Never have a bad meal

ZOMATO RESTAURANTS REVIEW ANALYSIS

Subham Sekhar Sarangi

Table Of Content

About Zomato

- Zomato is a well-known Indian multinational corporation that mostly serves the restaurant and culinary sectors. Zomato, which was founded in 2008 by Deepinder Goyal and Pankaj Chaddah, has become one of the top websites for finding restaurants and ordering food online.

- The goal of Zomato is to make sure that no one eats a terrible meal.

- Users can find restaurants, cafes, and diners using Zomato's platform based on location, cuisine, ratings, and reviews.

- Zomato is present in several nations throughout the world, including India, the US, the UK, Australia, the UAE, and many others.

Business Understanding

01 Business Problem

Gaining useful information from customer reviews of restaurants on Zomato is a business challenge.

02 Objective

Giving restaurant owners feedback on the calibre of their cuisine and service

03 Expectation

Improve the usability and suggestion precision of its platform

BUSINESS UNDERSTANDING

Sentiment Analysis & Customer Satisfaction

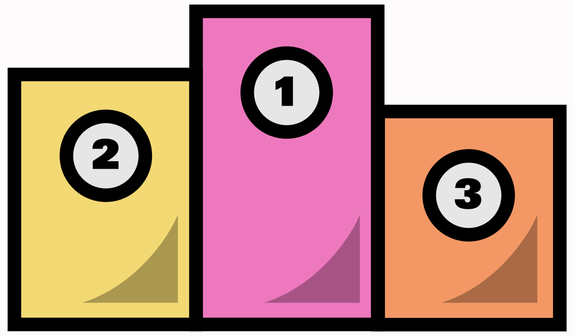
-Phrases
“Delicious & Must Try”

Feature Importance Analysis

Factors:
Ambiance, Service, Food Quality

Time-Series Analysis

Season of High and Low Demand

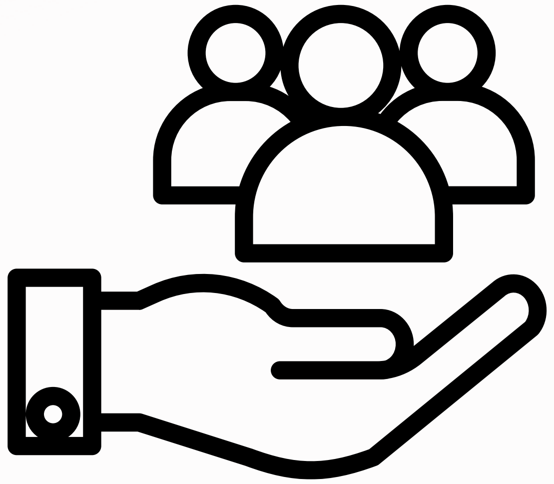


Competitive Analysis

Benchmarking to compare performance

Menu and Offering Analysis

Highlighting and Promoting Specific Dishes



Customer Segmentation

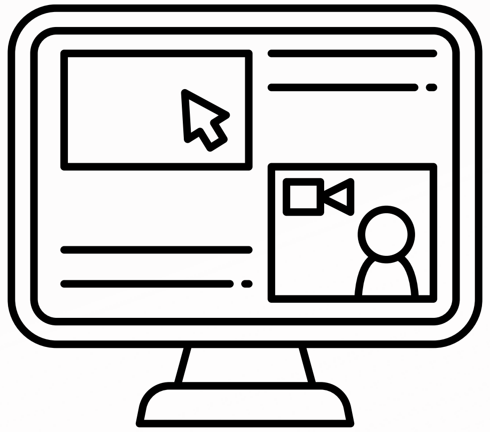
Loyalty Program, Tailored Promotion

Service Improvement Analysis

Service Quality, Staff Behaviour

Pricing, value Analysis and Geographical insights

Implementation of Discount & promotional strategy



Online Presence and Engagement

Follower count, social media engagerment

DATA UNDERSTANDING

Among this dataset's important characteristics are:

Restaurant: Specifics regarding the eatery under review.

Reviewer: Information regarding the person or thing who is writing t review.

Review: The review's textual material.

Rating: The restaurant's numerical rating or score.

