**Predicting the Rating of Restaurants in Bangalore**

**19 May 2019**

**1. Introduction**

**1.1 Background**

Going to Restaurants is quite popular among Bangaloreans and the turnout is quite high during weekends thus a lot of players have started tapping this potential revenue stream. Whenever someone tries to open a new venue they have a lot fo doubts regarding the footfall in the area, spending habits of people etc.

**1.2 Problem**

If the person who wants to open a new restaurant could get to know the amount ppl spend on avg in the restaurants in the vicinity/neighbourhood it’d be really easy for them to take a conscious decision about the place they would like to open their restaurant. To facilitate this understanding I’ll try to find about the venues in every neighbourhod of Bangalore and their features like cuisines, avg. cost for 2, location,etc and the effect these features produce on the popularity of the restaurants via ratings given by patrons.

**2. Data Acquisition & Cleaning**

**2.1 Data Sources**

Most restaurant’s ratings and features can be found on popular websites/apps like zomato, swiggy, uber eats but making a distinctive dataset out of that information is not easy as they don’t provide rating on whole for every restaurant, we need to give restaurant names one by one and then get the results. What we can do instead is to use Zomato API which returns a list of restaurants for every location when given an input in form of neighbourhood name and the radius from that location. The location of all neighborhoods can be retrived from Foursquare API or geacoders API of Python when fed with the name of the locality. Then names we can get by locating a website with a list of pincodes of Bangalore and the particular name of the localities inside that pincode as such are codes can have multiple societies under one pincode .

**2.2 Data cleaning**

The location coordinates got from the geocoders API are not exactly coorect as some coordinates are wrong as we can see that by plotting them all on a map thus we need to eliminate the rows which have values not corresponding to Bangalore’s coordinated area. We also need to remove any other rows which might not be useful for the analysis as we go on.

**2.3 Feature Selection**

The zomato API returns a JSON with a lot of values which might not be required at all so we’ll take ony these fields :

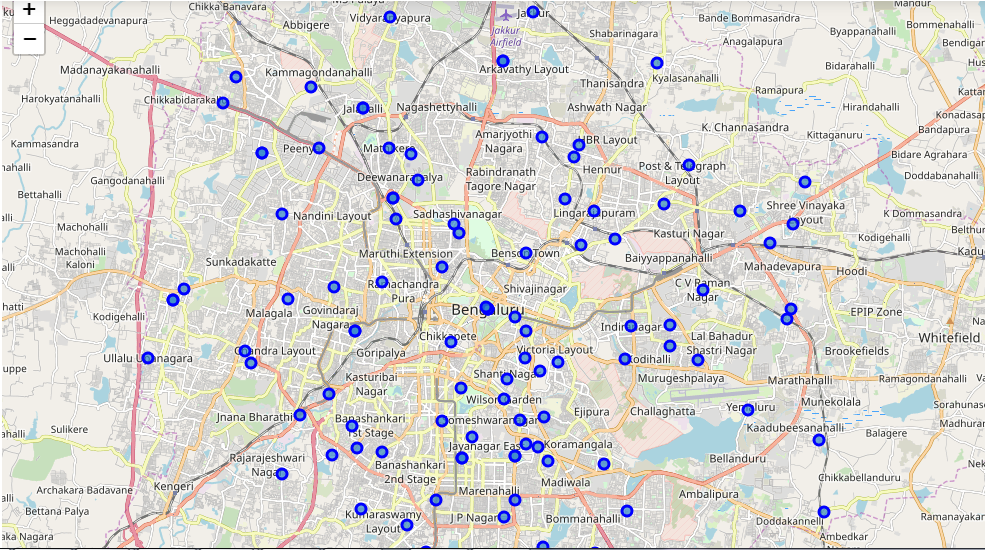
'Neighbourhood','Neighborhood Latitude', 'Neighborhood Longitude', 'restaurant', 'restaurant id', 'Cuisines', 'avg cost for two', 'Rating', 'has online delivery', 'has table booking', 'Total Rest. in area'

While making the list for model training and prediction we would require only 'Neighborhood Latitude', 'Neighborhood Longitude', 'avg cost for two', 'Rating', 'has online delivery', 'has table booking' as we are going to predict based on only these parameters.

**3. Exploratory Data Analysis**

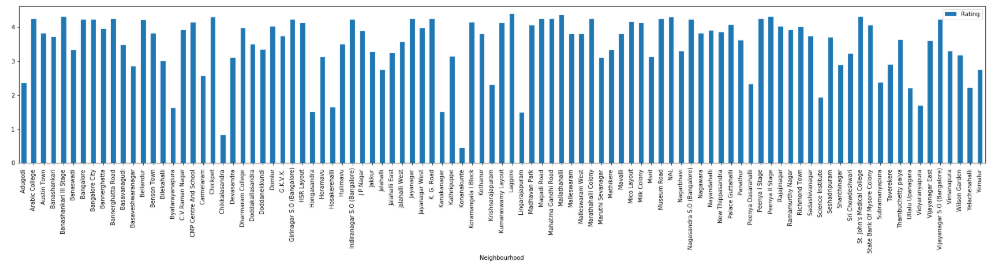
**3.1 Location Coordinates**

I plotted the location coordinates on a map to see where exactly they are and to get a fair idea of the neighbourhoods.



