INTRODUCTION

Business Problems

New York city and Toronto are the business capitals of their respective countries and the problem posted here is to find the best among the two locations to open a restaurant. We will analyse the neighbourhoods in each of the city to find the best neighbourhood in the best city selected to set up the business.

Audiences

The right audience for this work is the people who may want to site or start up a business in the neighbourhooods earlier-mentioned.

The project or solution is important to the stakeholders and other audiences because it provides the best neighbourhood to be considered in citing a business in the city. Besides, it helps the audience or stakeholders in making or reaching decision quickly.

DATA

Source of Data

The dataset or set of data available for this project are:

- Demographics of New York City from U.S. Census of 2000 and the New York Department of City Planning
- Demographics of New York City from wikipedia
- List of Postal Codes of Canada from wikipedia
- Boroughs of New York City from wikipedia
- Demographics of Toronto Neighbourhoods from wikipedia which is in turn taken from 2006 Canadian Census
- Foursquare API Geolocation data

Description of Data

The attributes or characteristics of the data are:

- identity numbers,
- longitudes,
- latitudes,
- boroughs,
- neighbourhoods,
- venue,
- categories,
- population,
- average income,
- gross domestic product,
- tips etc.

The latitude is the line running east and west of the earth measured in numeric while the Longitude is an imaginary line running north and south of Greenwich meridian also measured in numeric. The borough is a subset of a city. The neighbourhood is the community area name.

The target value (variable) or label of the data set is the neighbourhood which is the dependent variable. The other variables like longitudes, latitudes, venues, etc. are the independent (predictor) variables.

METHODOLOGY

Data Analysis (Exploratory Data Analysis) of the Manhattan and the Downtown Toronto

- The data sets of the two neighbourhoods, the Downtown Toronto and the Manhattan were visualized to gain knowledge of the distribution of variables and the neighbourhoods in the Boroughs, the Downtown and the Manhattan, using folium library to see how the neighbourhoods in this borough are spatially distributed.
- The resulting data is subjected to initial step of data analysis – the exploratory data analysis (EDA) to gain insight on how correlated or trending the data are.
- After the data sets have been subjected to exploratory data analysis using Folium, Seaborn and Matplotlib libraries to see how the variables of the data are distributed, trending and insight gained, the data was then subjected to pre-processing stage.

Inferential Statistical Test (Machine Learning) of the Manhattan and the Downtown

- Here the encoded data during the analysis are fed into the machine learning algorithm - K-mean Clustering - to group or segment the neighbourhoods.
- At the end, the neighbourhoods that have similar characteristics are grouped together. Then the similarities and dissimilarities of the Downtown and the Manhattan are observed.

DISCUSSION

Observations

- The two neighbourhoods in the Manhattan of the New York City and the Downtown Toronto of the Toronto City - are similar in having banks, grocery, market, farmers market, waterfall, college, transport station and buildings but dissimilar in other venues listed in the python notebook
- These venues that are similar indirectly represent availability of capital made available by bank, raw materials provided by the farmers markets and groceries, water provided by waterfall, labour provided by the colleges and the universities, efficient management provided by the college and the competitors which are in the restaurant businesses in the neighbourhoods, transport facilities provided by the presence of bus or train stations in the neighbourhoods.

Recommendations

Even though there are factors or variables that are not available during the research, the research still recommend that the Manhattan neighbourhoods should be considered in siting a restaurant business if the two boroughs given to be considered are the Manhattan and the Downtown.

CONCLUSION

- In the end, the Manhattan neighbourhoods are better than the Downtown neighbourhoods in siting a restaurant business based on the available data. Why? Though the two neighbourhoods are similar in factor that influence the location of restaurant business yet they are glaringly dissimilar in the population densities and hence the market shares of a potential restaurant business.
- Any potential restaurant businesses sited in the Manhattan will have large market share because of the population or population density of the area than if it is to be sited in the Downtown.