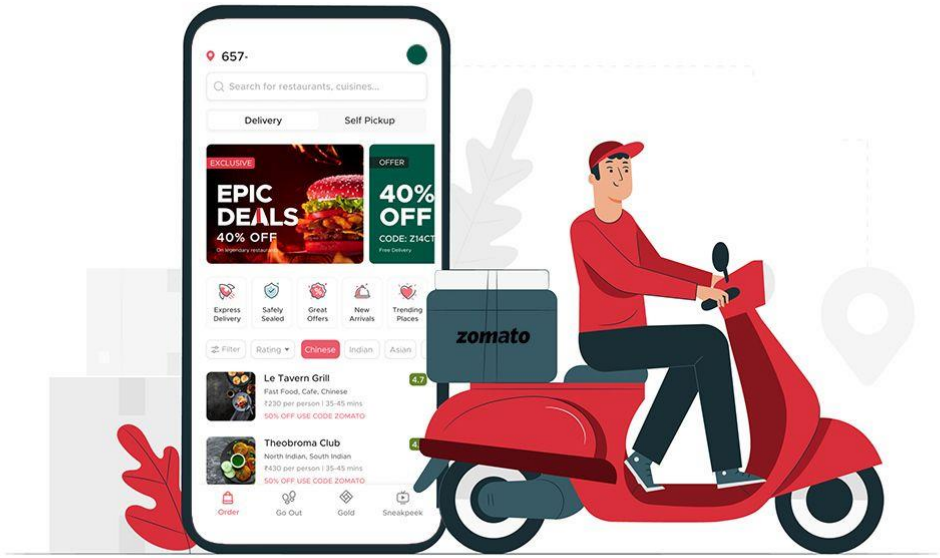
Capstone Project
(UNSUPERVISED ML)

ZOMATO RESTAURANT CLUSTERING AND SENTIMENT ANALYSIS



TEAM

- IQBAL BABWANE
- SAMEER ANSARI
- LUKMAN HAIDER KHAN

~ UNDER THE GUIDANCE OF TEAM ALMABETTER

AI

CONTENT

- INTRODUCTION OF PROJECT
- PROBLEM STATEMENT
- DATA DESCRIPTION
- EDA
- FEATURE ENGINEERING
- NLP OPERATIONS
- MACHINE LEARNING MODELS
- CONCLUSION

INTRODUCTION

Zomato is an Indian restaurant aggregator and food delivery start-up founded by Deepinder Goyal and PankajChaddah in 2008. Zomato provides information, menus and user-reviews of restaurants, and also has food delivery options from partner restaurants in select cities.

PROBLEM STATEMENT

- Zomato is an Indian restaurant aggregator and food delivery start-up founded by Deepinder Goyal and Pankaj Chaddah in 2008. Zomato provides information, menus and user-reviews of restaurants, and also has food delivery options from partner restaurants in select cities.
- India is quite famous for its diverse multi cuisine available in a large number of restaurants and hotel resorts, which is reminiscent of unity in diversity. Restaurant business in India is always evolving. More Indians are warming up to the idea of eating restaurant food whether by dining outside or getting food delivered. The growing number of restaurants in every state of India has been a motivation to inspect the data to get some insights, interesting facts and figures about the Indian food industry in each city. So, this project focuses on analyzing the Zomato restaurant data for each city in India.
- The Project focuses on Customers and Company, you have to analyze the sentiments of the reviews given by the customer in the data and made some useful conclusion 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 solve 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.
- This could help in clustering the restaurants into segments. Also the data has valuable information around cuisine and costing which can be used in cost vs. benefit analysis
- Data could be used for sentiment analysis. Also the metadata of reviewers can be used for identifying the critics in the industry.

DATA DESCRIPTION

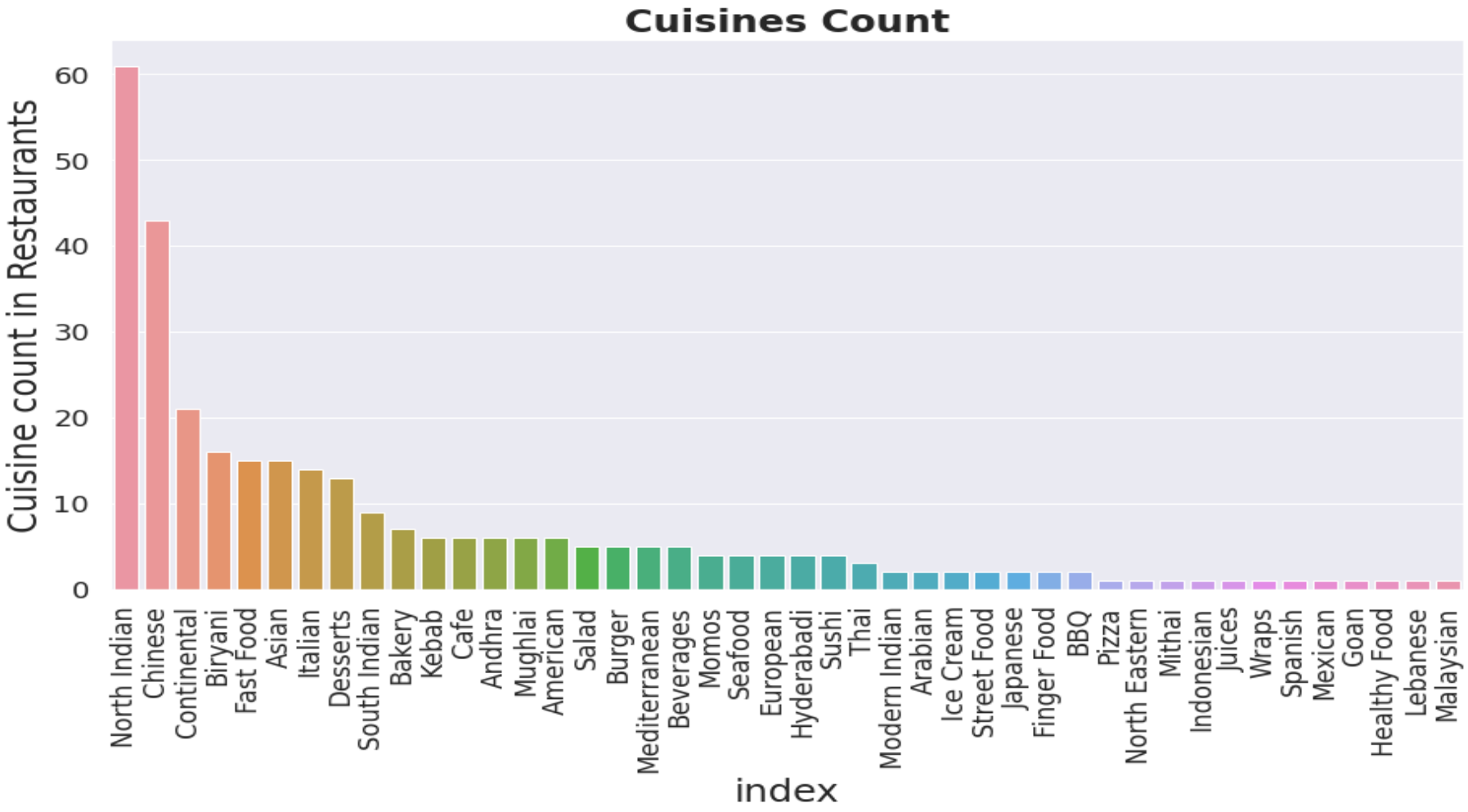
-The dataset contains two data **Zomato Restaurant names and Metadata** and **Zomato Restaurant reviews**

Data Description – 1

- Zomato Restaurant names and Metadata Attribute Information:

- Name : Name of Restaurants
- Links : URL Links of Restaurants
- Cost : Per person estimated Cost of dining
- Collection : Tagging of Restaurants w.r.t. Zomato categories
- Cuisines : Cuisines served by Restaurants
- Timings : Restaurant Timings

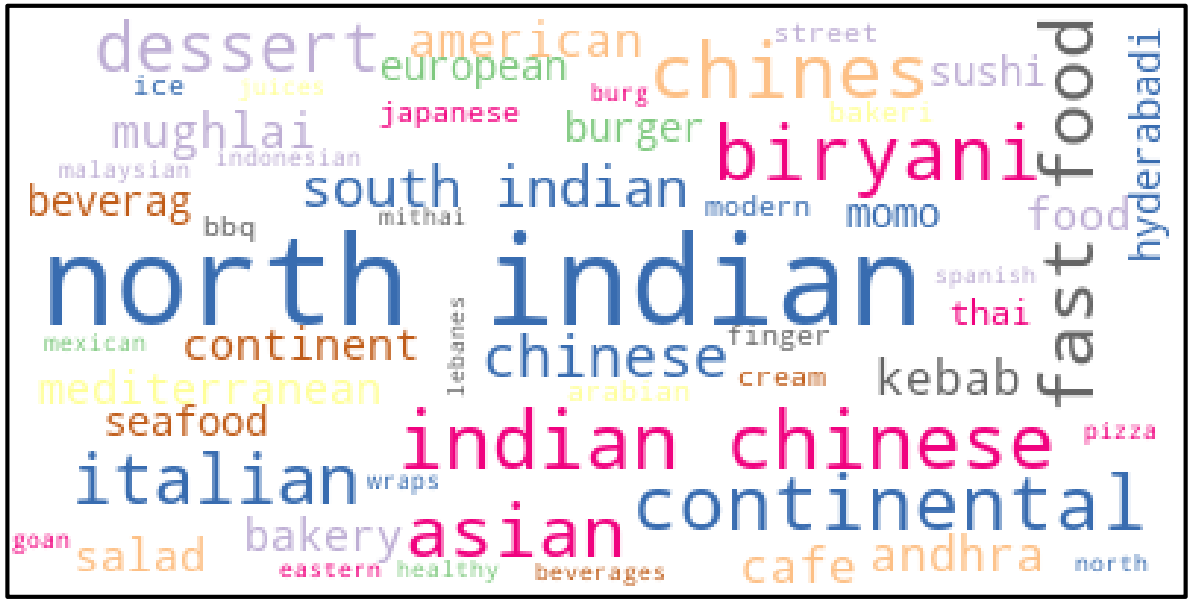
EDA



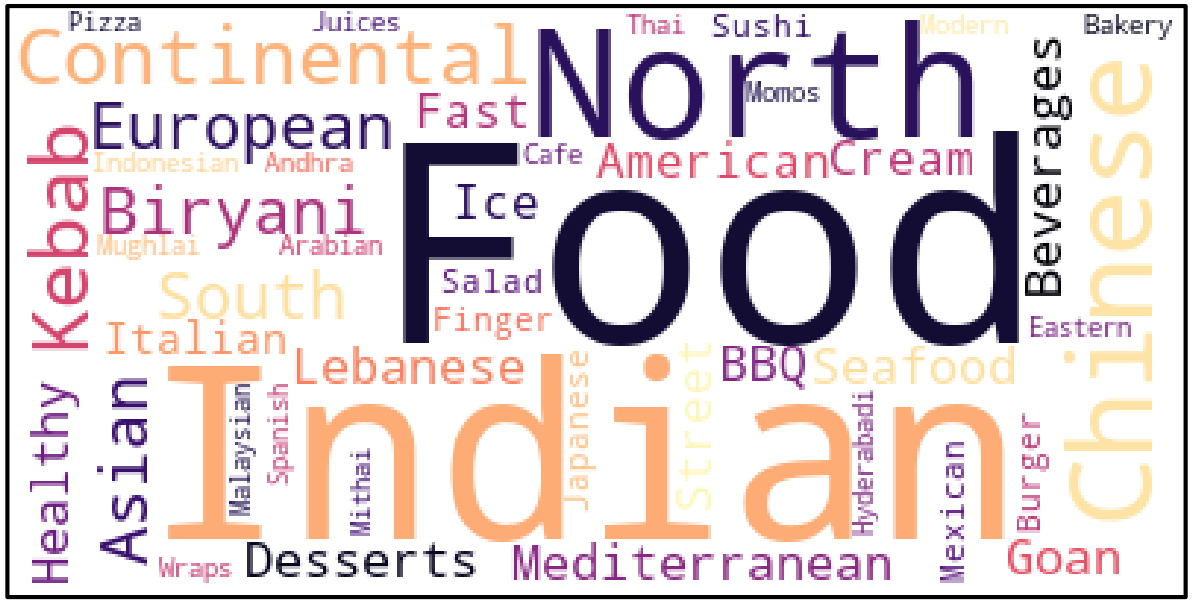
North Indian is the most popular cuisine

