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

**Rajan Bhatia**

**13 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.

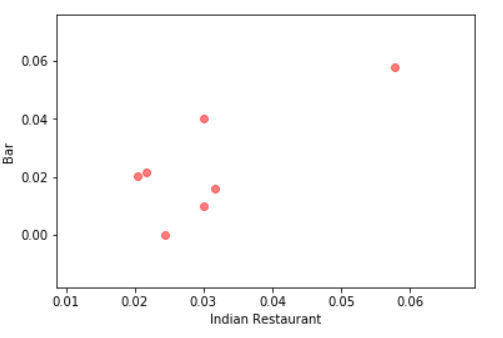
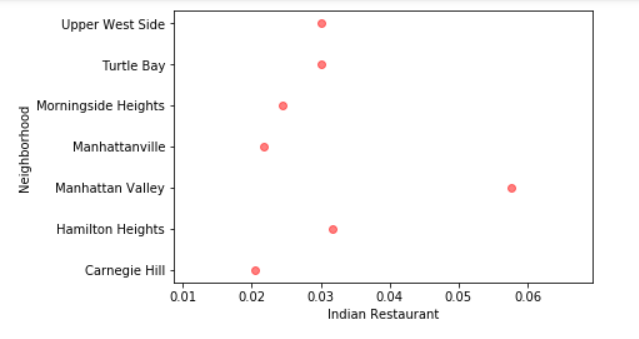
* 1. **Feature Selection**

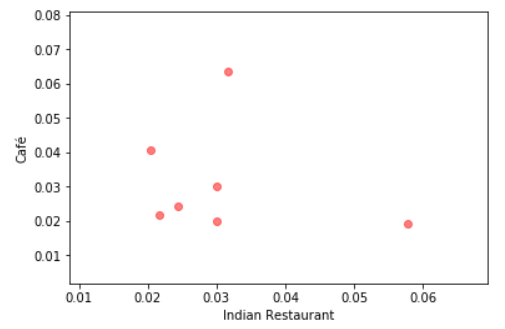
The venue categories were taken as the Features that would help in identifying what are the neighbourhoods that have potential for Indian Restaurants. Based on these categories. There were 40 neighbourhoods and 335 venue categories as features.

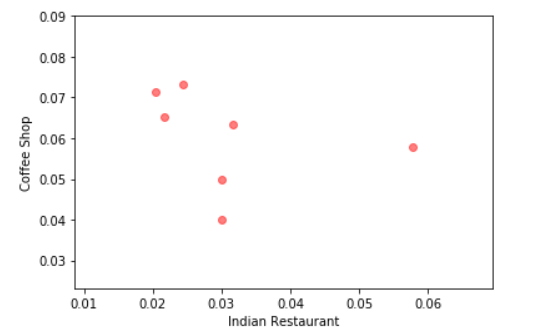
In order to evaluate all features, all were included in the modelling, testing an predictions.

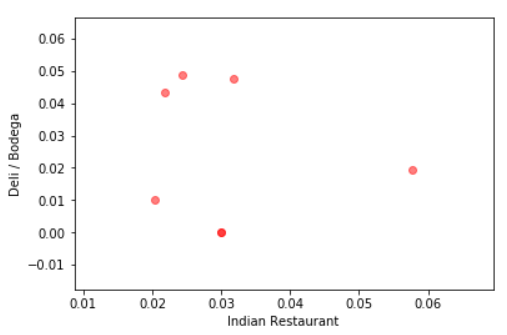
1. **Exploratory Data Analysis**
   1. **Calculation of Target Variable**

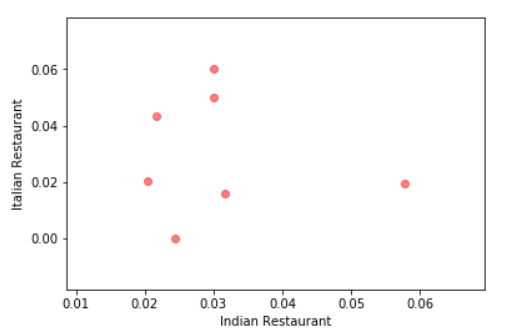
The presence of Indian Restaurant in neighbourhood would be the target output to be predicted based on the surrounding other venues. Positive cases are marked as having mean presence more than 0.2 in a neighbourhood. All neighbourhoods with Indian Restaurant mean > 0 are filtered nd split into 20% Test and 80% Train data. Once the model is trained, it is executed on the set of neighbourhoods with Indian Restaurant = 0, and the output indicates the potential areas favouring setup of Indian Restaurants. The below graphs show the distribution of Indian Restaurants across neighbourhoods and across various other nearby venues. This is to provide a view on favourable venues around which Indian Restaurants are generally coming up - so that we can then try to find the similar patterns in other locations.

