**Predicting the suitable locations for Opening a New Indian Restaurant in Manhattan**

**Rajan Bhatia**

**10 Feb 2020**

# **Introduction**

* 1. **Background**

Opening a new restaurant needs lot of planning with respect to the locations that have favourable venues that will ensure there is sufficient demand to turn the business into a profitable venture. This requires good understanding of the local neighbourhood, venues nearby in the areas where Indian Restaurants have been successful. Theare are various categories of venues like schools, food joints, malls, parks etc where we can expect visitors to come. Such locations could provide good opportunity for the catering and restaurant business. It is important to analyse which of the neighbour hoods have more Indian restaurants, factors encouraging Indian restaurants, and then analysing those areas which have less or no Indian restaurants but have similar favouring venues that could potentially a business opportunity for opening Indian Restaurants.

* 1. **Problem**

This planning requires access to data of various neighbourhoods and the venues in each neighbourhood. Once the data is accessible, it requires analysis and co-relation of various venues where Indian restaurants have flourished. The project aims to study the neighbourhoods of Manhattan and the venues around these, and identify the pattern or similarity of venues that are observed in these areas with higher presence of Indian Restaurants. We would need to predict and recommend potential neighbourhoods to open Indian Restaurants based on the analysis of data.

* 1. **Interest**

It would be of interest to Indian entrepreneurs who want to invest money in opening Indian Restaurants. Also this would be of interest to existing Restaurant chains for expanding their business in other neighbourhoods.

1. **Data Acquisition and Cleaning**
   1. **Data Sources**

There are two data sources that have been utilised for the project :

1. The Borough and Neighbourhood Data (<https://cocl.us/new_york_dataset>)
2. The Venues data from FourSquare

For this I have registered with FourSquare in their developer environment and utilised their public APIs to access the Venue data.

* 1. **Data Cleaning**

The Neighbourhood data contained data for many Boroughs. From this, the neighbourhoods of Manhattan were filtered so that we could focus on its neighbourhoods.

The venues were fetched from FourSquare website, within the distance range of 500m, with a limit of 100 venues.

Extracted the venue categories from the data. The data for neighbourhoods and venues were combined.

The venue categories were converted in columns and measure of frequency of each category was taken for each neighbourhood. Onehot encoding was done and means were calculated.

* 1. **Feature Selection**

The venue categories were taken as the Features that would help in identifying what are the neighbourhoods that have potential for Indian Restaurants. Based on these categories. There were 40 neighbourhoods and 335 venue categories as features.

In order to evaluate all features, all were included in the modelling, testing an predictions.