WORDCLOUD

- Visualization of words from 'Cuisines' feature



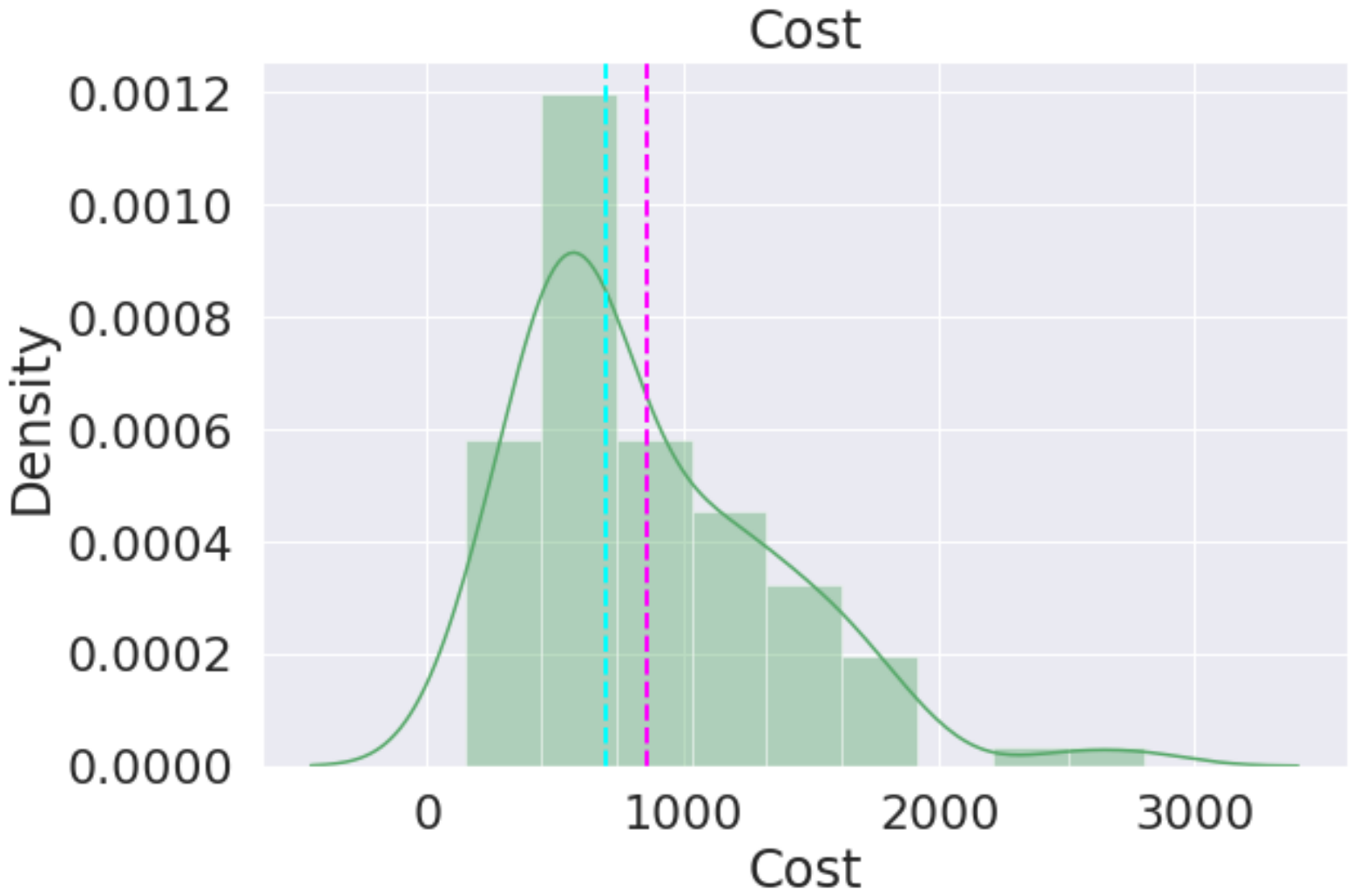
- Most nominated words from 'Cuisine' feature



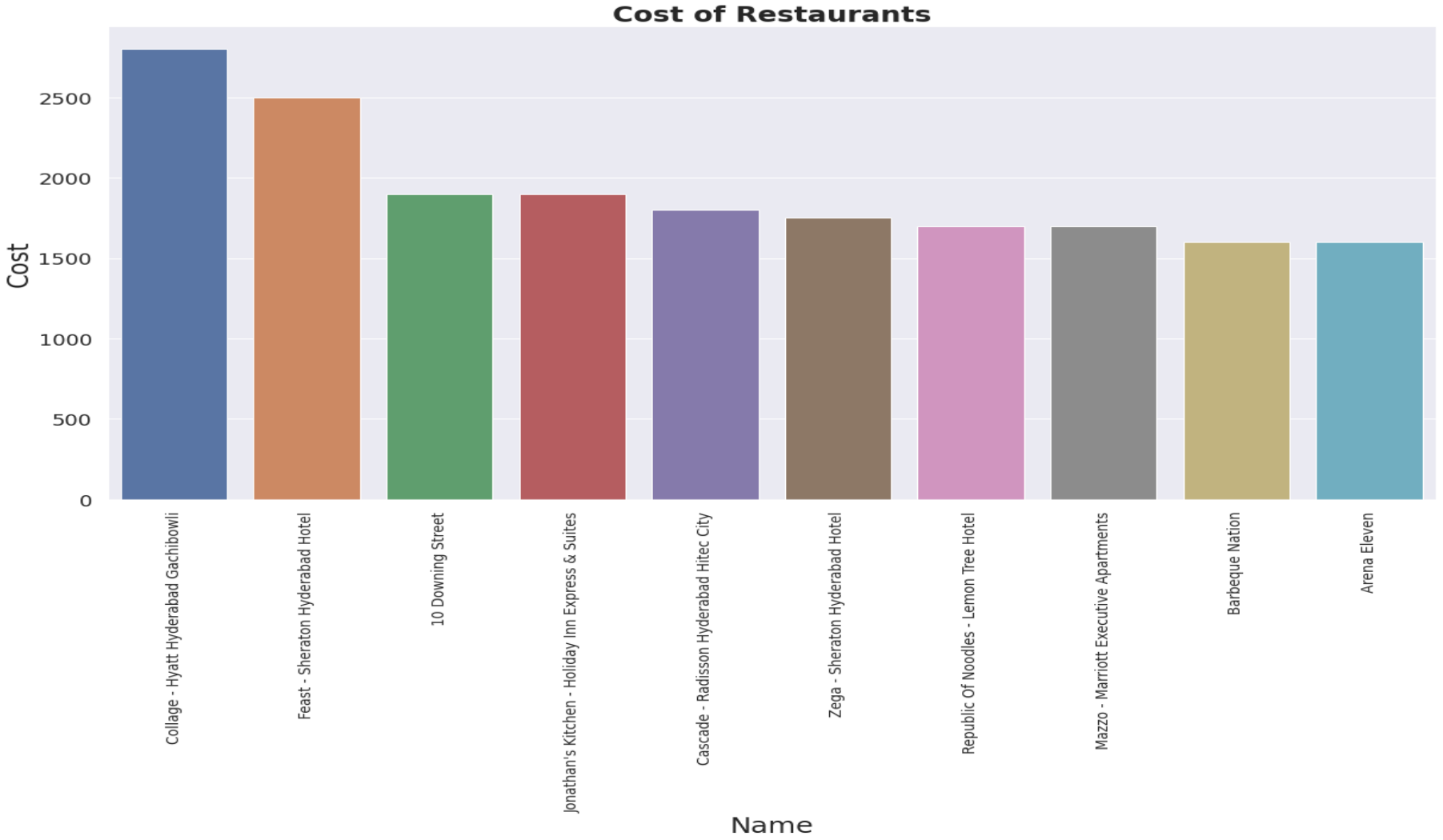
EDA

AI

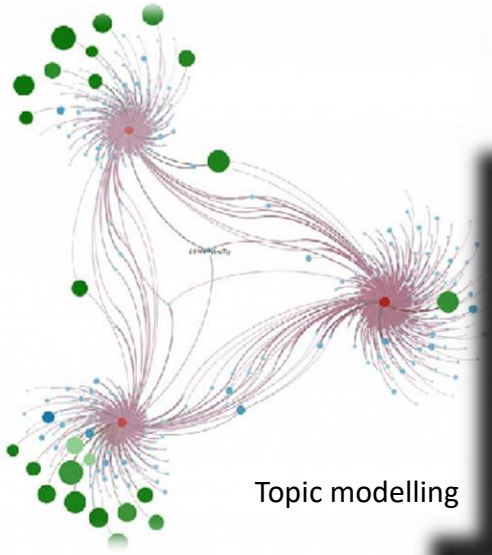
Distribution of cost column



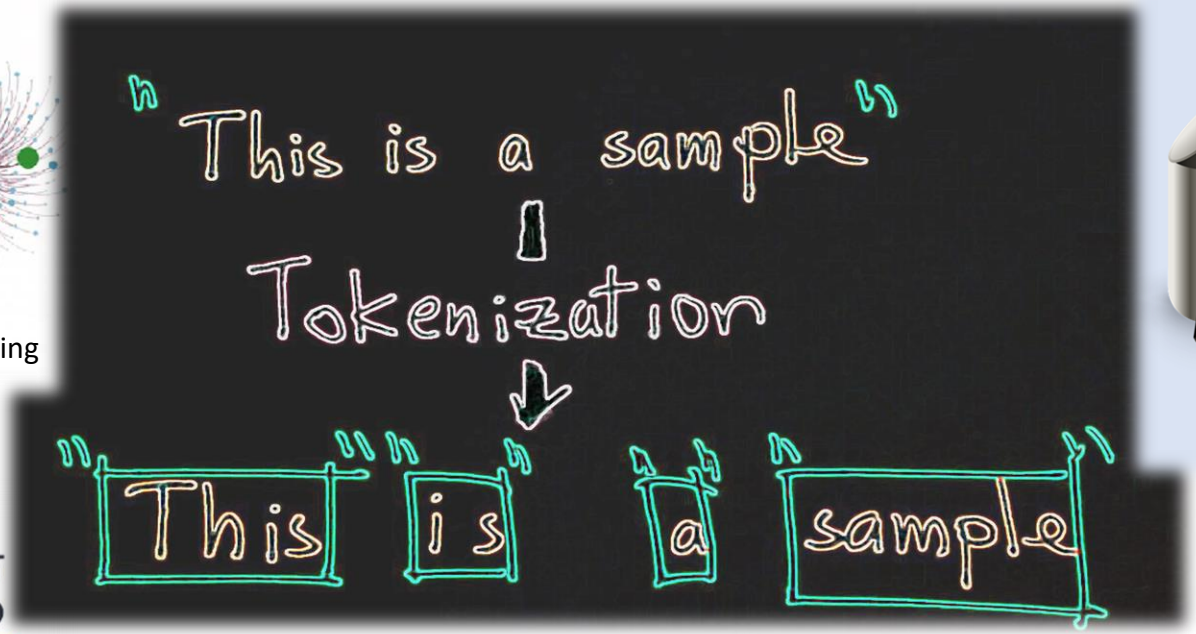
EDA

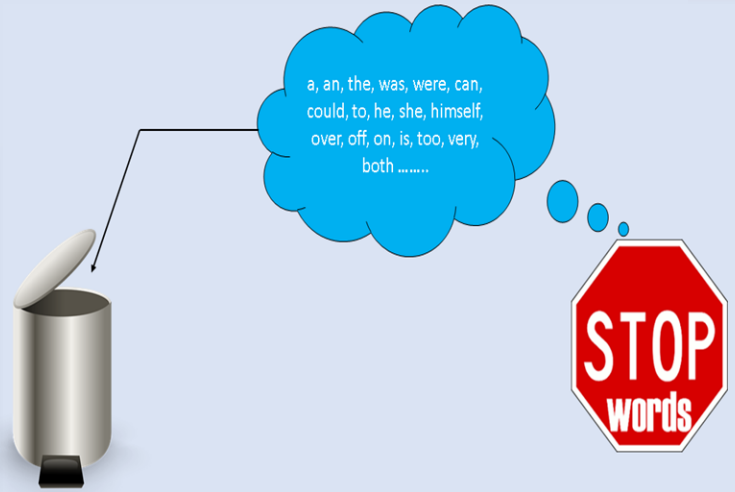


- Collage - Hyatt Hyderabad Gachibowli is the expensive restaurant with cost of 2800



