Battle of Neighbourhoods

Chirag Patel March 2020

1 Introduction

1.1 Background

It's trend of globalization and technology. Data is generated at a tremendous rate. Data is a great tool which can be used to solve several problems. Data can be used to predict outcome of certain event which can help humanity to a greater extend. Here also data will be used to generate an investment plan which will help an organization economically.

1.2 Business Problem

An investment group is looking for investing in Two biggest cities of North America, New York and Toronto respectively. They need an investment plan where they can invest their money. They are not sure which business are flourishing in that cities. They need to which neighbourhood would be best for their investment in those cities and which business would be best to invest in.

Data

2.1 Data Sources

Data needed for such business problem can be found on the web. The data for Toronto neighbourhood is available on Wikipedia and can be gathered using data scrapping. Data for New York neighbourhood is available in JSON format and is needed to be converted into tabular. Foursquare API will also be used to get venues by particular neighbourhood.

2.2 Data Cleaning and Formatting

Data for Toronto neighbourhood can be gathered using Beautiful Soup from Wikipedia and needs to be cleaned for further use and data for New York Neighbourhood need to be converted from JSON format to tabular format