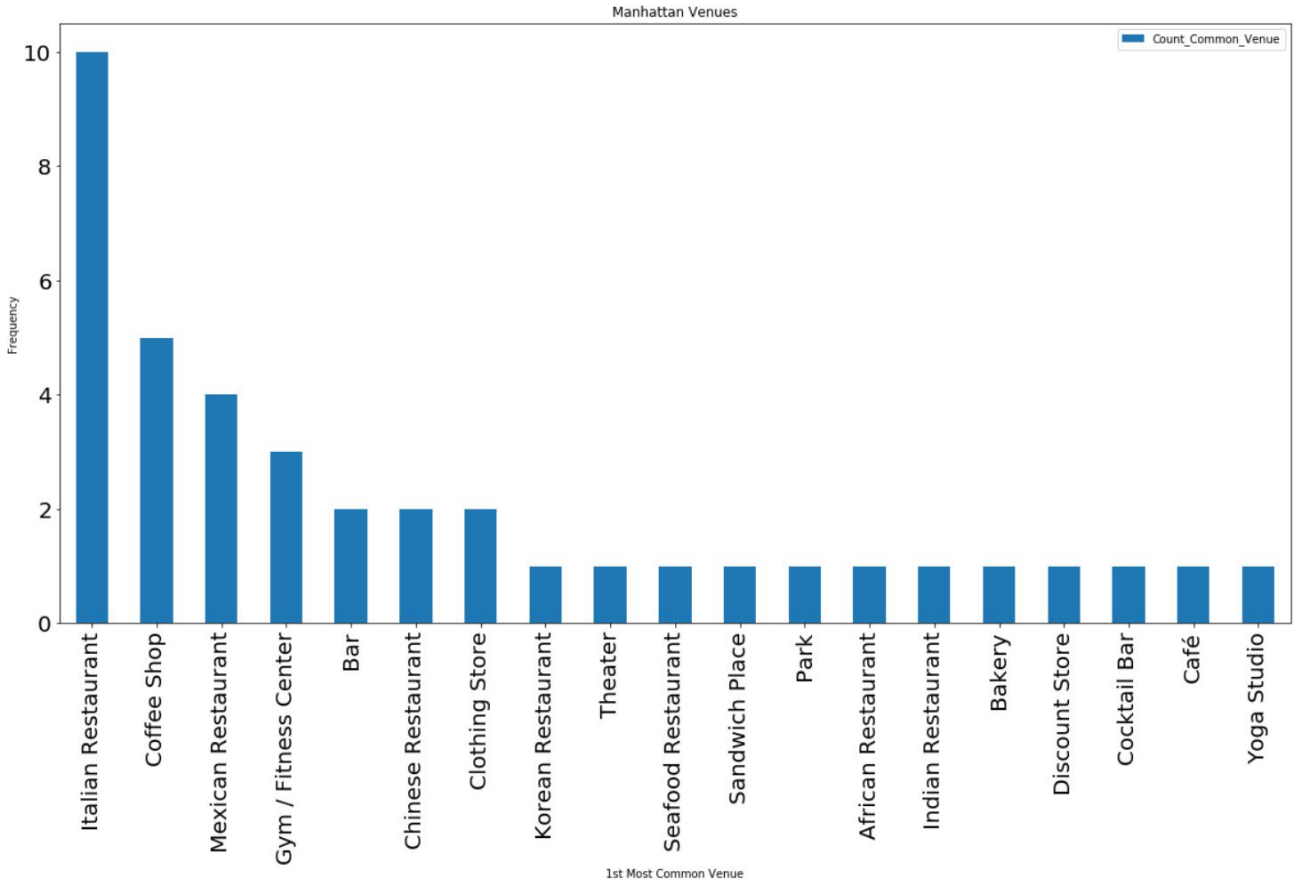
The algorithm is then used to assign labels to each of the clusters and finally, the 1st most common venue for each of the neighborhood is appended to a new dataset.