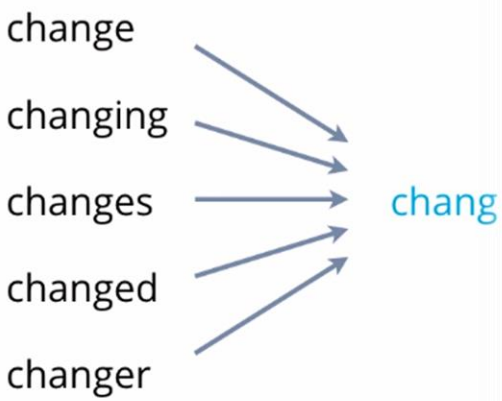
Topic modelling



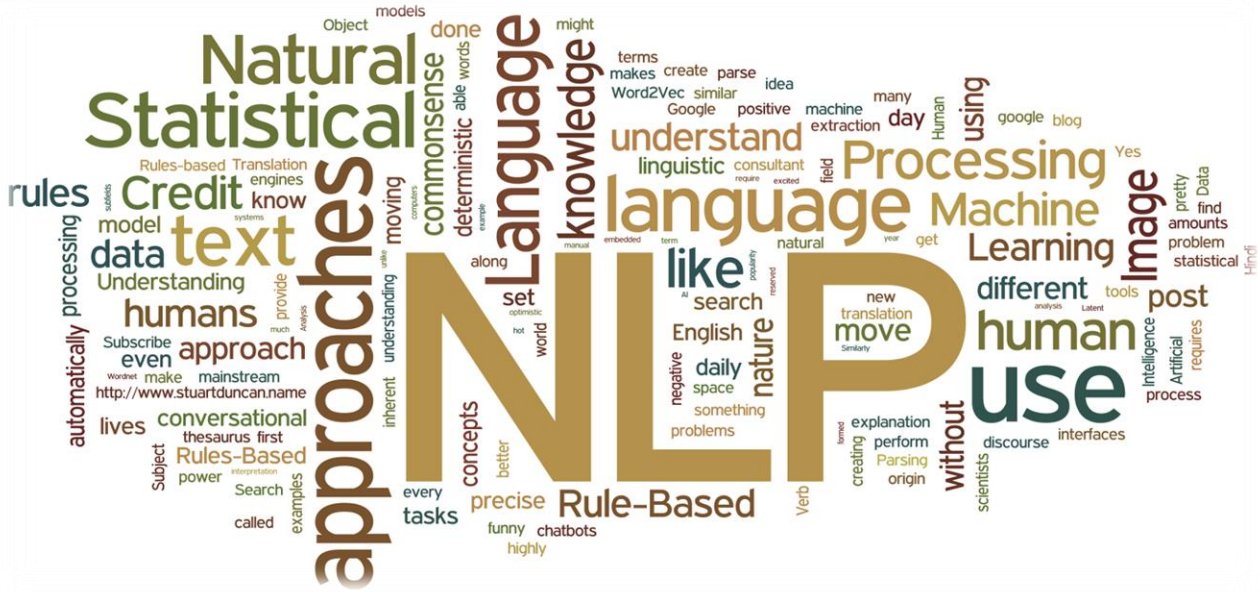


Stop words

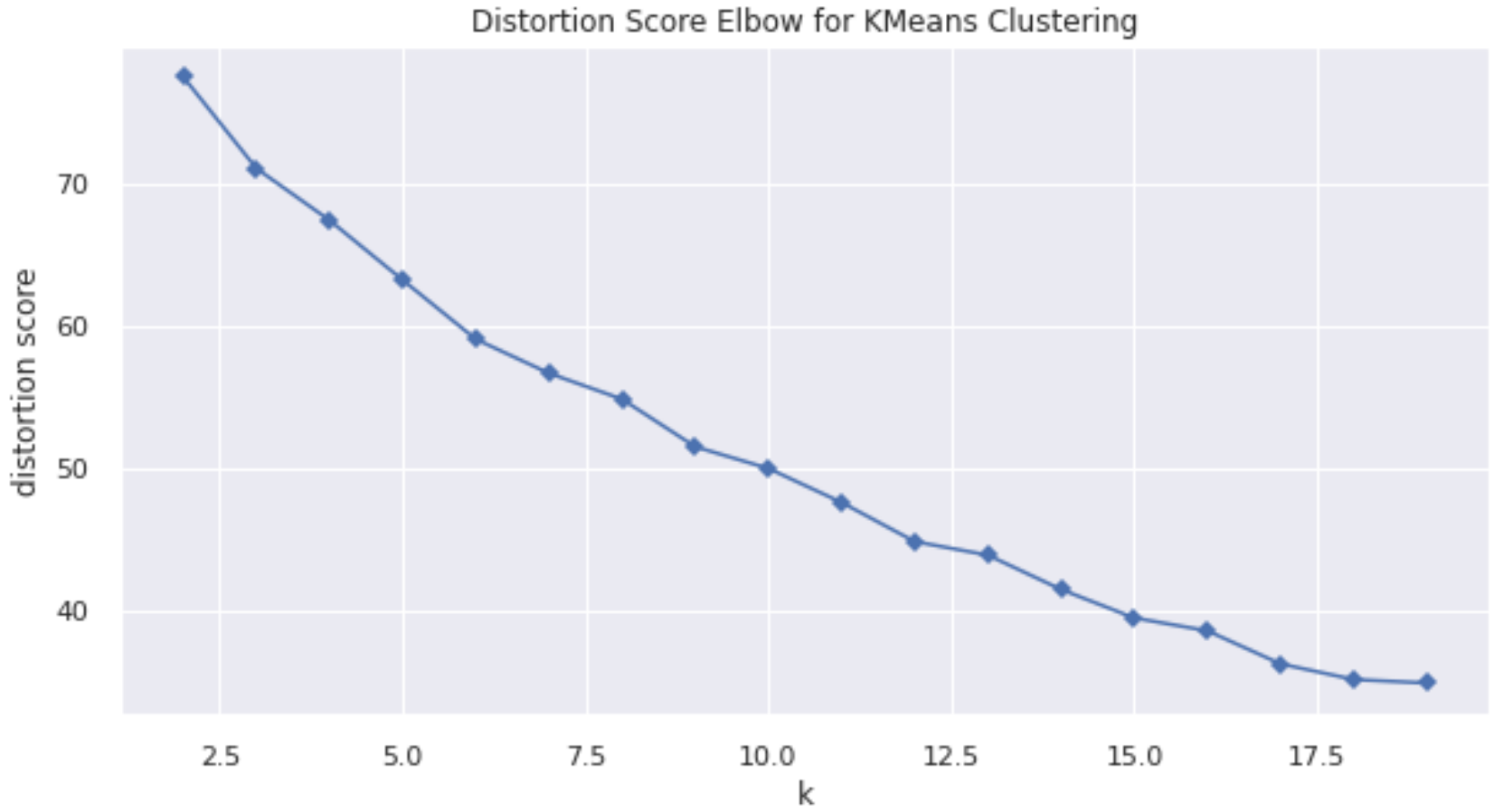
Stemming



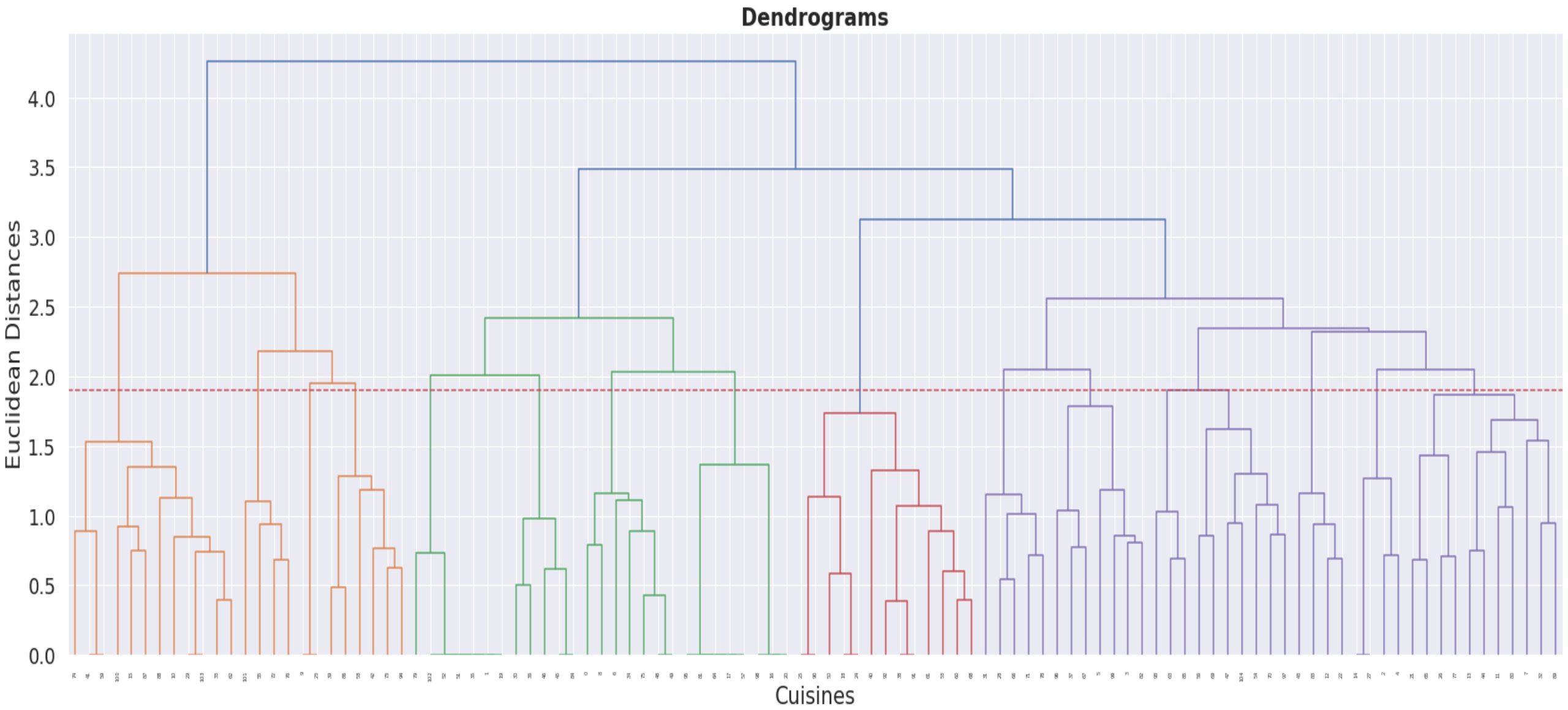
NLP
Techniques



Elbow method to find appropriate 'K' value

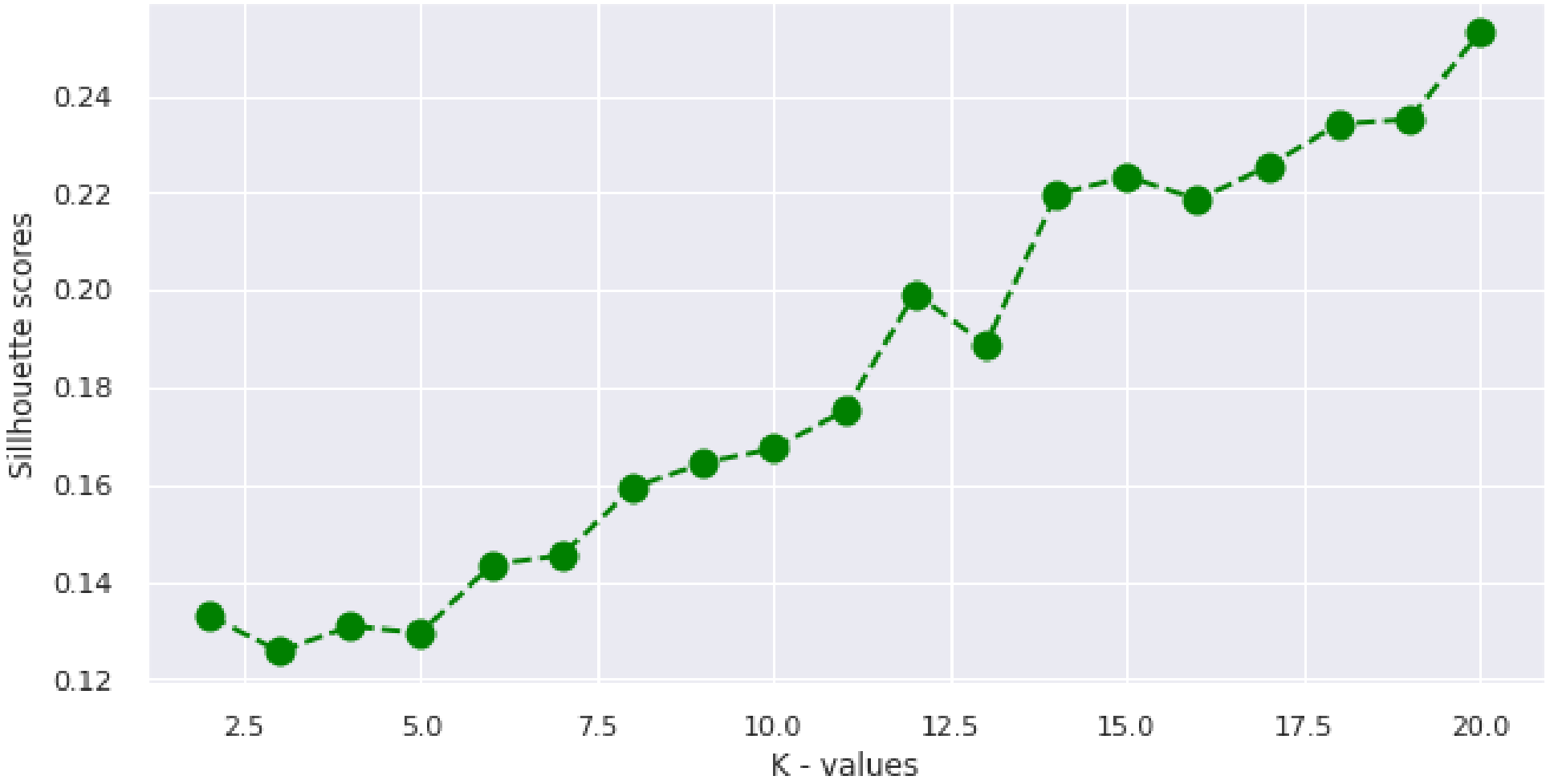


- Since here elbow occurs from 10 onwards, we can take k = 15 as optimum value



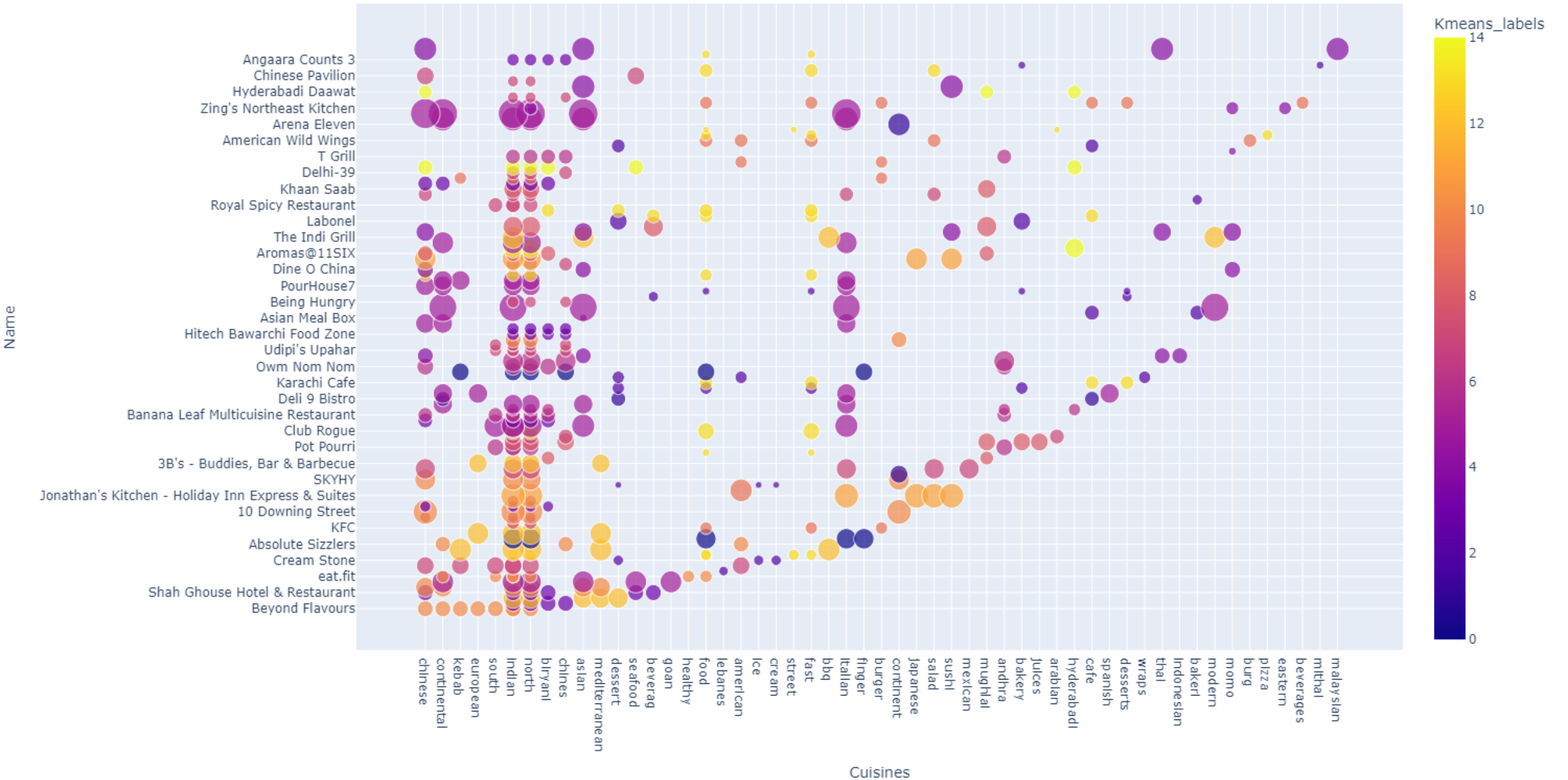
- By using Agglomerative Clustering, we came with the above dendrograms. At Euclidean Distance of 1.8, we got best cluster 15 with silhouette score of 0.238

