Comparison of Neighborhoods of New York and Toronto

INTRODUCTION:

New York and Toronto are both most populous cities in the Unites States and Canada respectively. New York - the most densely populated city in the US, a Global power city which has been described as the Cultural, Financial and media capital of the World. On the other hand, Toronto is also a global city which lies as the centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.

Both being the populous cities, tourism world is always interested in these places. Lets not get deeper into the tourism world but will analyze the attractive venues in and around the neighborhood of both cities to get to know well about the places.

With this analysis we will get to know which one is better over the other or which city has more venues in nearby distance and details on those venues in which one may be interested to have a visit. We will get to know if both cities are similar with respect to these venues or totally different!!!.

DATA:

We need to follow first 3 basic steps here

1. What Data is required?
2. From where to Collect the Data?
3. Is data collected is sufficient and is in correct format? If not, what next?

What Data is required?

To do the comparison of NY and Toronto Neighborhoods, we first need the neighborhood details of NY and Toronto. All the neighborhoods with their Boroughs and postalcodes to get the proper location of the each neighbourhood. Then the venues in those neighbourhoods. Type of venues, number of venues and details on those venues.

From where to collect the data?

NY and Toronto neighbourhood details are readily available in various websites. We can get the details directly from the trusted websites. Regarding the venues in each neighbourhood, **FourSquare API** comes handy. This has all the venue details in each neighbourhood.

Verify the data

Next step is to verify if the data is sufficient enough to start the analysis. Yes, for our analysis these data is fair enough. Now check the data format. Is it in proper format to start the processing. Partially Yes. Because NY data is in proper format which can be processed. But Toronto data is available in Wikipedia page from where we have to segregate and pick the required neighborhood data alone. So, pre-processing of data is required. Once the pre-processing is done, we will have both dataframes ready to start the processing and analysis.