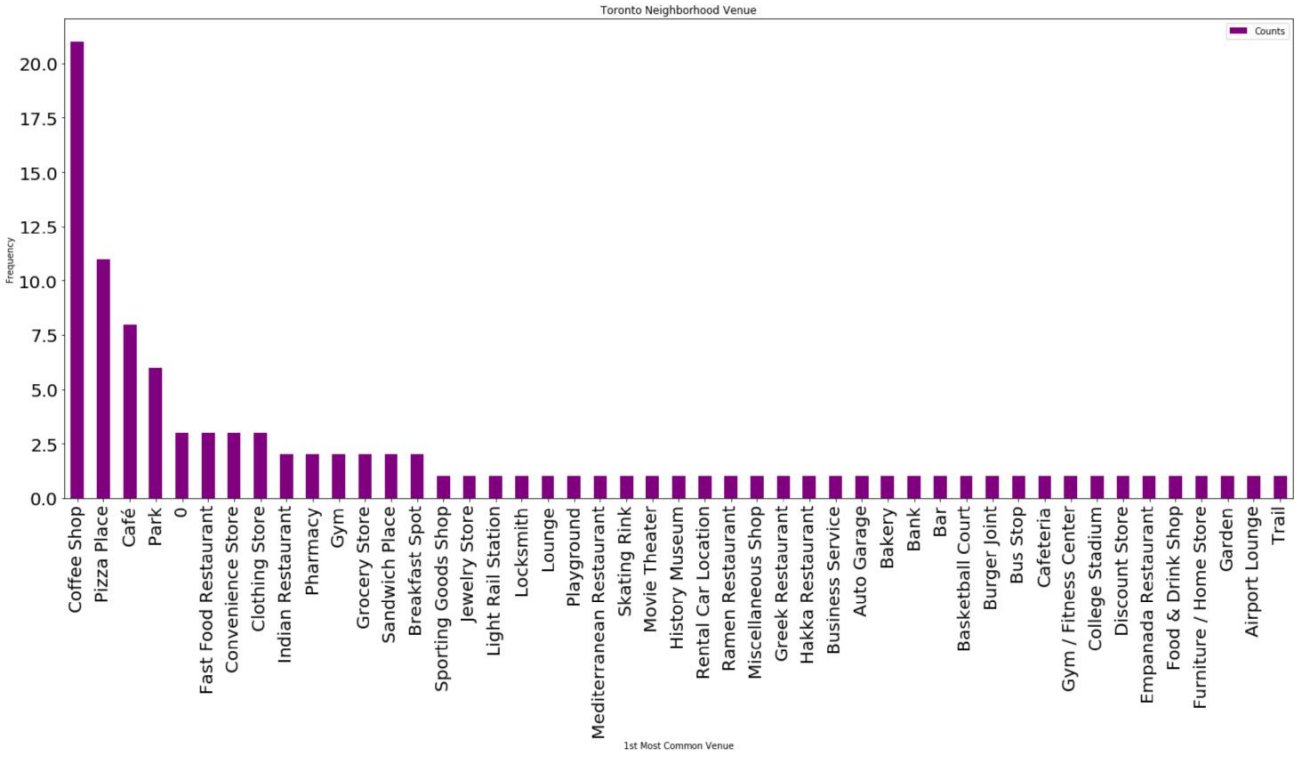
5. Results

The data is grouped on the basis of the 1st most common venue and then we count the top 4 most common venues among all the neighborhoods.

Furthermore, we plot the most common venues on a bar chart for visualizing the same. After plotting the same, we find the neighborhoods corresponding to the top 4 common

venues for each of the city in order to predict that what is the startup that can be built and in which location of the city.





The above two histograms shows the distribution of the most common venues in both the cities.

6. Discussions

Based on the data that we have visualized above, the neighborhoods corresponding to the venues are extracted from both the datasets. For the Toronto Dataset,

```
----------------------------------------------------
The nearest neighborhood to the Coffee Shop is :
['Woburn' 'Bathurst Manor, Downsview North, Wilson Heights' 'The Beaches'
 'North Toronto West'
 'Deer Park, Forest Hill SE, Rathnelly, South Hill, Summerhill West'
 'Cabbagetown, St. James Town' 'Church and Wellesley'
 'Harbourfront, Regent Park' 'St. James Town' 'Berczy Park'
 'Central Bay Street' 'Adelaide, King, Richmond'
 'Harbourfront East, Toronto Islands, Union Station'
 'Design Exchange, Toronto Dominion Centre'
 'Commerce Court, Victoria Hotel' 'Bedford Park, Lawrence Manor East'
 'The Annex, North Midtown, Yorkville' 'Stn A PO Boxes 25 The Esplanade'
 'First Canadian Place, Underground city' "Queen's Park"
 'Canada Post Gateway Processing Centre']
----------------------------------------------------
The nearest neighborhood to the Pizza Place is :
['Guildwood, Morningside, West Hill'
 "Clarks Corners, Sullivan, Tam O'Shanter" 'Willowdale West'
 'Victoria Village' 'Woodbine Gardens, Parkview Hill' 'Glencairn'
 'The Junction North, Runnymede' 'Alderwood, Long Branch'
 'Bloordale Gardens, Eringate, Markland Wood, Old Burnhamthorpe'
 'Westmount'
 'Kingsview Village, Martin Grove Gardens, Richview Gardens, St. Phillips']
----------------------------------------------------
The nearest neighborhood to the Café is :
['Bayview Village' 'Don Mills North' 'Studio District'
 'Harbord, University of Toronto'
 'Chinatown, Grange Park, Kensington Market' 'Christie'
 'High Park, The Junction South' 'Runnymede, Swansea']
----------------------------------------------------
The nearest neighborhood to the Park is :
['York Mills West' 'East Toronto' 'Lawrence Park' 'Rosedale'
 'Caledonia-Fairbanks' 'The Kingsway, Montgomery Road, Old Mill North']
----------------------------------------------------
```

For the Manhattan Dataset,

```
----------------------------------------------------
The nearest neighborhood to the Italian Restaurant is :
['Upper East Side' 'Yorkville' 'Lenox Hill' 'Upper West Side'
 'Greenwich Village' 'Tribeca' 'West Village' 'Noho' 'Turtle Bay'
 'Hudson Yards']
----------------------------------------------------
The nearest neighborhood to the Coffee Shop is :
['Murray Hill' 'Chelsea' 'Morningside Heights' 'Financial District'
 'Carnegie Hill']
----------------------------------------------------
The nearest neighborhood to the Mexican Restaurant is :
['Inwood' 'Hamilton Heights' 'East Harlem' 'Tudor City']
----------------------------------------------------
The nearest neighborhood to the Gym / Fitness Center is :
['Lincoln Square' 'Civic Center' 'Sutton Place']
----------------------------------------------------
```

So, from the extracted neighborhoods and the venues for both the cities, we can see that **"**Coffee Shop**"** is the best and the most common venue for the people residing in both the cities as it accounts to the 1st most common venue in Toronto and 2nd most common venue in Manhattan.

Since, the nearest neighborhoods to the Coffee shop in both the cities are seen in the above figures, so the best startup idea would be to open a **"**Coffee Shop**"** in either one of the neighborhoods as shown above.

7. Conclusion

In this study, I analyzed the neighborhoods for the cities of Manhattan, New York and Toronto, Canada to determine the most common venue in both of these cities and then find the neighborhoods corresponding to these venues and hence, was ultimately able to recommend the best startup idea and also recommend the neighborhood which is best feasible for the business to grow.