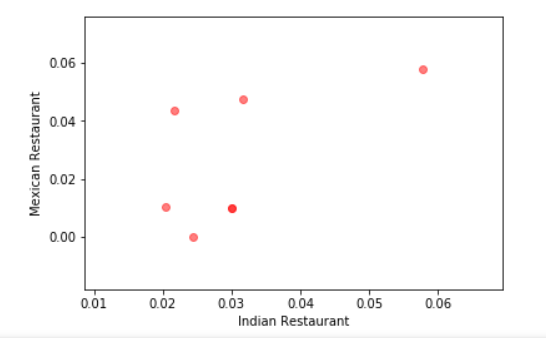


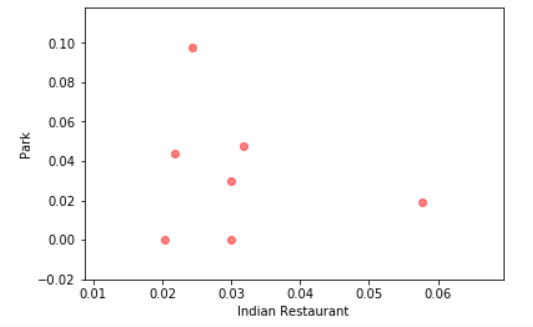


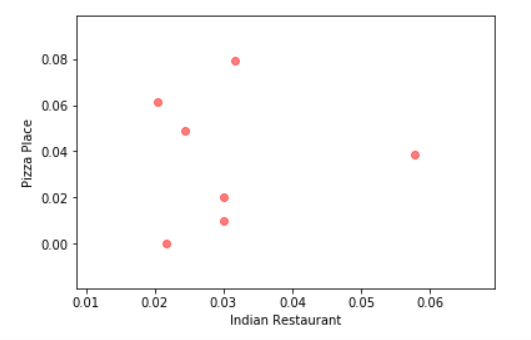


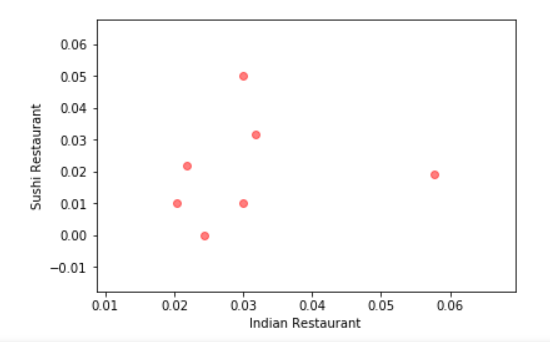












1. **Predictive Modelling**

Classification model was used to predict the possibilities of setting up Indian Restaurants in the neighbourhoods. In the assignment,

\* The data set is filtered where Indian Restaurants are present (Mean !=0)

\* Test data is set to 20%, Train data is set to 80%

\* The model is trained for K = 1 to 10

\* Model is tested on Test data, and accurany score is calculated

K=10, is taken as the chosen best score for actual prediction on the remaining data set (Indian Restaurant mean = 0)

Tudor City and Inwood were found to be the closest match with the neighbourhoods that had similar venues where Indian Restaurants are doing business.

The graph below shows the similarity in the venues across the cities along with Tudor and Inwood.



1. Conclusions

In this study, the venue patterns were analysed across the neighbourhoods, where Indian Restaurants are doing business. These patterns were identified by KNN regression model. This model was then used to identify more similar neighbourhoods where Indian Restaurants can be targeted. The data related to ethnicity of residents in these neighbourhoods was not available - that could be an influential factor to decide about the suitable location for opening Indian Restaurants.

