Capstone Project-4

ZOMATO RESTAURANT CLUSTERING AND SENTIMENT ANALYSIS

by-Bharath Kumar A

Contents

- Introduction
- Problem Statement
- Dataset description
- Exploratory Data Analysis
- Treating missing values and outliers
- Natural Language Processing
- Clustering
- Sentimental analysis
- Building models
- Conclusion

Introduction

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. 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.

Problem Statement

The Project focuses on analyzing the Zomato restaurant data. You have to analyze the sentiments of the reviews given by the customer in the data and make some useful conclusion in the form of Visualizations. Also, cluster the zomato restaurants into different segments.

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.

Dataset -1 decription

Description of Zomato Restaurant names and Metadata Dataset

1. Name : Name of Restaurants
2. Links : URL Links of Restaurants
3. Cost : Per person estimated Cost of dining
4. Collection : Tagging of Restaurants with respect to Zomato categories
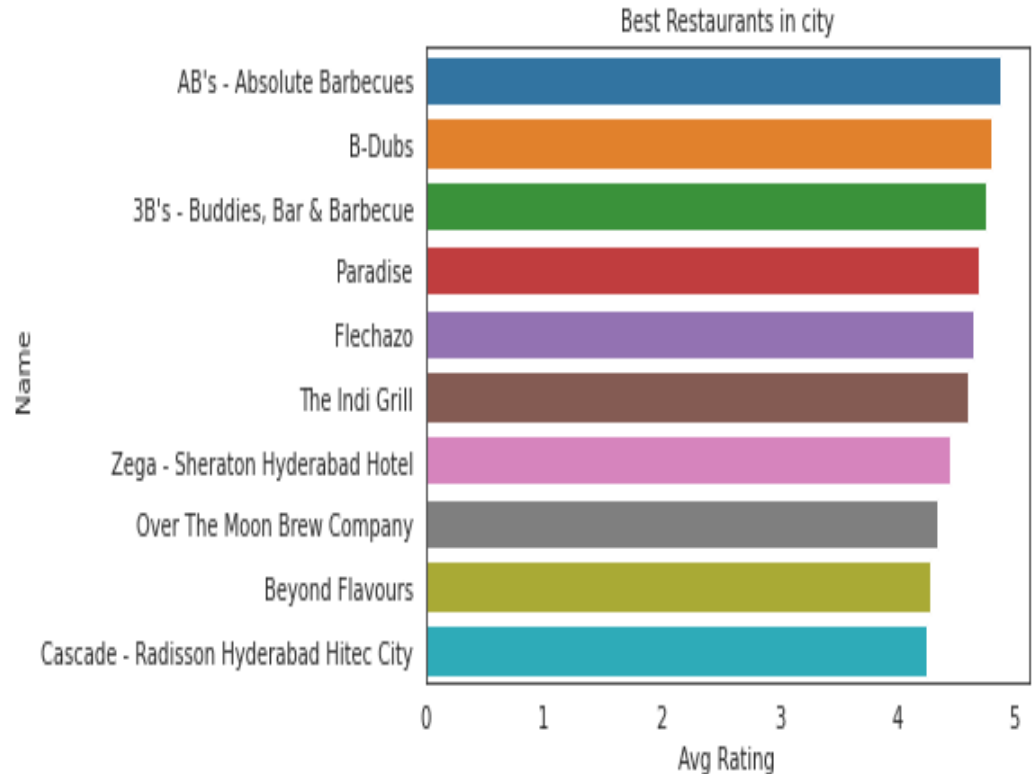5. Cuisines : Cuisines served by Restaurants
6. Timings : Restaurant Timings

Dataset -2 decription

1. Restaurant : Name of the Restaurant
2. Reviewer : Name of the Reviewer
3. Review : Review Text
4. Rating : Rating Provided by Reviewer
5. MetaData : Reviewer Metadata - No. of Reviews and followers
6. Time: Date and Time of Review
7. Pictures : No. of pictures posted with review

