**The battle of Neighborhood**

Swati Pandey August 09, 2019

**1. Introduction**

* 1. **Background**

Since lots of people nowadays find interest in music and parties to get a take away for sometimes from their regular life. So, if a borough with most of the neighborhoods can be selected for the piano bar if it does not exist there, then it can be really beneficial from business perspective as well as for public convenience.

* 1. **Problem**

In this project we will try to find an optimal location for a Bar. Specifically, this report will be targeted to stakeholders interested in opening a Piano Bar in a Borough of Toronto, Ontorio, Canada which has most neighborhoods.

Since there are lots of bars in Totonto we will try to detect locations or clusters that are not already crowded with bars. We are also particularly interested in areas with no Piano Bars in vicinity.

* 1. **Interest**

This report will be targeted to stakeholders interested in opening a Piano Bar in a Borough of Toronto, Ontorio, Canada which has most neighborhoods.

**2. Data acquisition and cleaning**

**2.1 Data sources**

Based on definition of our problem, factors that will influence our decision are:

* number of existing bars in the neighborhood (any type of bar)
* number of Piano brs in the neighborhood, if any

Following data sources will be needed to extract/generate the required information:

* Data of Postal code with Borough and Neighborhoods can be scraped or used as CSV from <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>
* For cordinates of each neighborhood, we can get the data from <https://cocl.us/Geospatial_data>
* number of bars and their type and location in every neighborhood will be obtained using **Foursquare API**
* coordinate of Toronto center will be obtained using geopy

**2.2 Data cleaning**

Data with Toronto Boroughs can be taken as CSV file and then imported in pandas dataframe. Then all the rows having ‘Not Assigned’ values should be analyzed and treated accordingly.

Name/Rename all the columns correctly. Data for the latitude and longitude of all the postal codes can be taken from the link mentioned in ‘data sources’ section and then it should be cleaned properly. Then merge the data of the two dataframes to get the final dataframe to work upon.