SILHOUETTE SCORE OF CLUSTERS

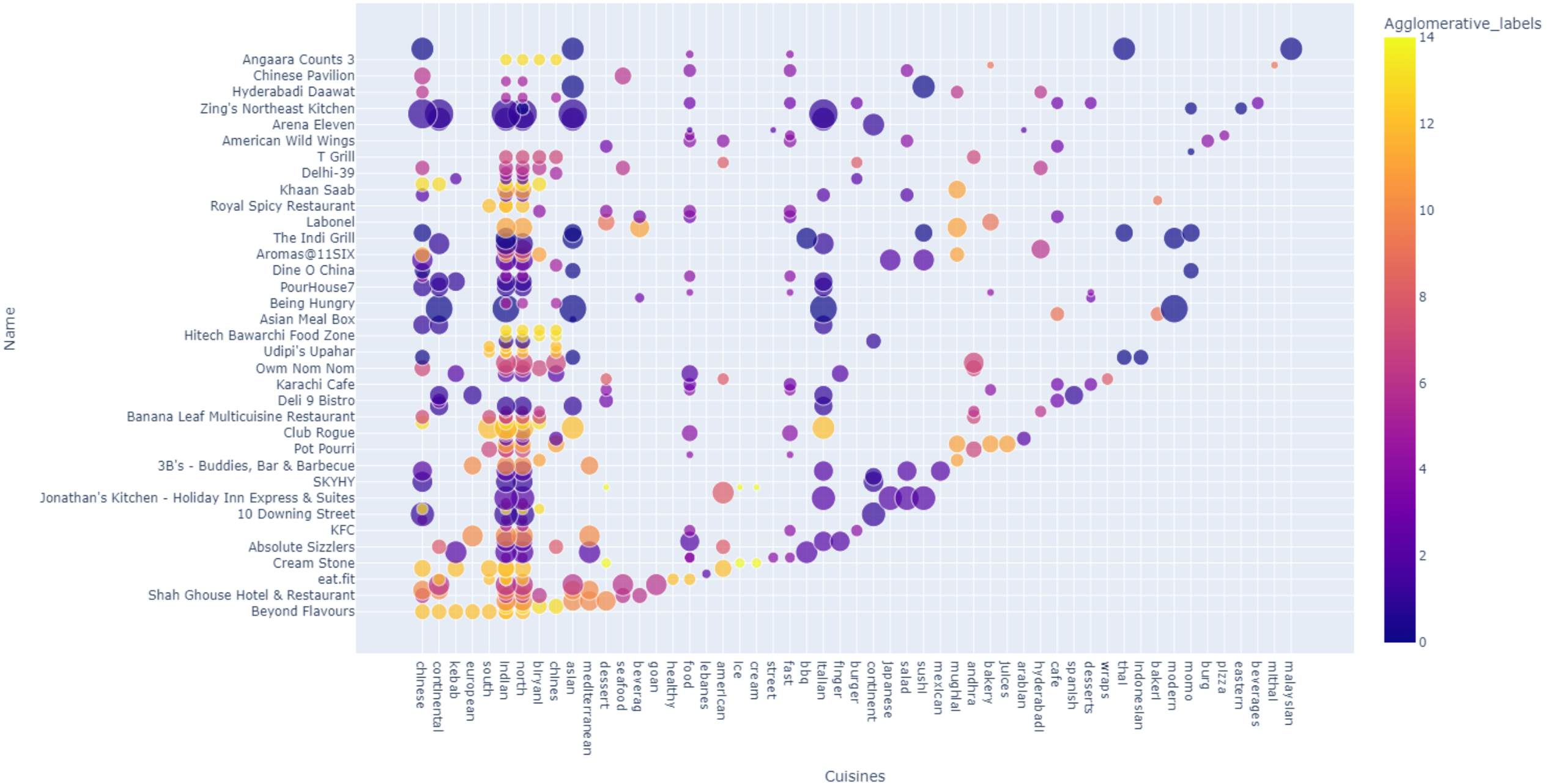


- From the above graph we can take k = 15 as optimum value of silhouette score

Clusters of K-Means Clustering



Clusters of Agglomerative Clustering



Data Description – 2

- Zomato Restaurant reviews

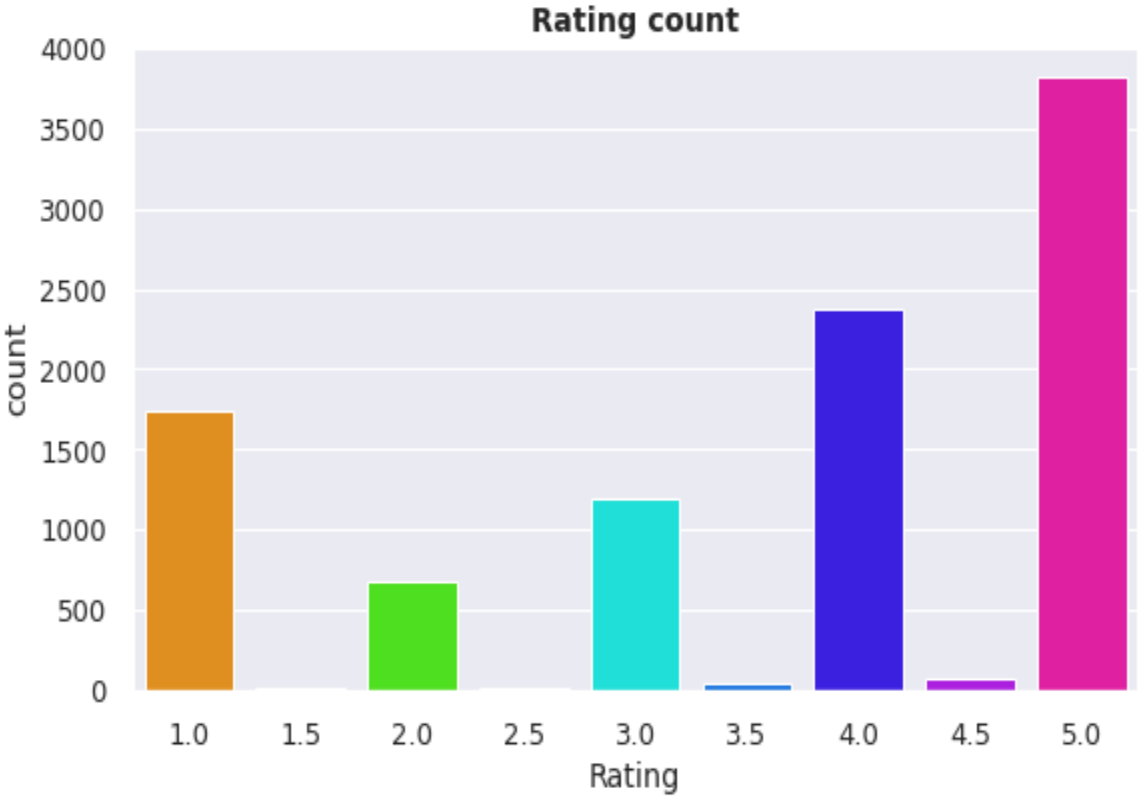
- Restaurant : Name of the Restaurant
- Reviewer : Name of the Reviewer
- Review : Review Text
- Rating : Rating Provided by Reviewer
- MetaData : Reviewer Metadata - No. of Reviews and followers
- Time: Date and Time of Review
- Pictures : No. of pictures posted with review