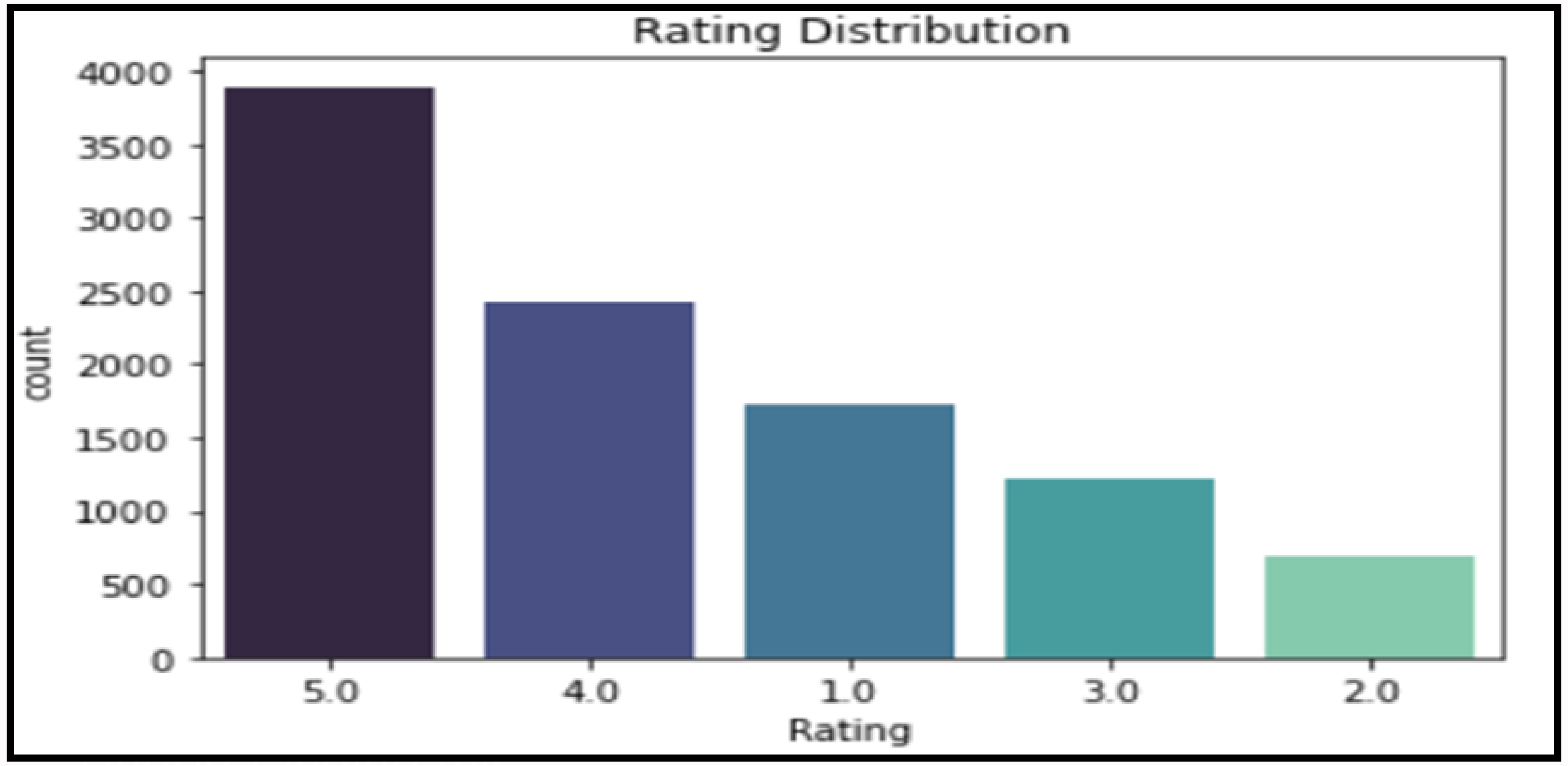
Metadata: Additional details regarding the review, such as the time place as well as other pertinent information.