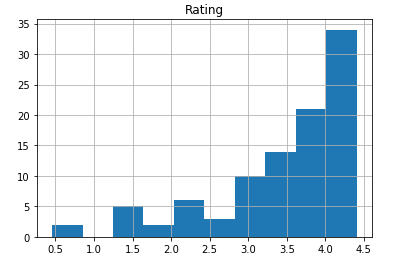
**3.2 Plot the ratings**

I plotted the average ratings of different neighbourhoods wrt to the ratings of the restaurants average ratings to get a good picture of the same.



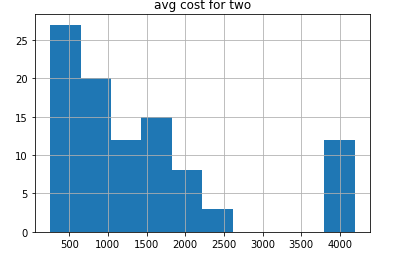
**3.3 Graph of absolute ratings**

I plotted a graph of the ratings of restaurants to be able to see the distribution of the ratings in a viewer friendly manner.



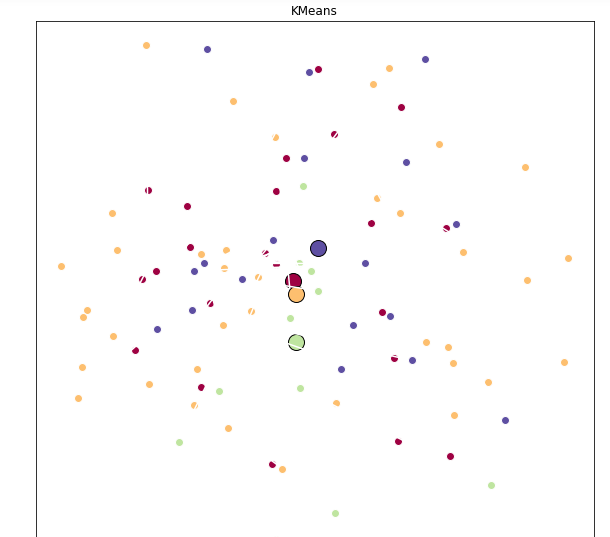
**3.4 Graph of Avg. cost**

I plotted a graph of avg. cost for two of restaurants to be able to see the distribution of the avlues in a viewer friendly manner.



**3.5 Clustering**

I applied KMeans algorithm to make clusters of the data visible so as to make different clusters of restaurants based on their ratings, avg cost, online presence , bookings , etc. I plotted these clusters then to get a visual confirmation of the palpability of these clusters and to see if they are that wide distributed.



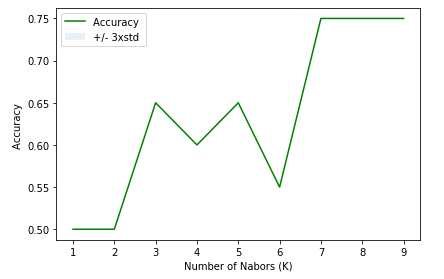
**Methodology**

**4. Predictive Analysis**

I applied the K nearest neighbors predictive algorithm & Support Vector Machines algorithem to predict values of the ratings with respect to the features we selected earlier.

**4.1 KNN algo**

I applied the algo and saw the result and there was an difference in the predicted values and the actual values, I used number of clusters as 7 because for the varied values of clusters the accuracy I got was of :



**Discussion**

The score that I observed were :

|  |  |
| --- | --- |
| Train set Accuracy | 0.7532467532467533 |
| Test set Accuracy | 0.75 |
| F1 score | 0.7497777777777777 |
| Jaccard index | 0.75 |

**4.2 Support Vector Machine**

For the value of gamma = 0.01 & C = 100 I observed the highest accuracy for the polynomial Kernel.

|  |  |
| --- | --- |
| Train set Accuracy | 1.0 |
| Test set Accuracy | 1.0 |
| F1 score | 1.0 |
| Jaccard index | 1.0 |

Here we see that the Support Vector Machine gave the best accuracy among these two and as the accuracy is of 1 there is no need to look for other predictive algorithms.

**Result**

**4.3 Actual Prediction for a restaurant**

Now I made an entry for a hypothetical restaurant which will come up at Adugodi, with avg. cost for two to be 250, with no online presence or bookings online and the result of rating came out as 1.

**5. Conclusion**

The predictive model has been made and trained for assessment with a very good value of accuracy. Now we can use it for prediction of values that can be fed to reach the decision of opening or not opening a restaurant at a certain location.