**Appendix**

For details the individual score for the Inwood and Tudor City are enclosed, as generated by the model. Clearly it shows that Tudor City is the best match recommended.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Neighborhood** | **Carnegie Hill** | **Hamilton Heights** | **Manhattan**  **Valley** | **Manhattanville** | **Morningside Heights** | **Turtle Bay** | **Upper West Side** | **Inwood** | **Tudor City** |
| **American Restaurant** | 0.0102041 | 0 | 0 | 0.0217391 | 0.0731707 | 0.01 | 0.02 | 0.0344828 | 0.0119048 |
| **Asian Restaurant** | 0 | 0 | 0 | 0 | 0 | 0.02 | 0.01 | 0 | 0.0238095 |
| **Bagel Shop** | 0.0102041 | 0 | 0 | 0 | 0 | 0 | 0.01 | 0 | 0.0119048 |
| **Bakery** | 0.0306122 | 0.031746 | 0.0192308 | 0 | 0 | 0.01 | 0.03 | 0.0344828 | 0 |
| **Bank** | 0.0102041 | 0.015873 | 0 | 0.0217391 | 0 | 0 | 0 | 0 | 0.0119048 |
| **Bar** | 0.0204082 | 0.015873 | 0.0576923 | 0.0217391 | 0 | 0.01 | 0.04 | 0.0172414 | 0.0119048 |
| **Bistro** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0172414 | 0 |
| **Boxing Gym** | 0 | 0 | 0 | 0 | 0 | 0.01 | 0 | 0 | 0.0119048 |
| **Bridge** | 0 | 0 | 0 | 0 | 0 | 0.01 | 0 | 0 | 0.0119048 |
| **Burger Joint** | 0.0102041 | 0.015873 | 0 | 0 | 0.0487805 | 0 | 0 | 0 | 0.0119048 |
| **Bus Station** | 0 | 0 | 0 | 0.0217391 | 0 | 0 | 0 | 0.0172414 | 0 |
| **Café** | 0.0408163 | 0.0634921 | 0.0192308 | 0.0217391 | 0.0243902 | 0.02 | 0.03 | 0.0517241 | 0.0595238 |
| **Caribbean Restaurant** | 0 | 0.031746 | 0.0192308 | 0 | 0 | 0 | 0 | 0.0172414 | 0 |
| **Chinese Restaurant** | 0.0102041 | 0.031746 | 0.0192308 | 0.0434783 | 0 | 0 | 0.01 | 0.0344828 | 0 |
| **Coffee Shop** | 0.0714286 | 0.0634921 | 0.0576923 | 0.0652174 | 0.0731707 | 0.05 | 0.04 | 0.0172414 | 0.0357143 |
| **Convenience Store** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0119048 |
| **Deli / Bodega** | 0.0102041 | 0.047619 | 0.0192308 | 0.0434783 | 0.0487805 | 0 | 0 | 0.0344828 | 0.0357143 |
| **Diner** | 0 | 0 | 0 | 0.0217391 | 0 | 0.01 | 0.01 | 0.0172414 | 0.0357143 |
| **Dog Run** | 0 | 0 | 0.0192308 | 0 | 0 | 0.01 | 0.01 | 0.0172414 | 0.0238095 |
| **Farmers Market** | 0 | 0 | 0 | 0 | 0.0243902 | 0.01 | 0 | 0.0172414 | 0 |
| **Fast Food Restaurant** | 0.0102041 | 0.015873 | 0 | 0.0217391 | 0 | 0 | 0 | 0.0172414 | 0 |
| **French Restaurant** | 0.0306122 | 0 | 0.0192308 | 0 | 0 | 0.03 | 0 | 0 | 0.0119048 |
| **Frozen Yogurt Shop** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0344828 | 0 |
| **Garden** | 0 | 0 | 0 | 0 | 0 | 0.02 | 0.01 | 0 | 0.0238095 |
| **Gift Shop** | 0.0102041 | 0 | 0 | 0 | 0 | 0.02 | 0.01 | 0 | 0.0119048 |
| **Greek Restaurant** | 0 | 0 | 0 | 0 | 0.0243902 | 0.02 | 0.01 | 0 | 0.0357143 |
| **Grocery Store** | 0.0204082 | 0 | 0.0192308 | 0 | 0.0243902 | 0.01 | 0 | 0.0172414 | 0 |
| **Gym** | 0.0306122 | 0.015873 | 0 | 0 | 0 | 0.01 | 0.01 | 0 | 0.0119048 |
| **Gym / Fitness Center** | 0.0204082 | 0 | 0.0192308 | 0 | 0 | 0 | 0.02 | 0 | 0.0119048 |
| **Hawaiian Restaurant** | 0 | 0 | 0.0192308 | 0 | 0 | 0 | 0 | 0 | 0.0119048 |
| **Heliport** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0119048 |
| **History Museum** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0172414 | 0 |
| **Hotel** | 0.0102041 | 0 | 0 | 0 | 0 | 0.02 | 0 | 0 | 0.0119048 |
| **Ice Cream Shop** | 0 | 0 | 0.0192308 | 0 | 0.0243902 | 0 | 0.02 | 0.0172414 | 0 |