WORDCLOUD

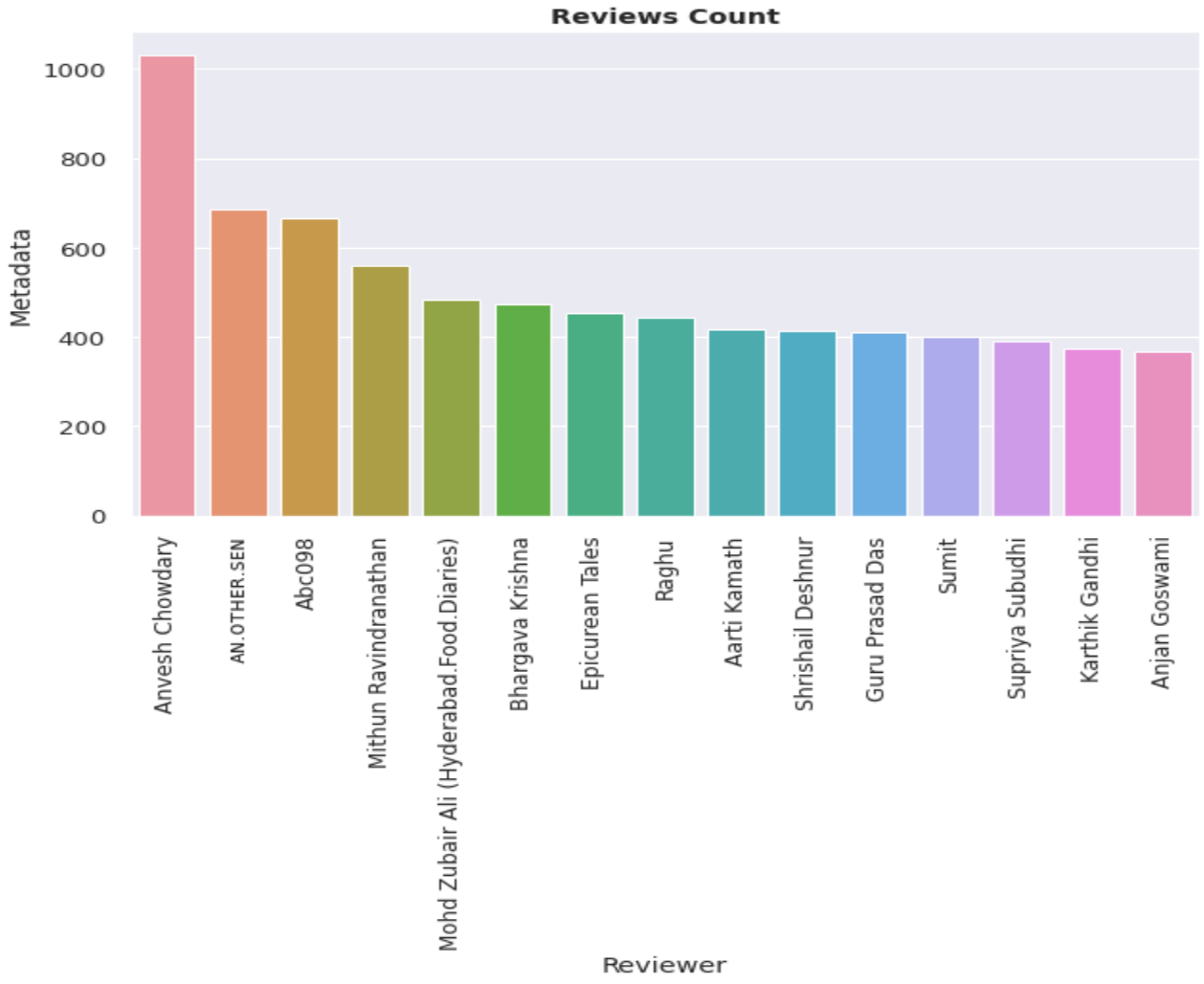
- Wordcloud for all Reviews

- Wordcloud for Reviews for more than 3 ratings

EDA

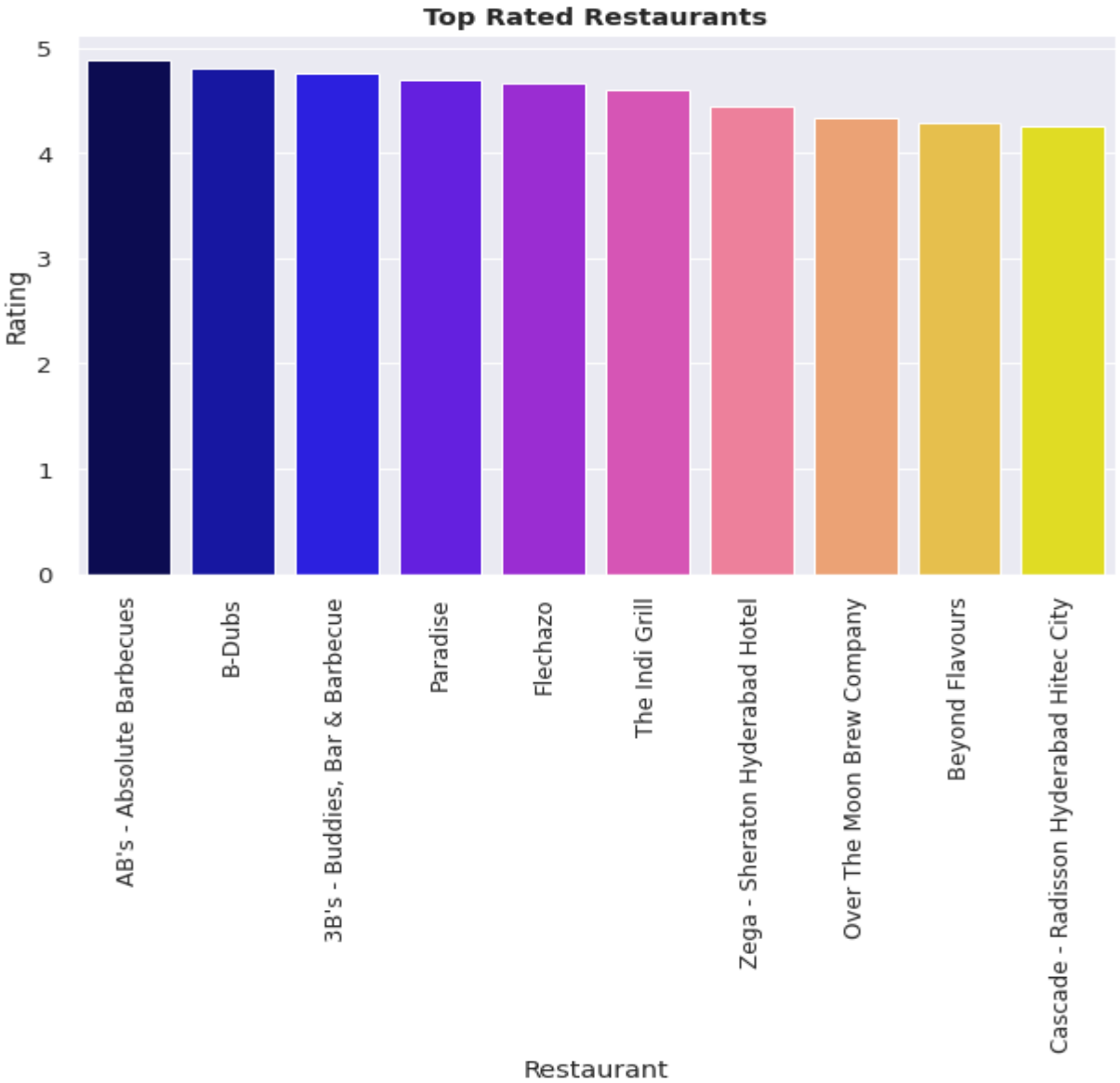


- 5.0 has maximum number of Ratings



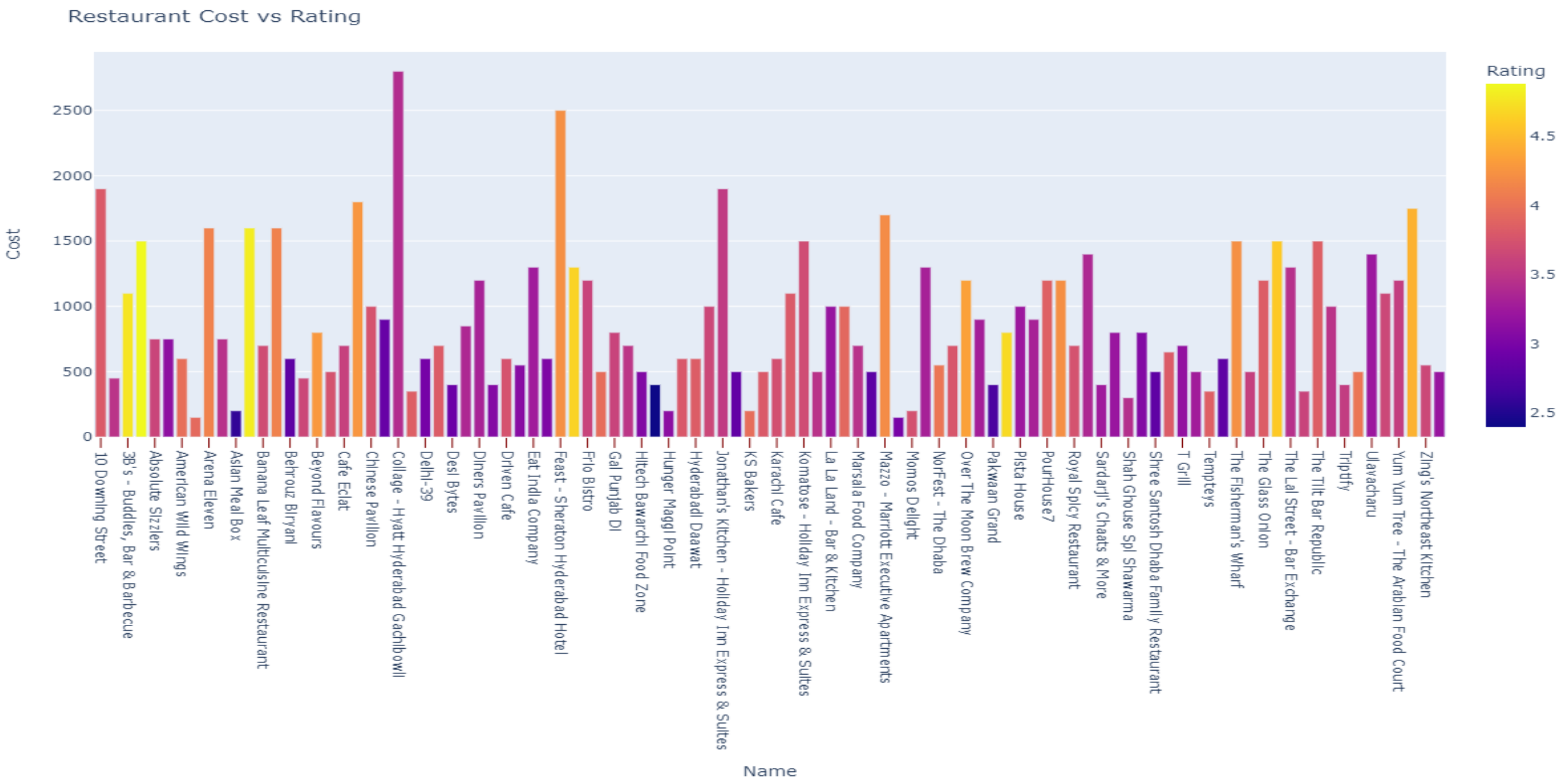
- Anvesh Chowdary is the customer which has given reviews most of the time

EDA



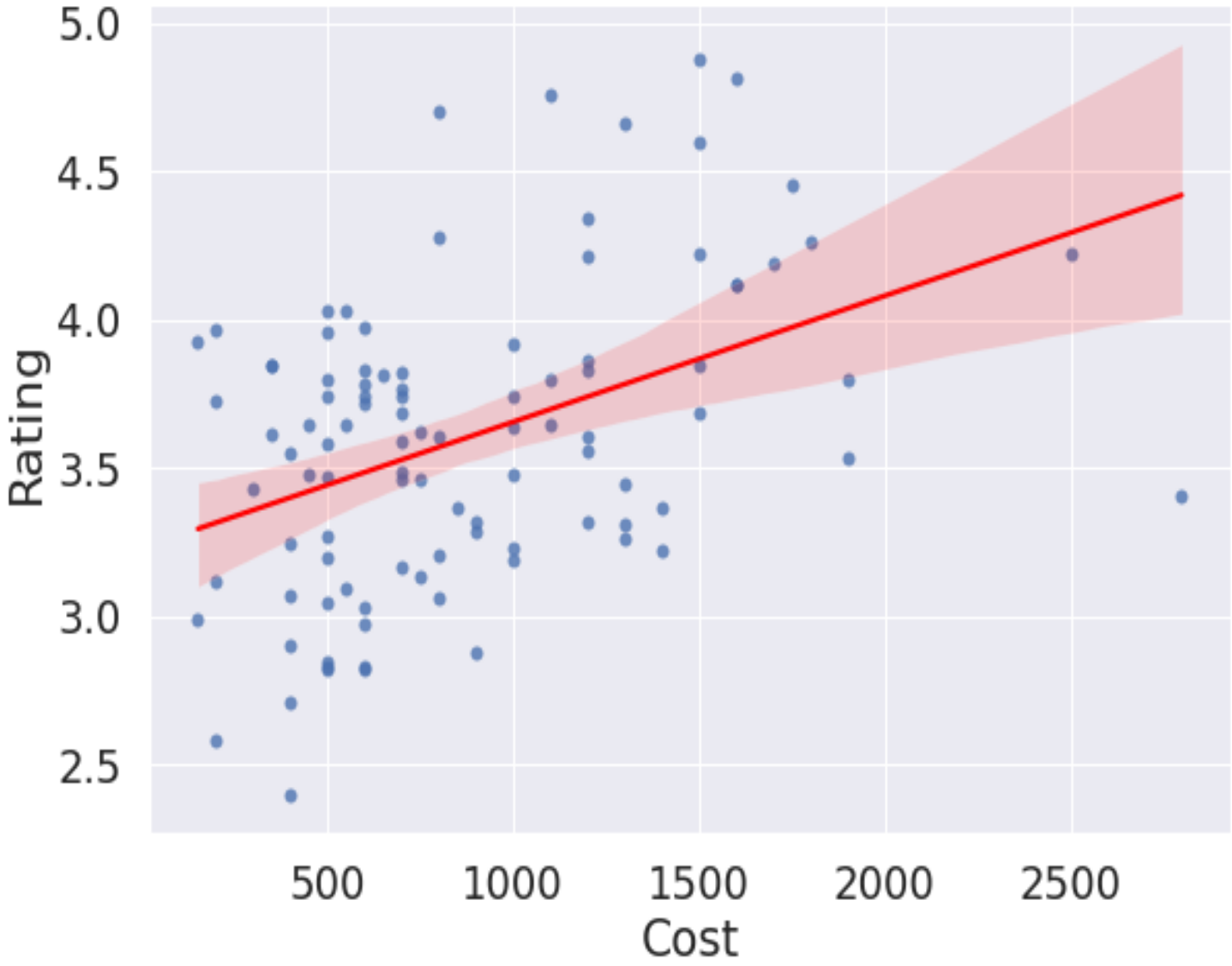
- Top 10 restaurants with highest rating, where, **AB's – Absolute Barbecues** is leading among all