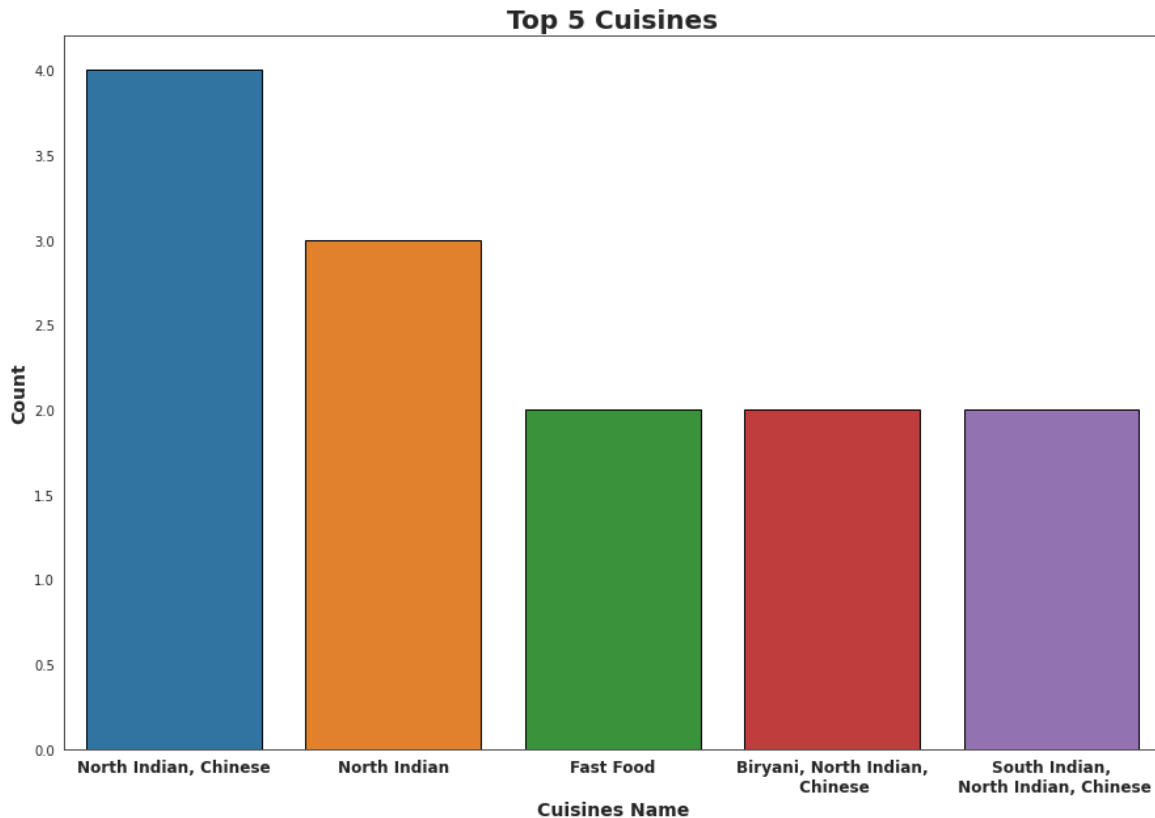# Exploratory Data Analysis

Top restaurants based on ratings

	Name	Avg Rating	Cost
0	AB's - Absolute Barbecues	4.88	1500
1	B-Dubs	4.81	1600
2	3B's - Buddies, Bar & Barbecue	4.76	1100
3	Paradise	4.70	800
4	Flechazo	4.66	1300
5	The Indi Grill	4.60	1500
6	Zega - Sheraton Hyderabad Hotel	4.45	1750
7	Over The Moon Brew Company	4.34	1200
8	Beyond Flavours	4.28	800
9	Cascade - Radisson Hyderabad Hitec City	4.26	1800

