**Capstone Project**

**Recommending Locations for starting a Indian Restaurant in Manhattan , New York**

Avinash Rao

October 15,2019

**Introduction**

Going to Restaurants is supposed to be one of the most loved activities worldwide.

Everyone would love to visit his or her favourite restaurants every now and then.

Food aficionados enjoy exploring new restaurants ,Going to a restaurant could be a

great family outing and everyone loves to spend quality time with friends and family

at a restaurant ,eating good food amidst a warm and friendly ambiance. Generally,

people love to go back to a restaurant again and again because of the food, the

service, the aesthetic ambiance and of course, the x-factor which could be from

chefs to simply the exclusivity

**Business Problem**

In this project we will try to find an optimal location for a restaurant. Specifically, this

report will be targeted to stakeholders interested in opening an Indian restaurant in

Manhattan ,New York.

Since there are lots of restaurants in New York we will try to detect locations that are

not already crowded with restaurants. We are also particularly interested in areas

with no Indian restaurants.

**Target Audience**

This project focuses on Business mans ,People who are willing to start a new Indian

restaurants in Manhattan ,New York . As there are many Indians and Asian people

live in New York Starting a Indian Restaurant at locations with no Indian

Restaurants will attract people and less competition from other Restaurants which

result in success of Restaurant.

**Data**

**Data Required**

* We require data of Neighbourhoods of Manhattan the most densely populated
* of the five boroughs of New York.
* We require geographical coordinates of Neighbourhoods of Manhattan
* We require list of Restaurants in each Neighbourhood

**Data sources needed to extract/generate the required information**:

* The json file <https://cocl.us/new_york_dataset> contains details about all
* boroughs and Neighbourhoods of New York.
* We will use Geopy package for getting Latitude and Longitude of Manhattan
* and New York .
* We use FoureSquare API for getting all venue details of nearby
* Neighbourhoods of Manhattan.

**Methodology**

First ,we need to Extract Neighbourhood data from https://cocl.us/new\_york\_dataset

Json file. After extracting Neighbourhoods of New York we need to create new data

frame consisting list of Neighbourhoods of New York. Now, we need to filter the

dataset so that final dataset should contain list of Neighbourhoods of Manhattan but

not New York .After cleaning data we need to use Geocoder package to get the

latitude and longitude values of Manhattan for visualization.

Now we need to get top venues in each Neighbourhood using Foursquare. After

getting venues we need to group them by Neighbourhoods and take mean of the

frequency of occurrence of each category then we need to filter Venues by Indian

Restaurants. Now, we need to perform clustering methods to cluster

Neighbourhoods With similar properties or similar number of Indian Restaurants.

We are using K-means algorithm which is a kind of clustering algorithm we define k

as 3 which is essential parameter which means number of clusters or number of

categories that the data will be divided into.

After clustering now we can say which Neighbourhoods have less or moderate or

nigh number of Indian Restaurants based on that we can decide which location or

Neighbourhood is suitable for opening a new Restaurant. We can use folium to

visualize our clusters on map

**Result**

* Clustering data into 3 Categories
* Cluster 0:Neighbourhoods with no Indian Restaurants.
* Cluster 1:Neigbourhoods with High number of Indian Restaurants.
* Cluster 2:Neighbourhood with moderate number of Indian Restaurants.

A close up of a map

Description automatically generated

Visualised Map after clustering Neighbourhoods based on number of Indian

Restaurants. Neighbourhood in Cluster 0 is represented by Red colour,

Neighbourhoods in Cluster 1 is represented by Purple colour,

Neighbourhoods in Cluster 2 is represented by Light Green Colour.

Discussion

From above map and data we can say that Neighbourhoods in Cluster 0 have no

Indian Restaurants so these locations are very suitable for opening new Indian

Restaurants as there are no Indian Restaurants for competition, Neighbourhoods in

Cluster 1 have high number of Indian Restaurants so these locations are highly not

suitable for opening new Restaurant as it may result in loss as there are many

Restaurants are already present. Neighbourhoods in Cluster 2 have moderate

number of Indian Restaurants. We can also observe that Neighbourhoods in east

Manhattan have low to none Indian Restaurants and Neighbourhoods in west of

Manhattan have high concentration on Indian Restaurants.

Conclusion

In this project we collected data about Neighbourhood venues of Manhattan.

By using unsupervised Machine Learning model k-means we clustered

Neighbourhoods into 3 groups or clusters. By analysing clusters we can conclude

that Neighbourhoods in Cluster 0 are suitable or ideal for starting new Indian

Restaurant. The findings of project will help Business man or people who are

interested in opening a Indian Restaurant