Sentiment Analysis

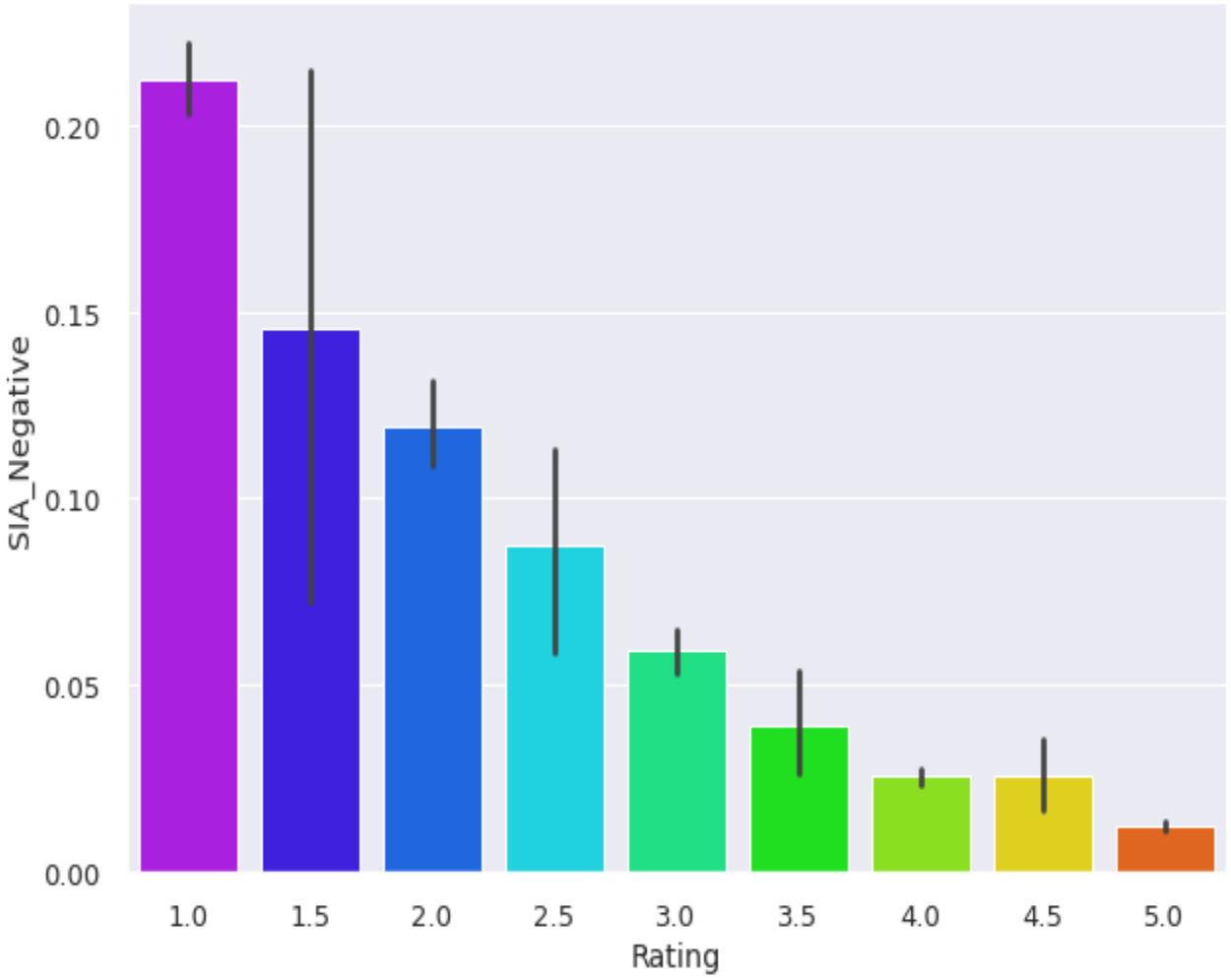


Correlation of Rating and Cost

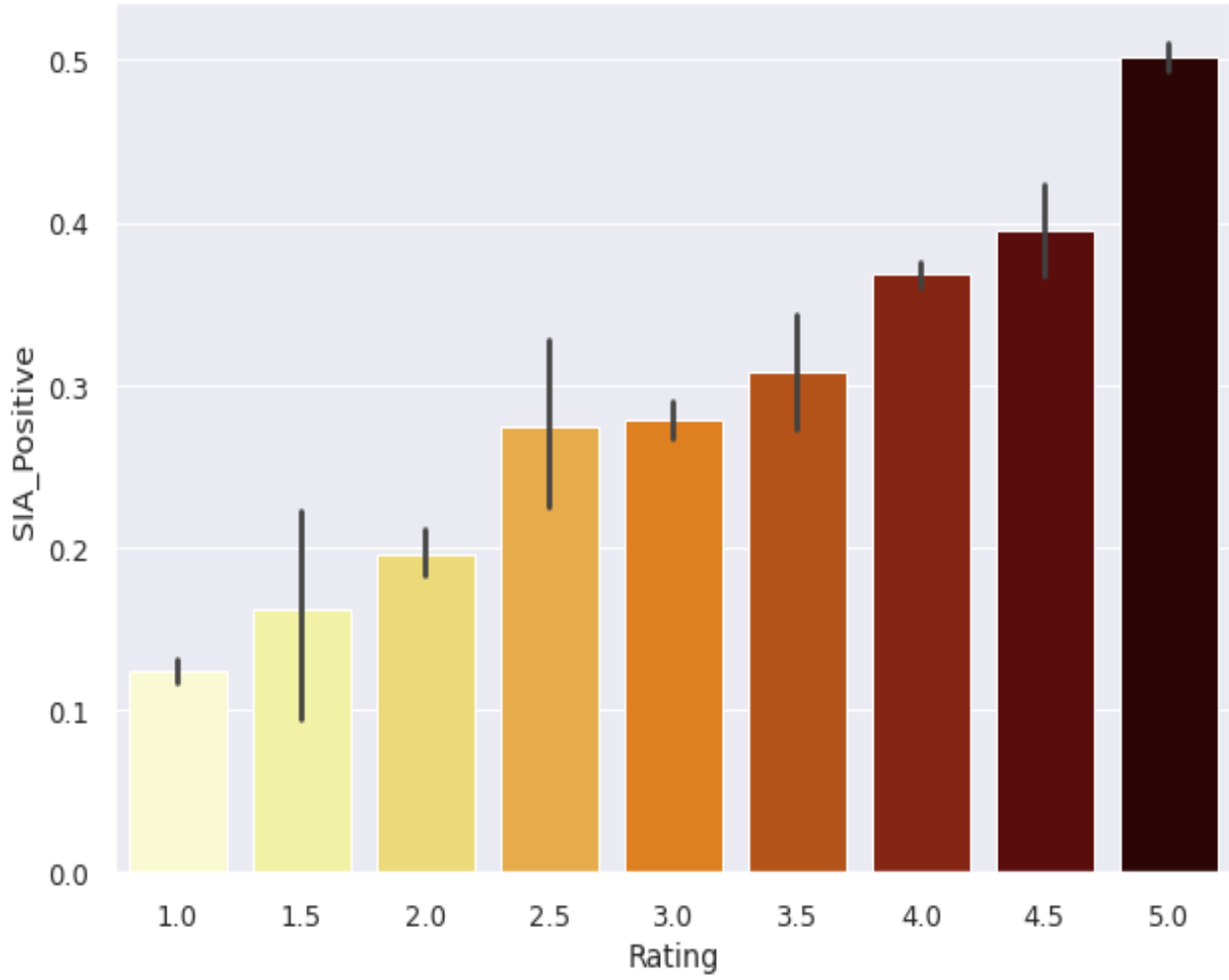
- Here we can see that rating is highly and positively correlated with cost