Best Restaurants in city

Top 5 cuisines

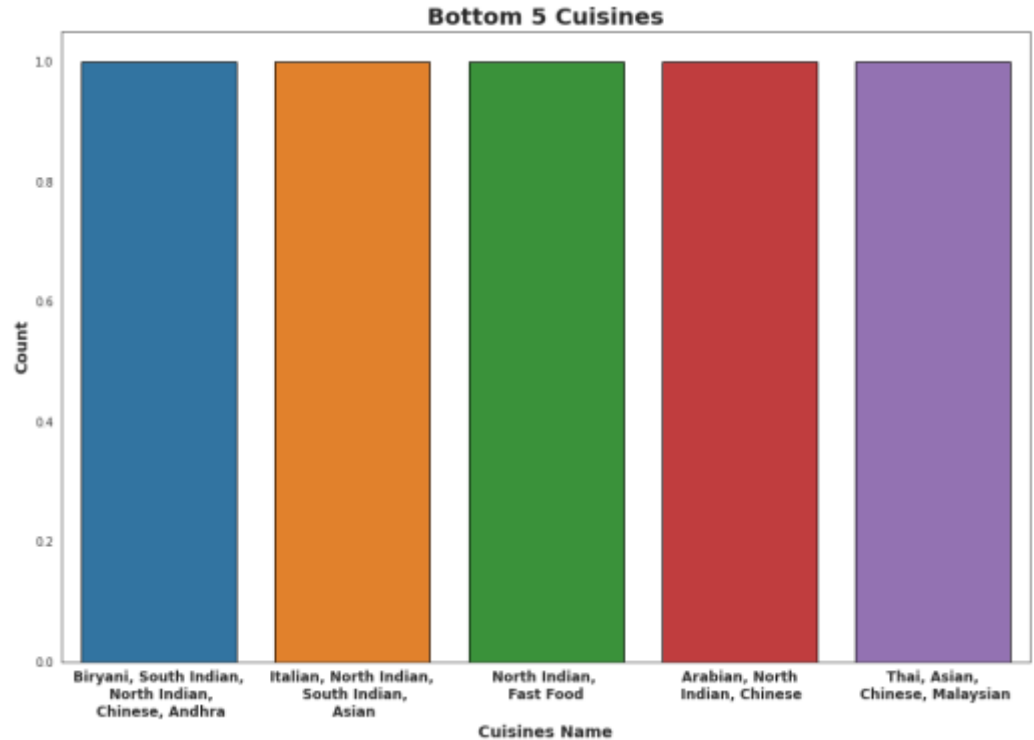|	Cuisines_name	Count
0	North Indian, Chinese	4
1	North Indian	3
2	Ice Cream, Desserts	2
3	Biryani, North Indian, Chinese	2
4	South Indian, North Indian, Chinese	2

Top 5 cuisines based on ratings

Bottom 5 cuisines

It includes Biryani, south Indian Italian, Arabian, Thai etc

	Cuisines_name	Count
87	Biryani, South Indian, North Indian, Chinese, ...	1
88	Italian, North Indian, South Indian, Asian	1
89	North Indian, Fast Food	1
90	Arabian, North Indian, Chinese	1
91	Thai, Asian, Chinese, Malaysian	1

Bottom 5 Cuisines

Top costliest and cheap restaurants

```
#top 5 costliest restaurants
rest[['Name','Cost']].sort_values(['Cost'],ascending=False).head()
```

