# **Project Title : ZOMATO RESTAURANT CLUSTERING AND SENTIMENT ANALYSIS**

Zomato is a fastest growing restaurant discovery website, established in 2008 by Deepinder Goyal and Pankaj Chaddah. Initially, it was named Foodiebay but in 2010, it was renamed as Zomato. It provides not only information related to nearby restaurants but also provides facilities such as online ordering, table reservations and management. Zomato currently serves 10,000 cities across 36 countries having 1.2 million popular restaurants with 80 million foodies every month. It is available in 10 different languages and has 10 million reviews with 18 million bookmarks. It gives a platform to the restaurant owners to serve a large number of users a good quality of food.

The Project focuses on Customers and Company, you have to analyze the sentiments of the reviews given by the customer in the data and make some useful conclusions in the form of Visualizations. Also, cluster the zomato restaurants into different segments. The data is visualized as it becomes easy to analyze data at instant. The Analysis also solves some of the business cases that can directly help the customers finding the Best restaurant in their locality and for the company to grow up and work on the fields they are currently lagging in.

We started our project by collecting two datasets from almabetter platform, we merged datasets on common columns 'name' and 'restaurant' and created a new dataset as df.When we moved toward missing values more than half of values from the collection column were missing, we haven't removed all values, instead of removing NULL value we tried to replace those null values by others.In preprocessing After extracting important info we removed unwanted columns like Restaurant, Timings, Reviewer

When we started Exploratory data analysis we picked up variable Hour for our analysis where we found Per person estimated Cost of dining in 2017 was too high as it was expensive year, further we reached to the cost per person variable where the cost was high avg.750RS in january and very low avg. 600RS in july and the collage-Hyatt Hyderabad Gachibowli' and 'Feast sheraton hyderabad hotel are most expensive hotel

By rating variable we found that North Indian, Chinese, Continental,European, Mediterranean are most rated cuisines.After extracting day from date variable we could conclude that Monday and Tuesday was the most expensive days in week and at the end of the analysis we came know the Most Cheapest Restaurants were Amul & Mohammedia Shawarma

Further we jumped to clustering algorithms where For 2 clusters silhouette\_score was 0.70 From elbow method we got 2 number of clusters was best among all.By using dendrogram we could found 2 as optimal number of cluster After applying Several Regression models such as MultinomialNB, LogisticRegression, Decision Tree Classifier

and Random forest Regression has yielded us Best Accuracy compared to all the other models which is of 99% for TFIDF Vectorizer

& for Bag of words we applied ,LogisticRegression,DecisionTreeClassifier and Random forest Regression and LogisticRegression gave us 98% accuracy for train dataset

Contributerts Roles:

1.Lalit Ahirrao:

1.Data Wrangling

2. Data collection of Zomato Restaurant names and Metadata

3.Merging dataset\_1 and dataset\_2

4.Expensive year from all the four years

5.Restaurants with maximum numbers of reviews

6.cuisines with highest ratings

7.Elbow Method for best cluster

2.Aniket Gajmal:

1.Data Wrangling

2. Data collection of Zomato Restaurant reviews

3.Expensive and cheapest month of the year

4.Most favorite cuisines

5. Numbers of pictures uploaded

6.Silhouette score

7.Model training

3.Rushikesh Pawar:

1.Data Wrangling

2.Handeling missing values

3.Most Expensive cuisines

4.Dendrogram diagram

5.Model training

6.Power BI Report

4.Prasad Ghegade:

1.Data Wrangling

2.Handling Duplicates

3.Expensive restaurants

4.Most reviewed Day in the week

5.Bag Of Words

6. Model training

5.Samarth Gangurde:

1.Data Wrangling

2.Lowest and Maximum rating receive

3.Expensive & Cheapest Day of the week

4.Most common words by word cloud

5.TFIDF Vectorizer

6. Model training

7.Power BI Report