Sentiment Intensity Analyzer

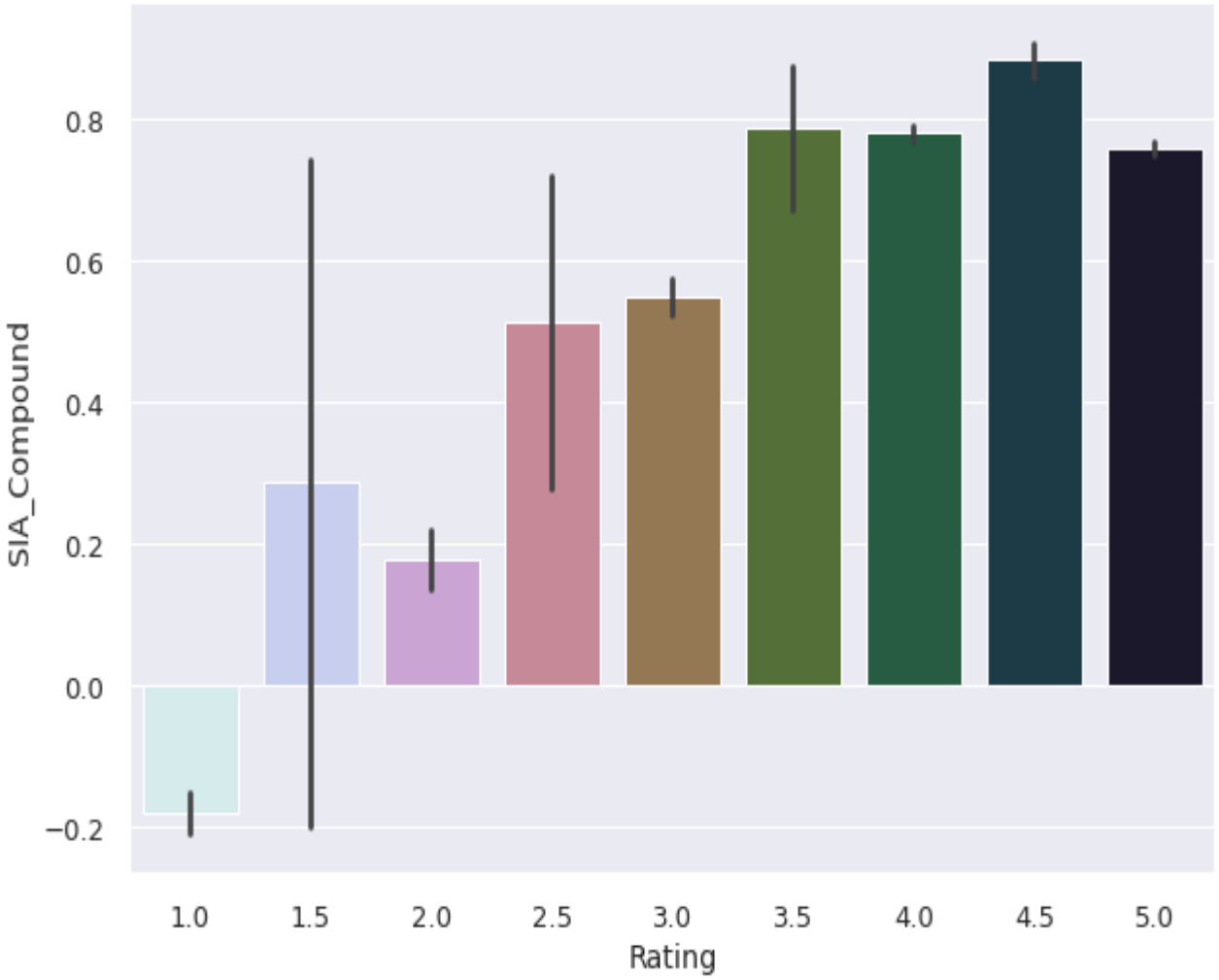


Negative polarity

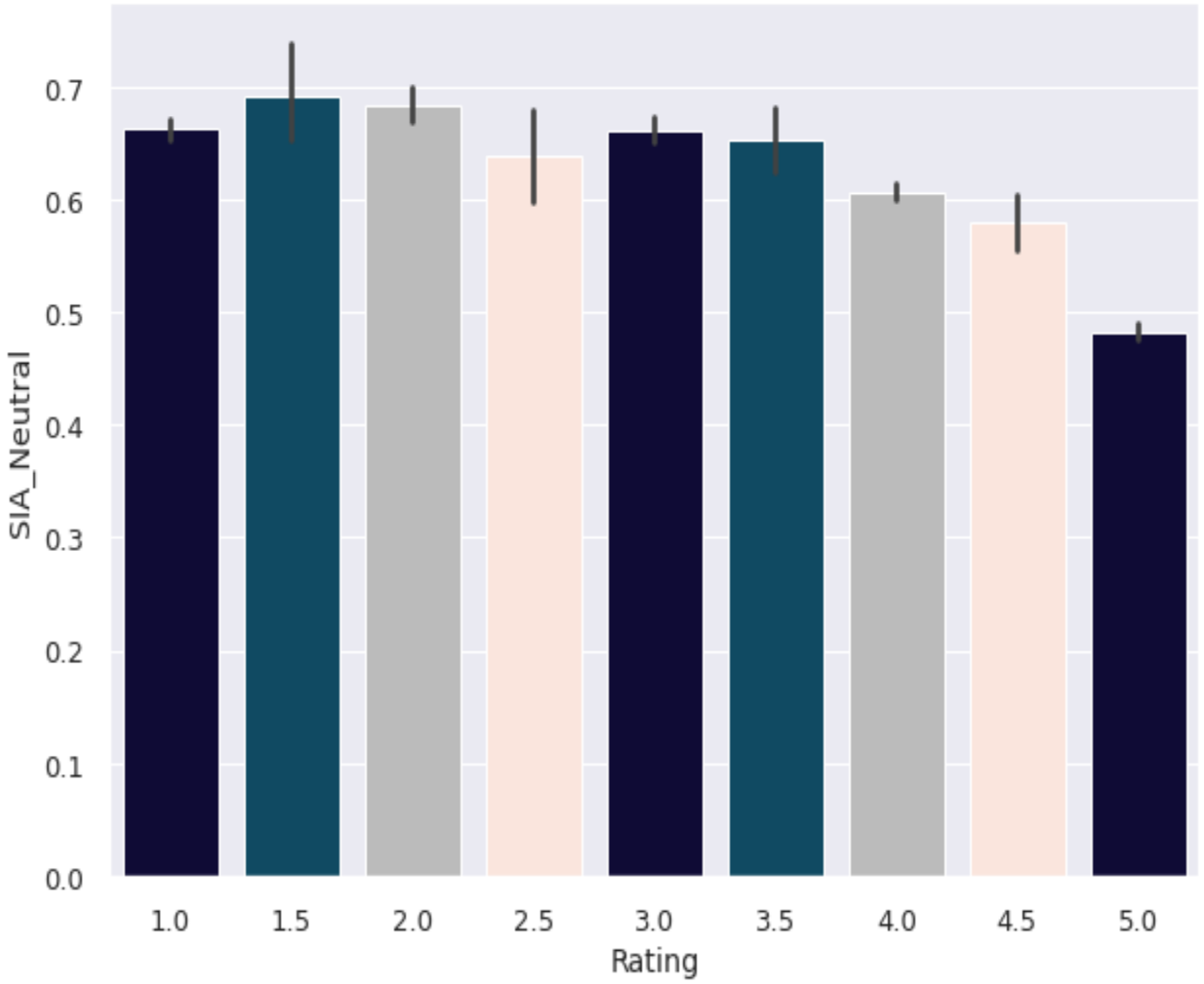


Positive polarity

Sentiment Intensity Analyzer

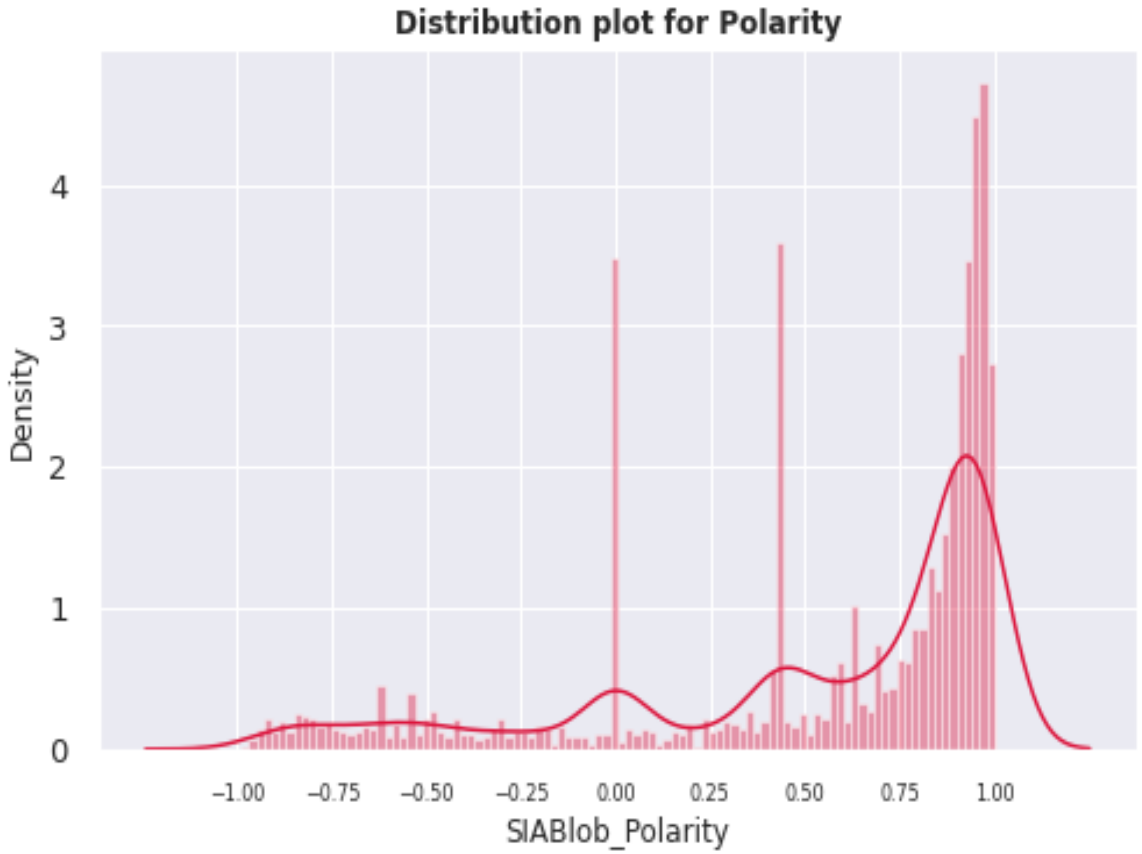


Compound polarity

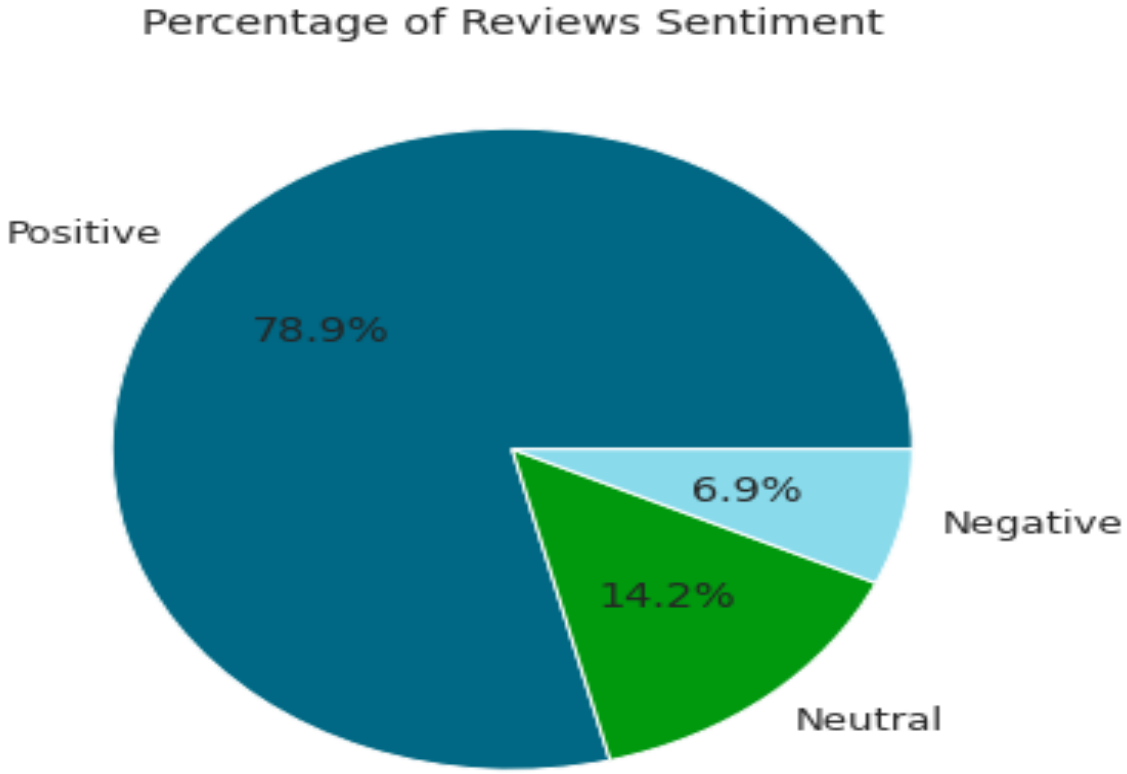


Neutral polarity

Distribution plot for polarity

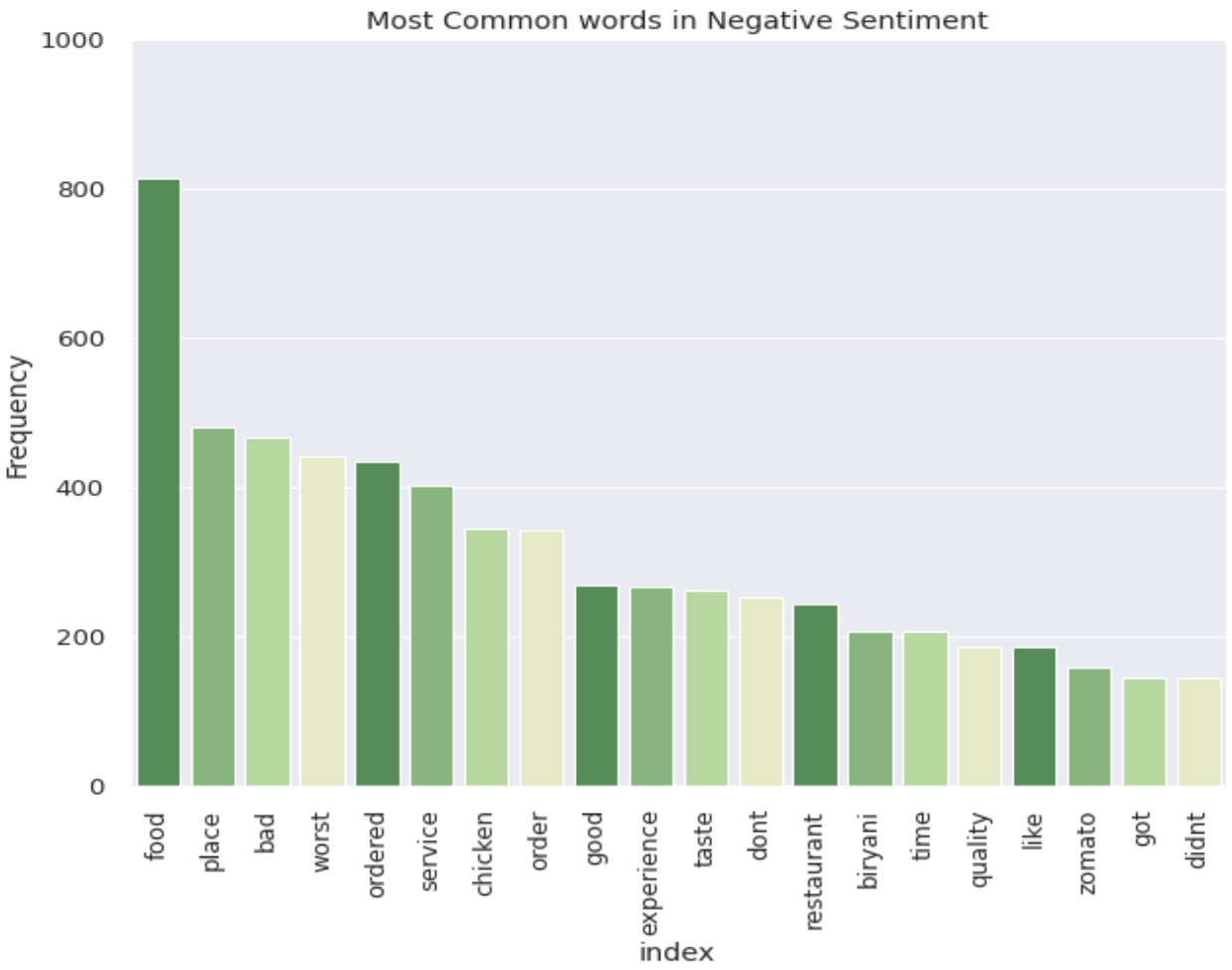


Percentage of Reviews Sentiment



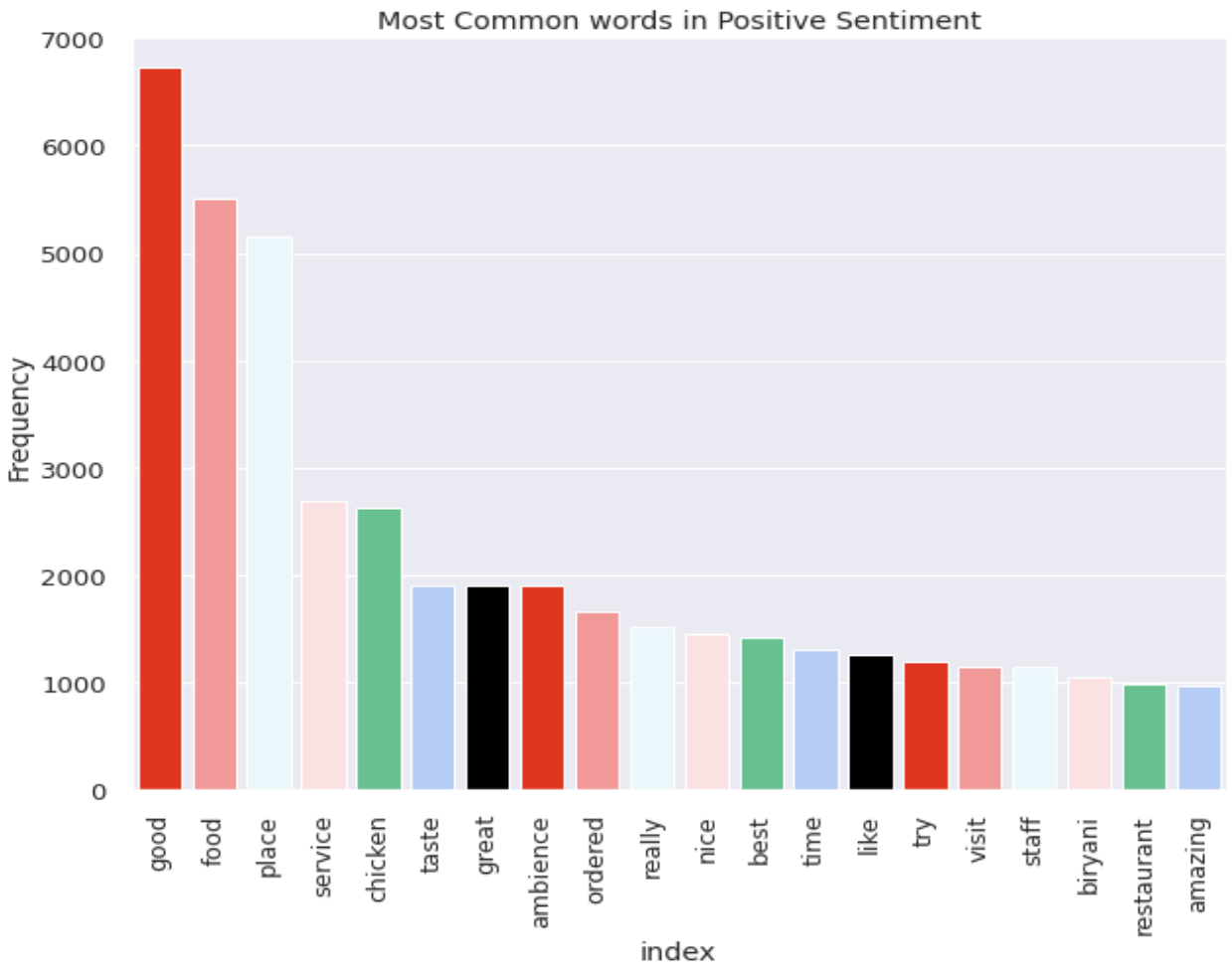
- ❑ Positive polarity is quite high followed by neutral
- ❑ Negative polarity density is very low i.e. 6.9%

Most Common words in Negative Sentiment



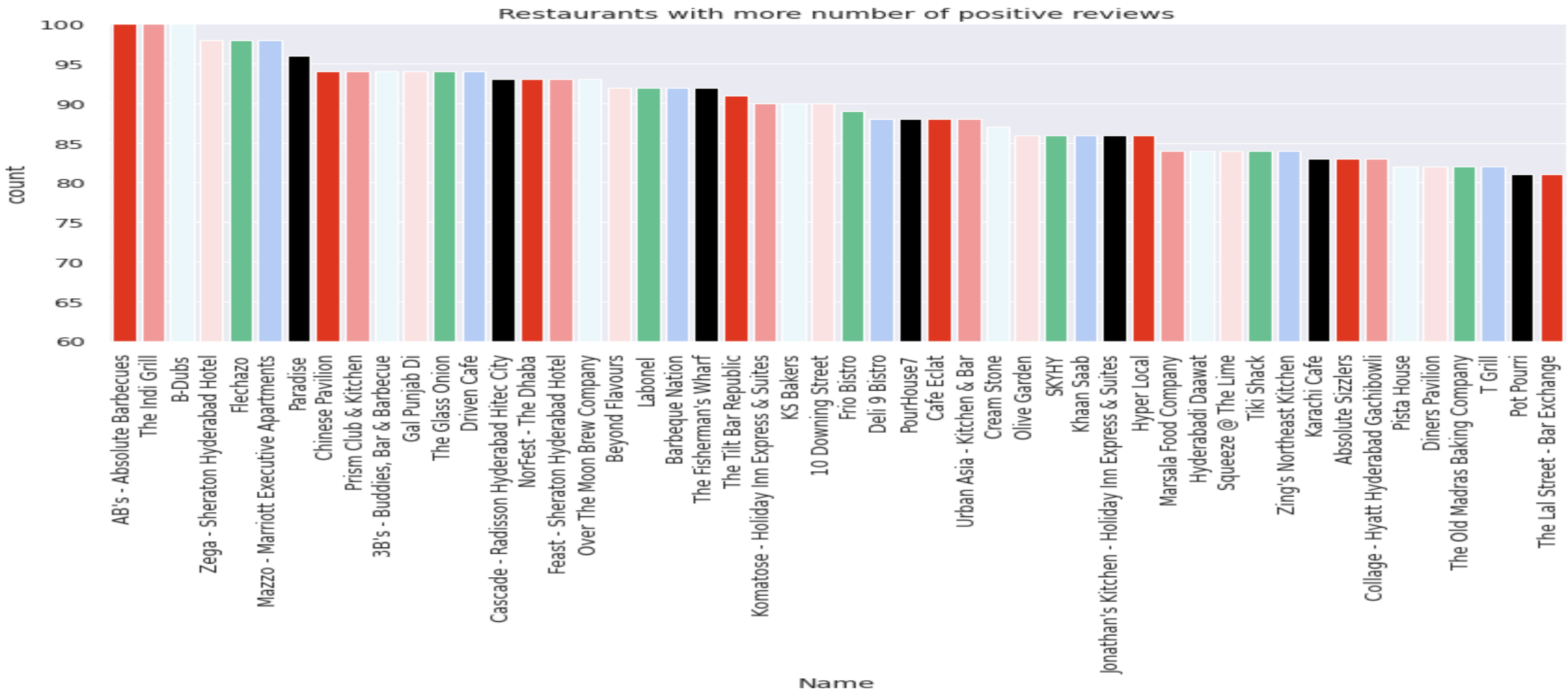
- From all the negative sentiments we separate the words which has repeated many time