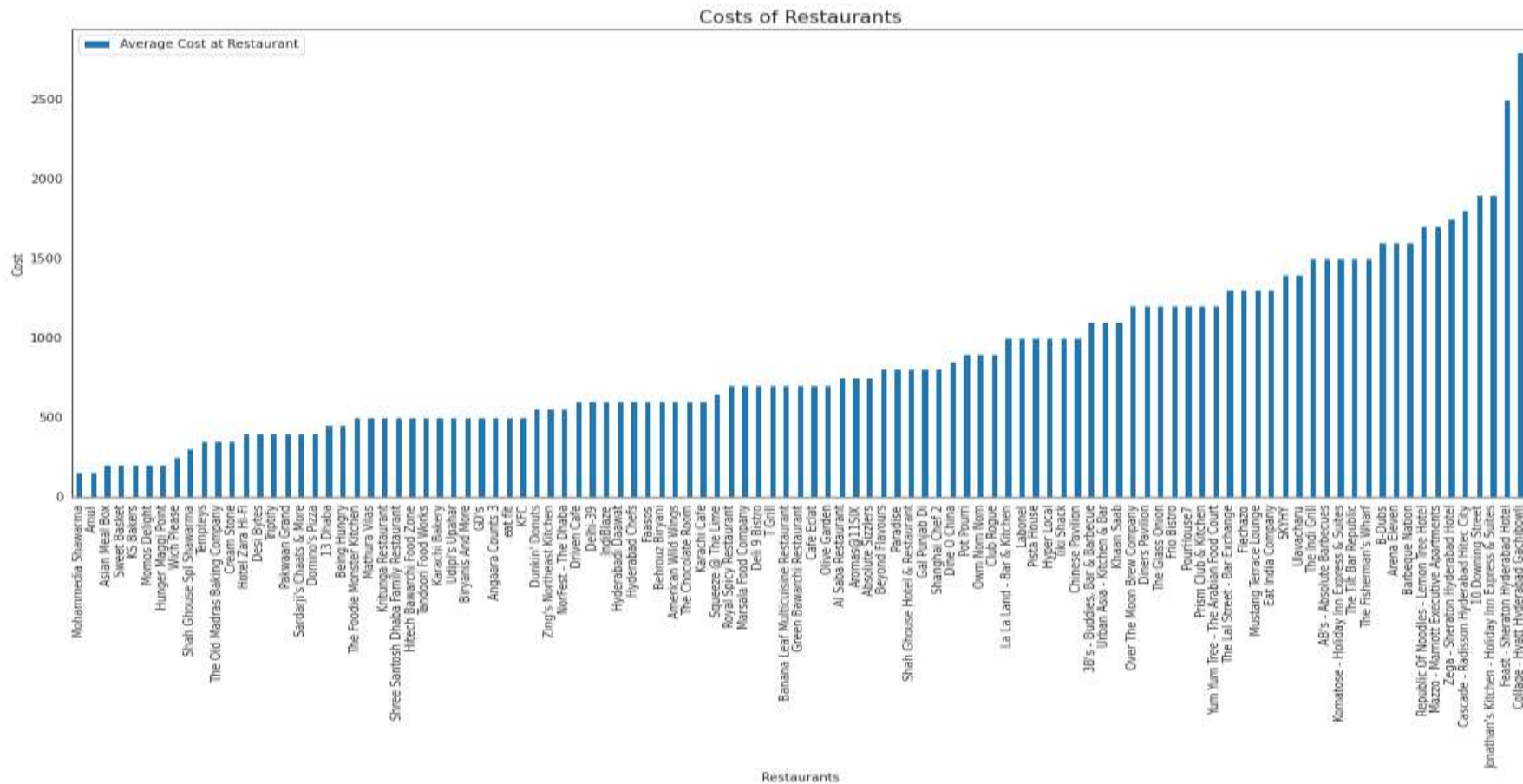|	Name	Cost
92	Collage - Hyatt Hyderabad Gachibowli	2800
56	Feast - Sheraton Hyderabad Hotel	2500
21	Jonathan's Kitchen - Holiday Inn Express & Suites	1900
18	10 Downing Street	1900
91	Cascade - Radisson Hyderabad Hitec City	1800

```
#top 5 cheapest restaurants
rest[['Name','Cost']].sort_values(['Cost']).head()
```

