**Best Areas in Mumbai to set up a restaurant**

1. **Introduction/Business Problem:**

**Opening a new restaurant involves a lot of planning and a lot of investment. There were numerous restaurants that were set up and closed within a few months due to less profitability and less footfall. It should be understood that there are a lot of factors that contribute to the success of a new place.**

**Choosing an optimal place to open a restaurant will give the owners a big boost and the right choice will also bring in new potential customers. This project aims to suggest the places that can be an optimal choice for opening a restaurant.**

1. **Data:**

Data for the analysis was primarily gathered using foursquare API, using the explore and the venue endpoints. Even before the API, we will have to gather a list of all the postal areas in Mumbai and then get their coordinates, that could be passed to the API.

Postal areas in Mumbai were available openly on the internet. The coordinates of all the postal areas were obtained by GeoNames Database. (<http://www.geonames.org/export/zip/>).

The data then was used to create an initial data frame and we can then start exploring the dataset.

1. **Methodology:**

The primary activity performed before any results were obtained was gathering data from Foursquare API. This was done in 2 stages:

* Using the explore API: Foursquare Explore API was used to get all the recommended places near the postal region latitude and longitude. The results of this section will include the Postal Area and the nearby places along with their place ID, which will be passed onto the foursquare API to get rich venue details.
* Using the venue API: Foursquare venue API was used to get all the rich details about a venue such as :
  + 1. Likes
    2. Ratings
    3. Tips
    4. Category associated with the place.

The above details will then be used to cluster the areas.

Also, the above-mentioned columns (excluding the Category column) were normalized using the Standard Scaler module of the sklearn library.

At relevant places, a heat map of the dataframes were generated to see if there are any null values.

One point worth noting is that there were many places that had no rating and tips and that was severely limiting the data that could be gathered. In such a case, the tips or ratings were initialized to zero.

This was done as a means to also penalise places that had no ratings or tips.

Also, the category column was a categorical column and had to converted to numerical column. This was done using the pandas get\_dummies method.

An important step was performed where the rows are grouped by Area Name, and mean is taken of the rows. This will give us the frequency of occurrence of each category and also normalizes / averages other columns (Ratings, likes, tips). The dataframe finally created will be used for clustering.

1. **Results:**

After running the K-means algorithm, we got 3 clusters.

1. Looking at Cluster 1:

This cluster is full of places with:

Hotels

Lounge

Coffee Shops

Restaurants

Spa

1. Looking at Cluster 0

Places in cluster 0 have very low ratings and likes.

This could be due to the fact that the ratings and tips were set to zero if they were not available.

This also infers that the places in these areas were not rated in the first place. This could be due to less footfall or the area just not being weekend friendly.

Looking at the top categories in the cluster, we see that there are very few restaurants / eateries in such places. Defining categories are:

* Home Service
* Outdoors and Recreation
* Flea Markets
* Lots of snack places

1. Looking at cluster 2:

Areas in cluster 2 have limited / decent ratings and likes but looking at the top categories of the nearby venues, we see:

* Grocery Store
* Restaurants
* Bakery
* Coffee Shops
* Womens Store
* Clothing Store
* Shopping Malls
* Train Station

1. **Discussions:**

After getting the clusters, the only remaining task is to get meaning out of the clusters and understand with clusters specifies the areas that is good to open a restaurant.

**For Cluster 0:**

Opening restaurants in such areas is tricky and can be done by understanding if there is a need to actually open one.

Looking at a few areas such as Seepz, Airport, Santacruz, we see that places that fall in this category as mostly commercial, which also is line with the notion that these places can be 'not weekend friendly'.

**For cluster 1:**

The places in this clusters will pose a heavy competition to any new restaurant until the restaurant want to serve only the top tier customers and have the money to fight the competition and sustain itself. Areas in this cluster can be deemed as UNSAFE for new restaurants.

**For cluster 2:**

Due to the presence of Stores and Train Stations and Shopping Malls, it can be inferred that these places should have higher average footfalls. This is a very good factor for a new restaurant.

This has already been understood as we see a lot of restaurants and eateries around these areas. New and upcoming restaurants in these areas will have high competition but the higher footfall should compensate that.

All in all, the areas in this cluster should be relatively safe for new restaurants, not forgetting the higher competition.

1. **Conclusion:**

The model has tried to cluster the places based on the data that was created using the Foursquare API and it has been inferred that areas in cluster 2 could be better than the areas in the other two clusters.

One should not forget that while a good area could definitely provide good boost to the business, there are other factors that play pivotal role to the success of a running business.