RESTAURANTS IN TORONTO

Akshay Mendiratta

May 12, 2019

**1. Introduction**

**1.1 Background**

Food industry is one of the biggest industries in the world. Everyday hundreds of new restaurants open worldwide and hundreds of them close down, because of many factors. Location of a restaurant is one of the key factors which determine its success and failure. A restaurant in a good, well-chosen location will get a large number of customers and will generate more revenue as compared to a restaurant in a bad location. Therefore, it would be advantageous to determine which location would help in the success of a restaurant.

**1.2 Problem**

The problem which stands in front of us is to how to determine the location which would help a restaurant to succeed. Data of locations of various restaurants in a city (in our case the city of Toronto), in which the new restaurant is to be opened might help us in solving this problem.

**1.3 Interest**

Any person who wishes to open a new restaurant would be interested in knowing which location would help him/her in succeeding in the business.

**2. Data Acquisition and Cleansing**

**2.1 Data Acquisition**

The data about the various boroughs of Toronto was collected from a Wikipedia page, which contained the latitude and longitude values of the borough along with their names, postal codes and neighborhoods.

The data about the restaurants of Toronto was collected from the Foursquare website using their API. The website returned a data of various venues out of which we used the data of the restaurants which included the restaurant’s name, category, and, latitude and longitude values

**2.3 Data Cleansing**

The data acquired from the Wikipedia page required to drop some postal codes to which did not have any boroughs assigned to them. Also we had to merge various postal codes.

The data returned by the Foursquare website consisted of different venue out of which we only required the data of restaurants so we dropped all the other data. Also the data of the restaurant had many attributes out of which we only kept the restaurant’s name, category, and, latitude and longitude values.

Our final data looked like this.