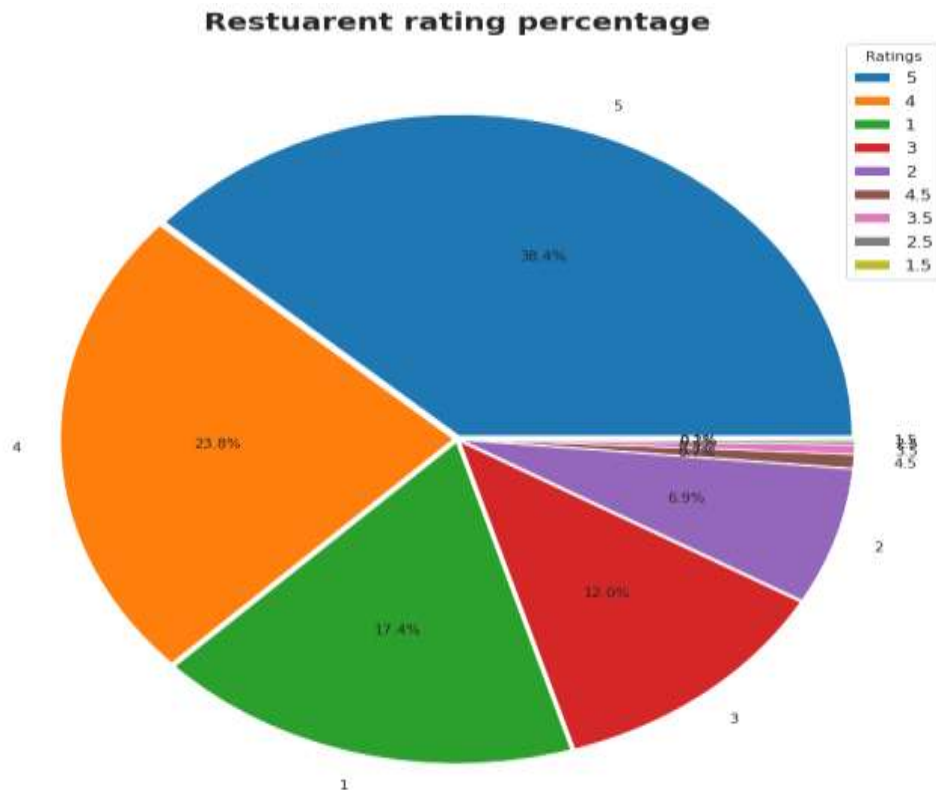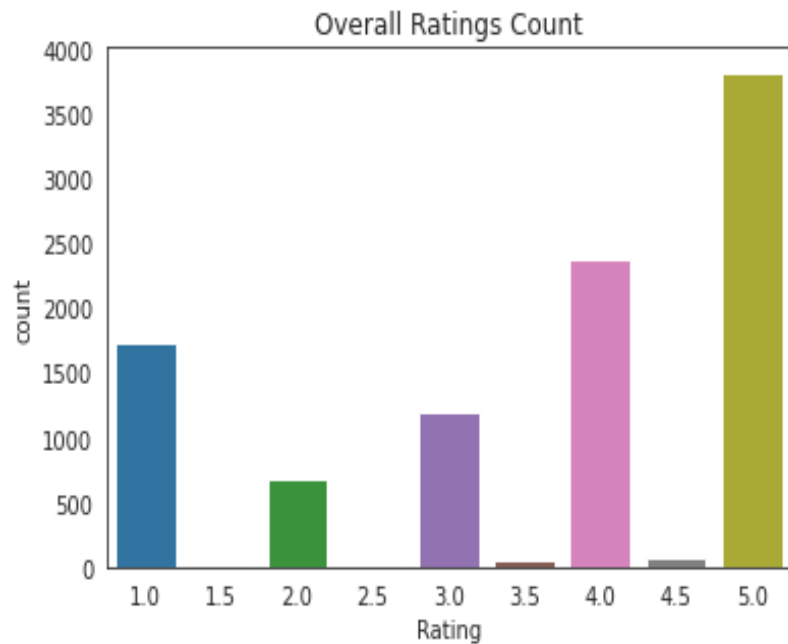|	Name	Cost
89	Mohammedia Shawarma	150
23	Amul	150
54	Asian Meal Box	200
101	Sweet Basket	200
59	KS Bakers	200

Costliest restaurants is College-Hyatt Hydrabad Gachibowli – 2800 Rs.

Cheap restaurants is the Mohammedia Shawarma -150 Rs.