Most Common words in Positive Sentiment



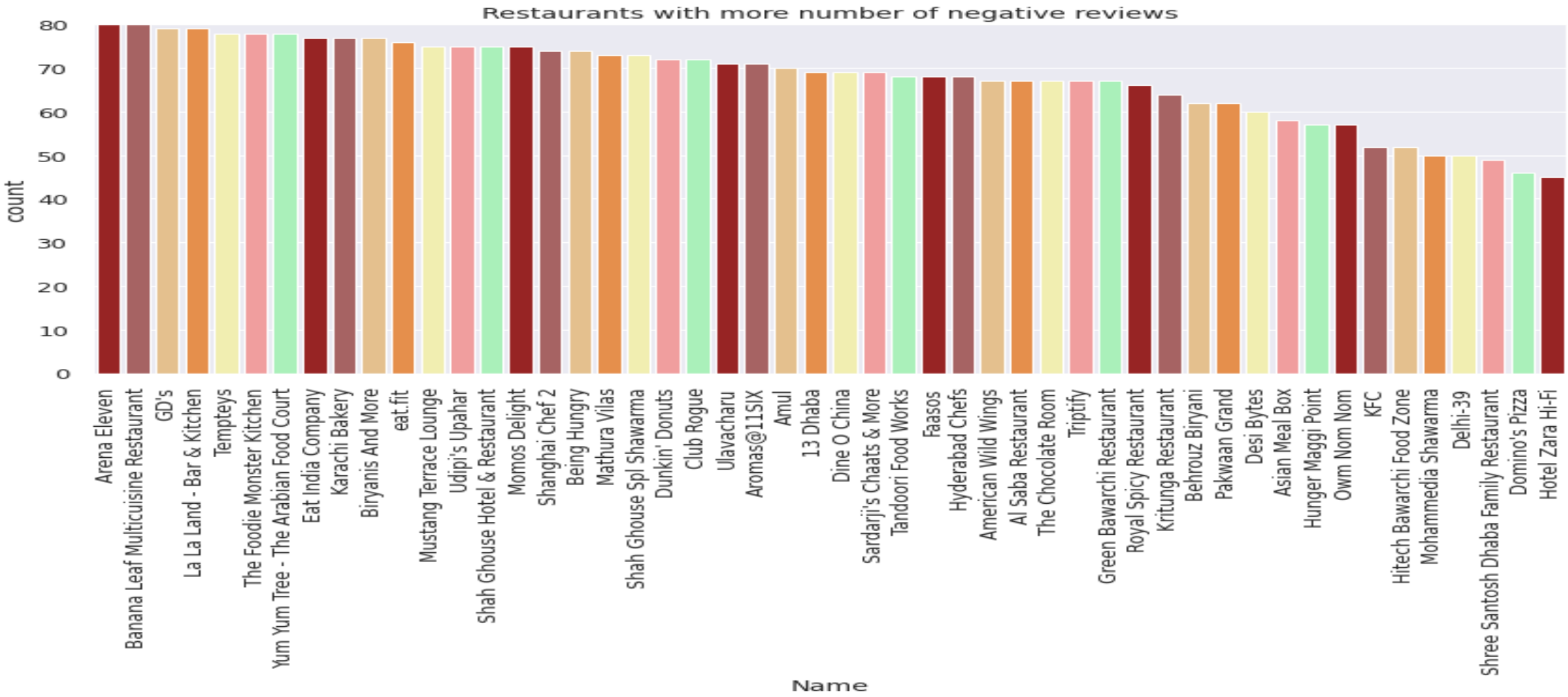
- From all the positive sentence we separate the words which has repeated many time

Restaurants with more number of positive reviews



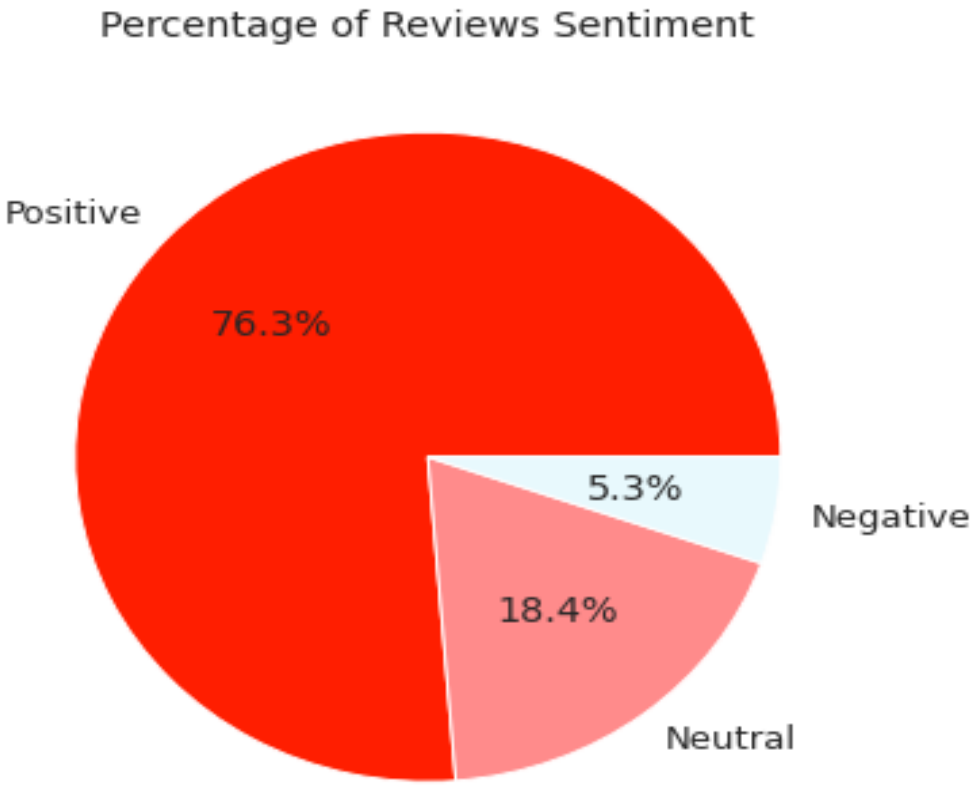
AB's – Absolute Barbecues was the restaurants having highest positive reviews

Restaurants with more number of negative reviews

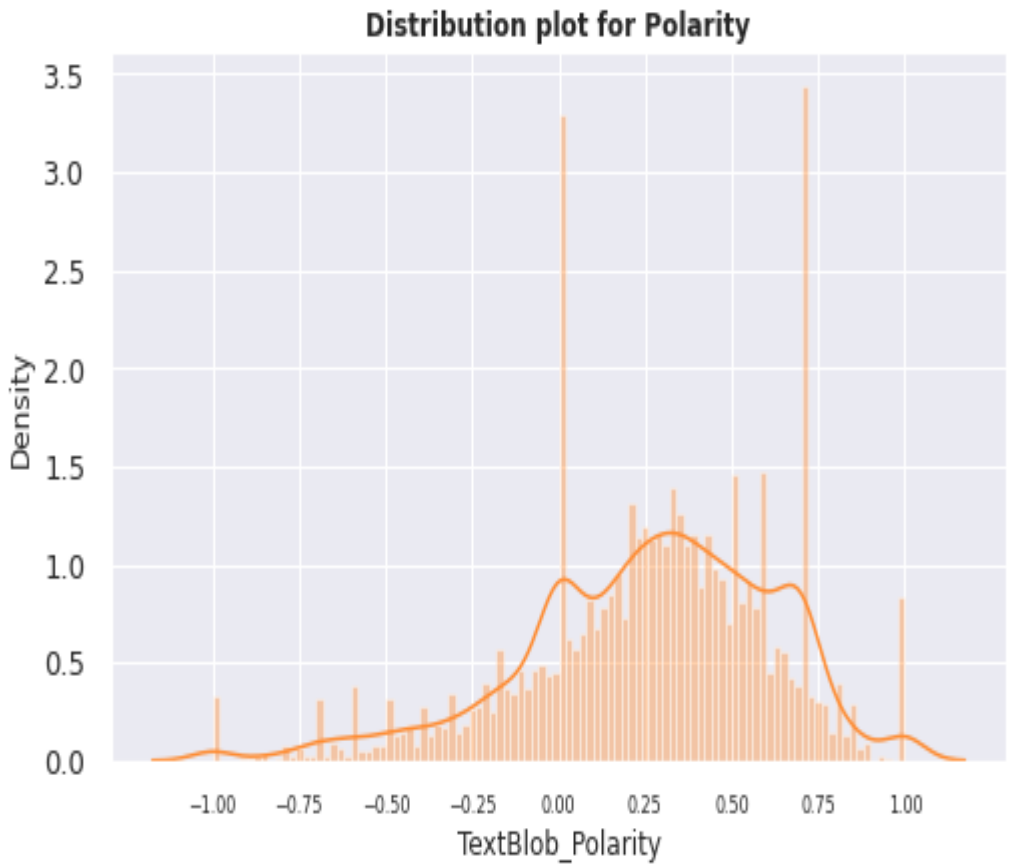


Arena Eleven was the restaurants having highest negative reviews

TextBlob Polarity

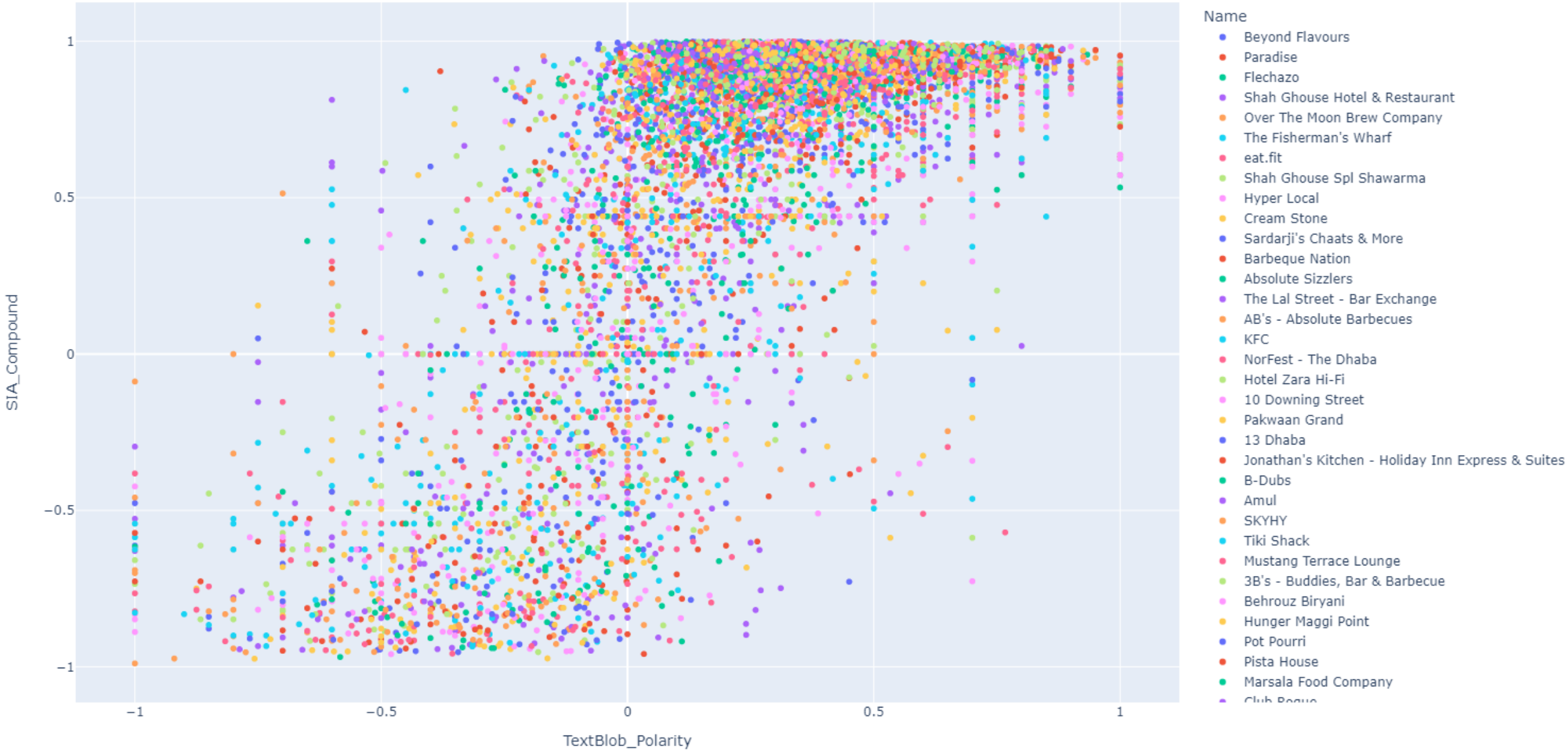


Reviews Sentiments



Distribution plot of polarity

Model Comparison



CONCLUSION

- **North Indian** Cuisine has the highest count among all
- **Collage - Hyatt Hyderabad Gachibowli** is the expensive hotel with cost of 2800
- Calculated **k = 15** as optimum value, for cuisine clusters.
- Most of the hotels are given **5 Rating**
- **Anvesh Chowdary** was the most reviewer and given many hotel rating.
- **AB's - Absolute Barbecues** has given most of the **positive reviews** and **Arens Eleven** got the **highest negative review**
- **Udipi's Upahar** is Affordable Restaurant with best ratings
- **Club Rogue** is Expensive Restaurant with worst ratings


THANK
YOU