| **Italian Restaurant** | 0.0204082 | 0.015873 | 0.0192308 | 0.0434783 | 0 | 0.05 | 0.06 | 0 | 0.0119048 |
| **Japanese Restaurant** | 0.0306122 | 0.015873 | 0.0192308 | 0 | 0 | 0.03 | 0.01 | 0 | 0.0119048 |
| **Jewish Restaurant** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0119048 |
| **Juice Bar** | 0 | 0 | 0 | 0.0217391 | 0 | 0 | 0.01 | 0.0172414 | 0 |
| **Latin American Restaurant** | 0 | 0.015873 | 0.0192308 | 0 | 0 | 0 | 0 | 0.0172414 | 0.0119048 |
| **Lingerie Store** | 0 | 0 | 0 | 0 | 0 | 0 | 0.01 | 0 | 0.0119048 |
| **Liquor Store** | 0 | 0.015873 | 0 | 0.0217391 | 0 | 0 | 0 | 0 | 0.0119048 |
| **Lounge** | 0 | 0 | 0 | 0.0217391 | 0 | 0.01 | 0 | 0.0517241 | 0.0119048 |
| **Mexican Restaurant** | 0.0102041 | 0.047619 | 0.0576923 | 0.0434783 | 0 | 0.01 | 0.01 | 0.0689655 | 0.0595238 |
| **Moving Target** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0172414 | 0 |
| **Office** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0119048 |
| **Park** | 0 | 0.047619 | 0.0192308 | 0.0434783 | 0.097561 | 0.03 | 0 | 0.0344828 | 0.0595238 |
| **Pet Store** | 0.0102041 | 0 | 0 | 0 | 0 | 0.01 | 0.01 | 0.0172414 | 0.0119048 |
| **Pharmacy** | 0 | 0 | 0 | 0 | 0.0243902 | 0 | 0 | 0.0344828 | 0.0119048 |
| **Pizza Place** | 0.0612245 | 0.0793651 | 0.0384615 | 0 | 0.0487805 | 0.01 | 0.02 | 0.0517241 | 0.047619 |
| **Playground** | 0.0102041 | 0.015873 | 0.0192308 | 0.0217391 | 0 | 0 | 0 | 0.0172414 | 0 |
| **Plaza** | 0 | 0 | 0.0192308 | 0 | 0 | 0.02 | 0 | 0 | 0.0119048 |
| **Restaurant** | 0.0102041 | 0 | 0 | 0 | 0 | 0.01 | 0.02 | 0.0517241 | 0.0238095 |
| **Salad Place** | 0.0102041 | 0 | 0 | 0 | 0.0243902 | 0 | 0.01 | 0 | 0.0119048 |
| **Salon / Barbershop** | 0.0102041 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0119048 |
| **Sandwich Place** | 0 | 0.031746 | 0 | 0 | 0.0487805 | 0.01 | 0 | 0 | 0.0238095 |
| **Seafood Restaurant** | 0 | 0.015873 | 0 | 0.0434783 | 0.0243902 | 0.02 | 0.02 | 0.0172414 | 0.0119048 |
| **Shanghai Restaurant** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0119048 |
| **Spa** | 0 | 0 | 0.0192308 | 0.0217391 | 0.0243902 | 0.01 | 0 | 0 | 0.0119048 |
| **Spanish Restaurant** | 0 | 0.015873 | 0 | 0.0217391 | 0 | 0.01 | 0 | 0.0344828 | 0.0238095 |
| **Steakhouse** | 0 | 0 | 0 | 0 | 0 | 0.05 | 0 | 0.0172414 | 0 |
| **Supermarket** | 0.0102041 | 0 | 0 | 0.0217391 | 0.0243902 | 0 | 0 | 0.0172414 | 0 |
| **Sushi Restaurant** | 0.0102041 | 0.031746 | 0.0192308 | 0.0217391 | 0 | 0.05 | 0.01 | 0 | 0.0238095 |
| **Taco Place** | 0.0102041 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0119048 |
| **Tattoo Parlor** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0172414 | 0 |
| **Tennis Court** | 0 | 0 | 0 | 0 | 0.0243902 | 0.01 | 0 | 0 | 0.0119048 |
| **Thai Restaurant** | 0 | 0 | 0.0192308 | 0 | 0 | 0.02 | 0.02 | 0 | 0.0238095 |
| **Trail** | 0 | 0 | 0 | 0 | 0 | 0.01 | 0.01 | 0 | 0.0119048 |
| **Veterinarian** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0172414 | 0 |
| **Vietnamese Restaurant** | 0.0204082 | 0 | 0.0192308 | 0 | 0 | 0 | 0.01 | 0 | 0.0119048 |
| **Wine Bar** | 0.0102041 | 0.015873 | 0 | 0 | 0 | 0.04 | 0.04 | 0.0344828 | 0 |
| **Wine Shop** | 0.0306122 | 0 | 0.0192308 | 0 | 0 | 0 | 0.01 | 0.0172414 | 0.0119048 |
| **Yoga Studio** | 0.0306122 | 0.031746 | 0.0384615 | 0 | 0 | 0 | 0.02 | 0.0172414 | 0.0119048 |