Average costs of Restaurants

Costs of Restaurants

Ratings

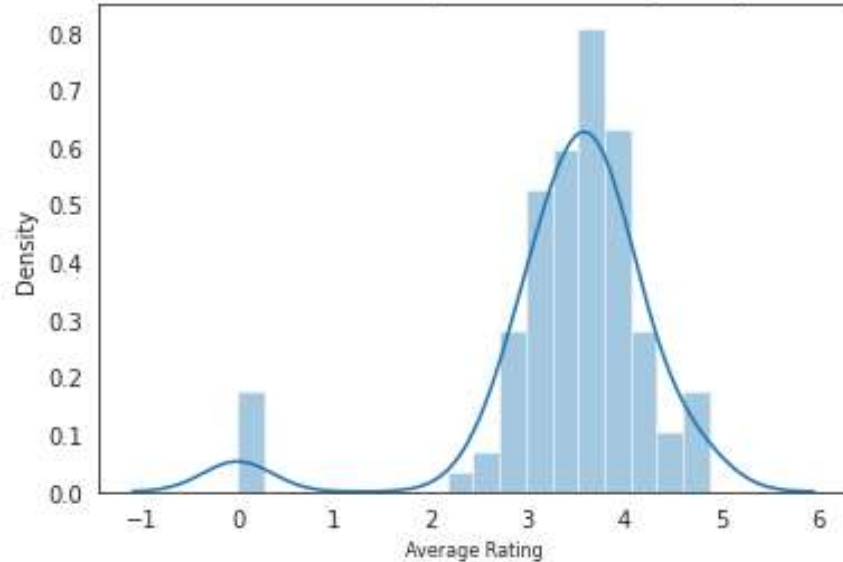# Restaurants with low ratings
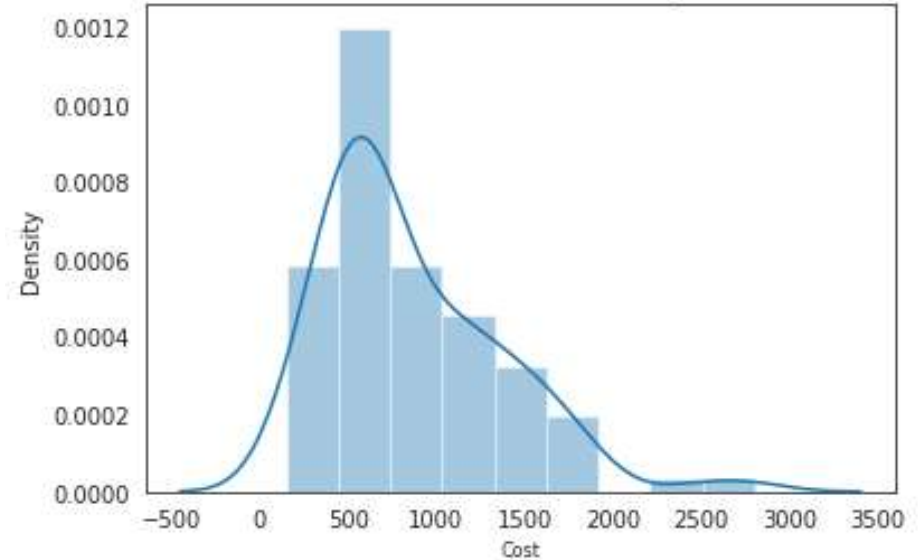

Restaurants with Low Ratings

Average distribution of ratings and cost
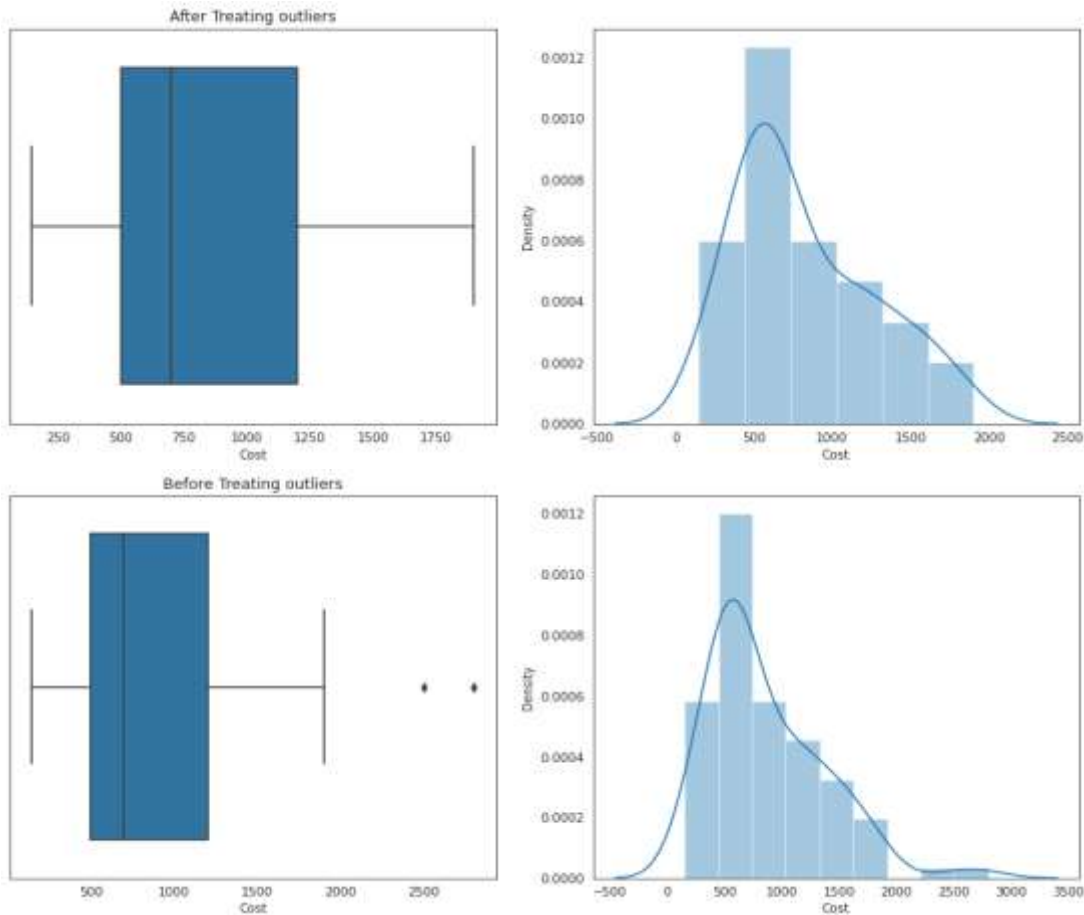

Distribution of Average Restaurant Ratings in city
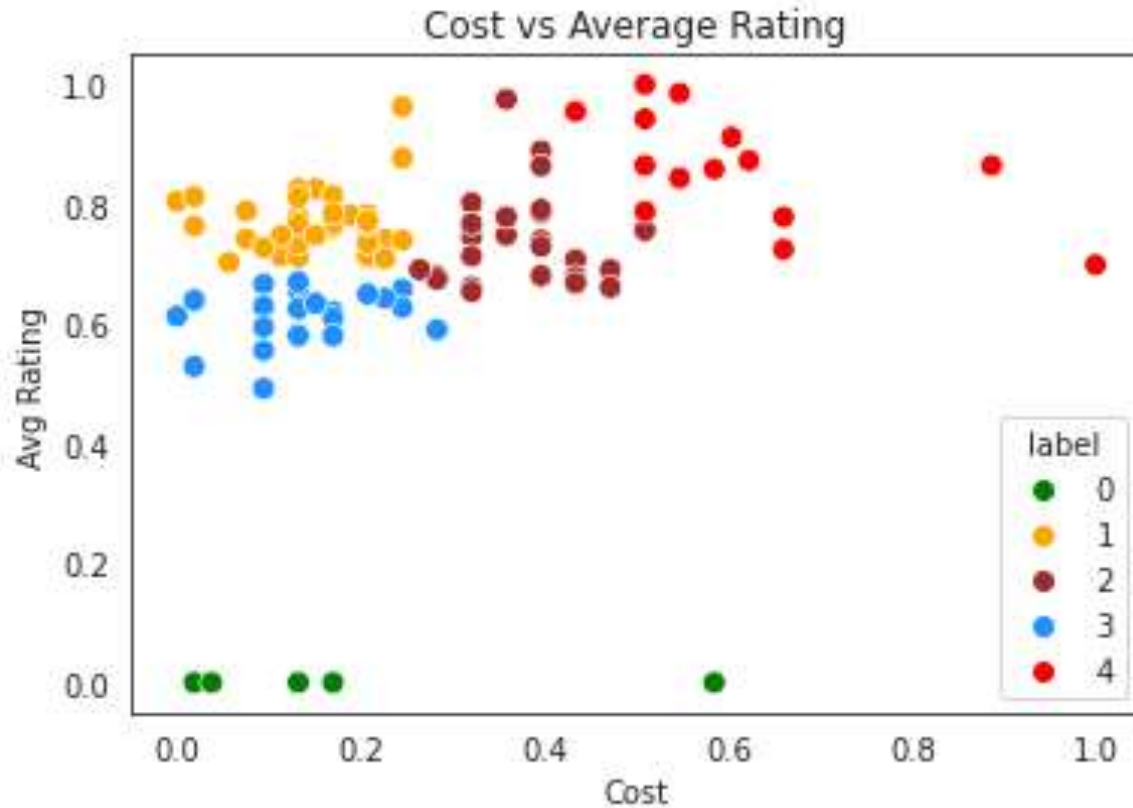
Distribution of Restaurant Costs in Hyderabad

Treating outliers in cost column

Positive and Negative words in reviews

-------------→Good words

Bad words ←------------------

Restaurant Clustering

Cluster 0:

Color: Purple

Cuisines:Fast food and Continental

Average Rating: 3.42 Average

Cost: 942 INR Median Cost: 600 INR

Cluster 1:

 Color: Red

Cuisines: North Indian and Complimentary

Average Rating: 3.63

Average Cost: 823 INR

Cluster 2:

Color: Blue

Cuisines: North Indian, Chinese and Continental

Average Rating: 3.77
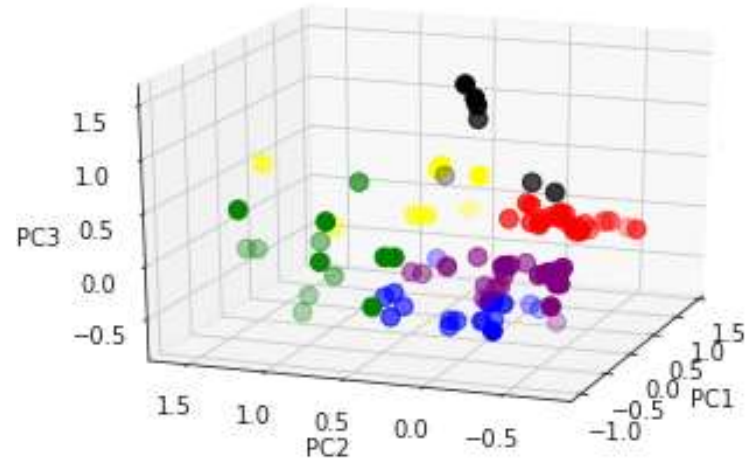
Average Cost: 1331 INR

Cluster 3:

Color: Green

Cuisines: Chinese, Thai, Asian, Malaysian etc

 Average Rating: 3.18

 Average Cost: 890 INR

Sentiment Analysis:

Evaluation: ● In the business problem, predicting the negative sentiments correctly is really important but is more important for the models to reduce the number of false positives.

```
classification_report for LogisticRegression
                precision    recall  f1-score   support

      average       0.89      0.36      0.51       948
          bad       0.89      0.90      0.90      1810
         good       0.88      0.98      0.93      4708

     accuracy                           0.88      7466
    macro avg       0.89      0.75      0.78      7466
 weighted avg       0.89      0.88      0.87      7466
```

● False positives indicate that the reviews were actually negative but they were categorized as positive and this will lead to missing a complaint to work on.

● Even though the number of false negatives is higher in the case of Logistic, it is performing better in terms of reducing False positives. This indicates that Logistic Regression is penalizing False positives more just as we want.

Conclusion

The conclusions drawn from the analysis are as follows

● The best restaurants in Hyderabad are AB's - Absolute Barbecues, B-Dubs, and 3B's - Buddies, Bar & Barbecue..
● The most popular cuisines are the cuisines which most of the restaurants are willing to provide. The most popular cuisines are North Indian, Chinese, Continental, and Hyderabadi.
● The restaurants have a flexible per person cost of 150 INR to 2800 INR. The cheapest is the food joint called Mohammedi Shawarma and the costliest restaurant is Collage - Hyatt Hyderabad Gachibowli.

Recommendations: ● Restaurants with negative reviews should be worked with in order to arrive at a good situation.
● Ratings should be collected on a category basis such as rating for packaging, delivery, taste, quality, quantity, service, etc. This would help in targeting specific fields.

Thank You