Time: Timestamp reflecting the posting date of the review.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 4 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Restaurant  10000 non-null  object
 1   Reviewer    9962 non-null   object
 2   Review      9955 non-null   object
 3   Rating      9962 non-null   object
dtypes: object(4)
memory usage: 312.6+ KB
```

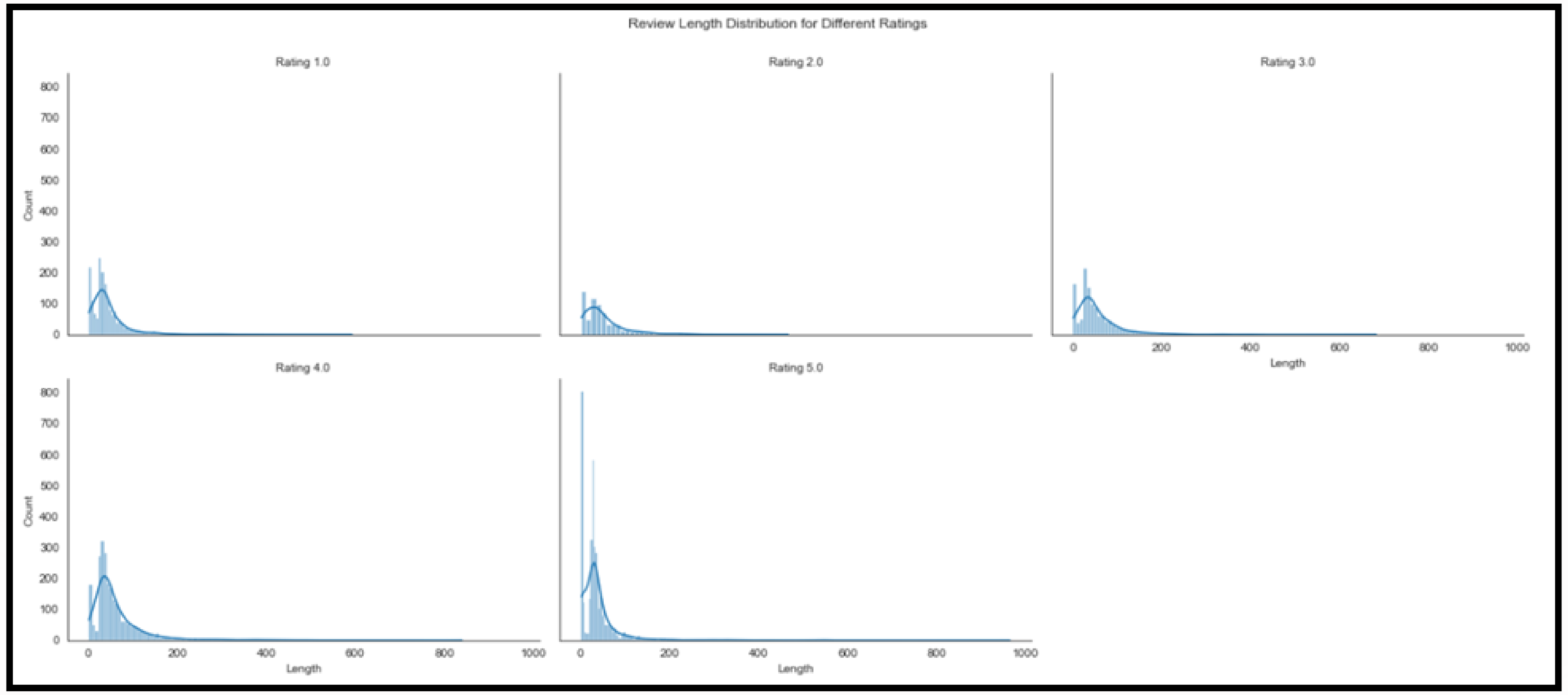
Data Visualization

Visualization of Target Variable



understanding the distribution can be invaluable for restaurant owners and analysts, as it provides insights into customer sentiment and can inform

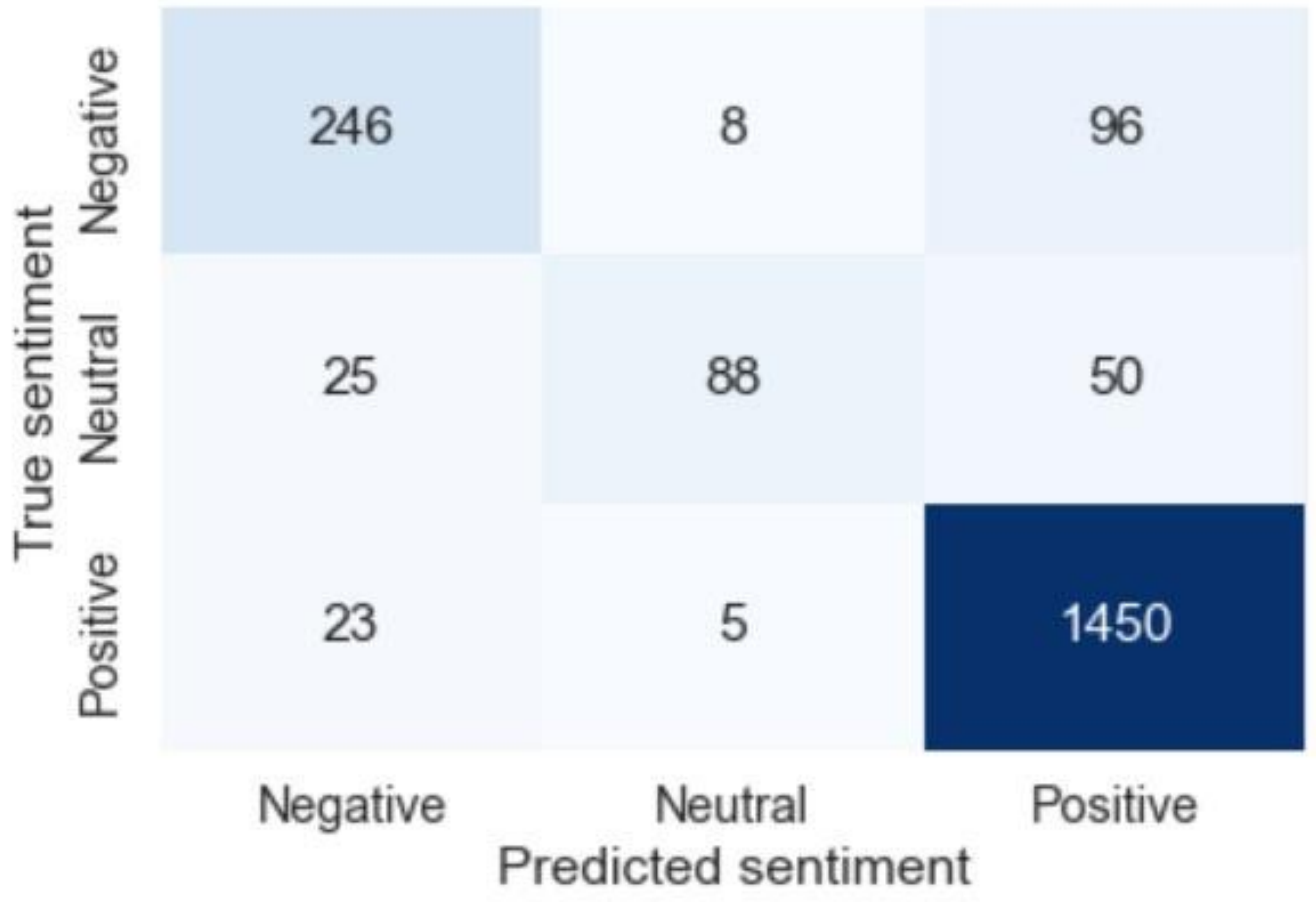
Review Length Distribution for Different Ratings



The spike is highest for rating 5, which means that people who are rating 5 stars

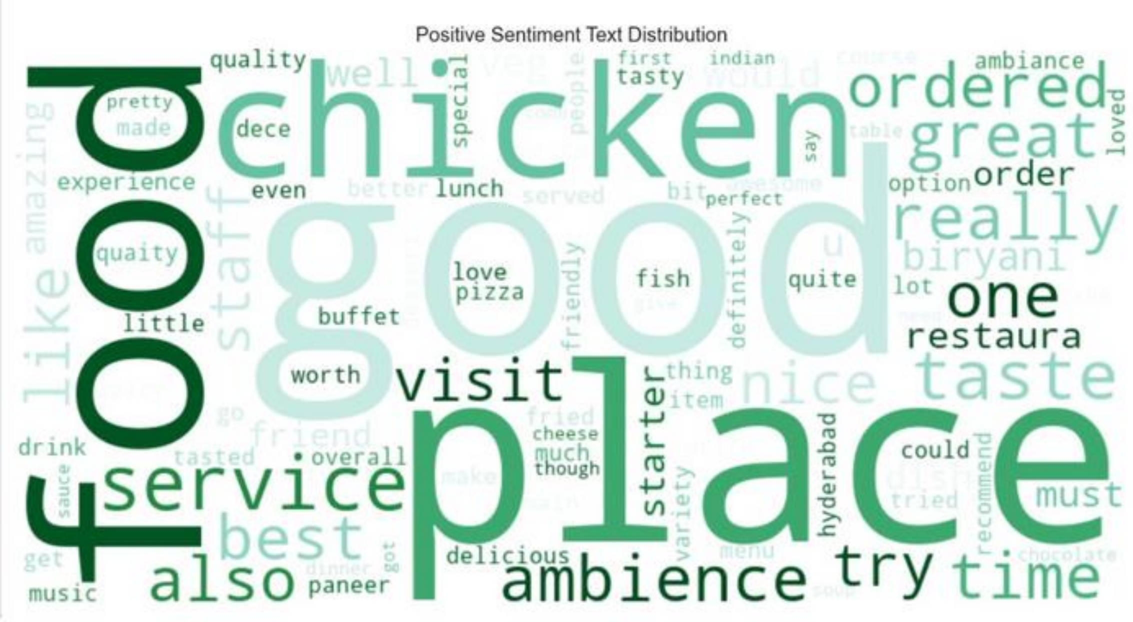
Sentiment Analysis-VADER approach & NRClex

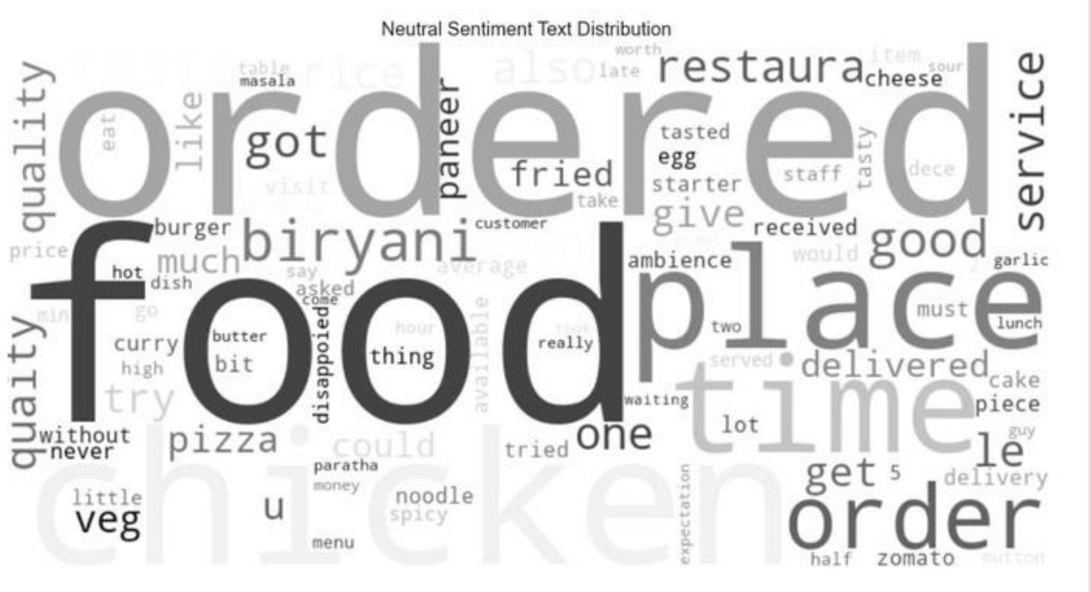
By contrasting the predicted labels with the actual labels in a collection of data, we may see how effectively your model is categorizing feelings (positive, negative, and neutral). Here is how a confusion matrix can be explained in relation to Zomato restaurant reviews:



- **True Positive (TP) :** Identified a review that said, "Thisrestaurant is amazing!" as being genuinely positive
- **Truly negative (TN):** Identified a review that stated, "The food was terrible" as negative
- **False Positive (FP) :** A false positive would occur, for instance, if a review stated, "The service was terrible," but your model misclassified it as positive
- **False Negative (FN) :** A false negative, for example, would be if a review stated, "The ambiance was great," but your model misclassified it as negative

Word Cloud





- **Positive reviews** - Compliments on the restaurant's ambiance,service, and food quality and suggest visiting
- **Negative Reviews** - Unfavorable evaluations may draw attention to problems like wrong orders, inadequate food quality, subpar service, or an overall unpleasant experience
- **Neutral Reviews** - Potential consumers who are looking for real information to make their dining decisions can learn more about what was ordered, how it was served, and how certain meals tasted

CLASSIFICATION (USE CASE & APPLICATION)

- Review classification can be used to leverage customer feedback for continuous improvement, enhanced customer satisfaction, and a competitive edge
- The classification of reviews allows restaurants to systematically categorize customer feedback into positive, negative, or neutral sentiments
- Quickly identify and respond to customer feedback and potential issues.
- Based on the sentiment classification results, develop strategies for **Quality Improvement, Performance Benchmarking, gain Marketing Insights, Strategic Decision-Making, to respond to customer feedback and gaining competitive advantage**
- **Positive sentiments:** Engage with positive feedback and thank customers for their support.
- **Negative sentiments:** Address negative feedback promptly, resolve issues, and demonstrate commitment to customer satisfaction.
- **Neutral sentiments:** Monitor for potential issues or areas where customer engagement might be improved.

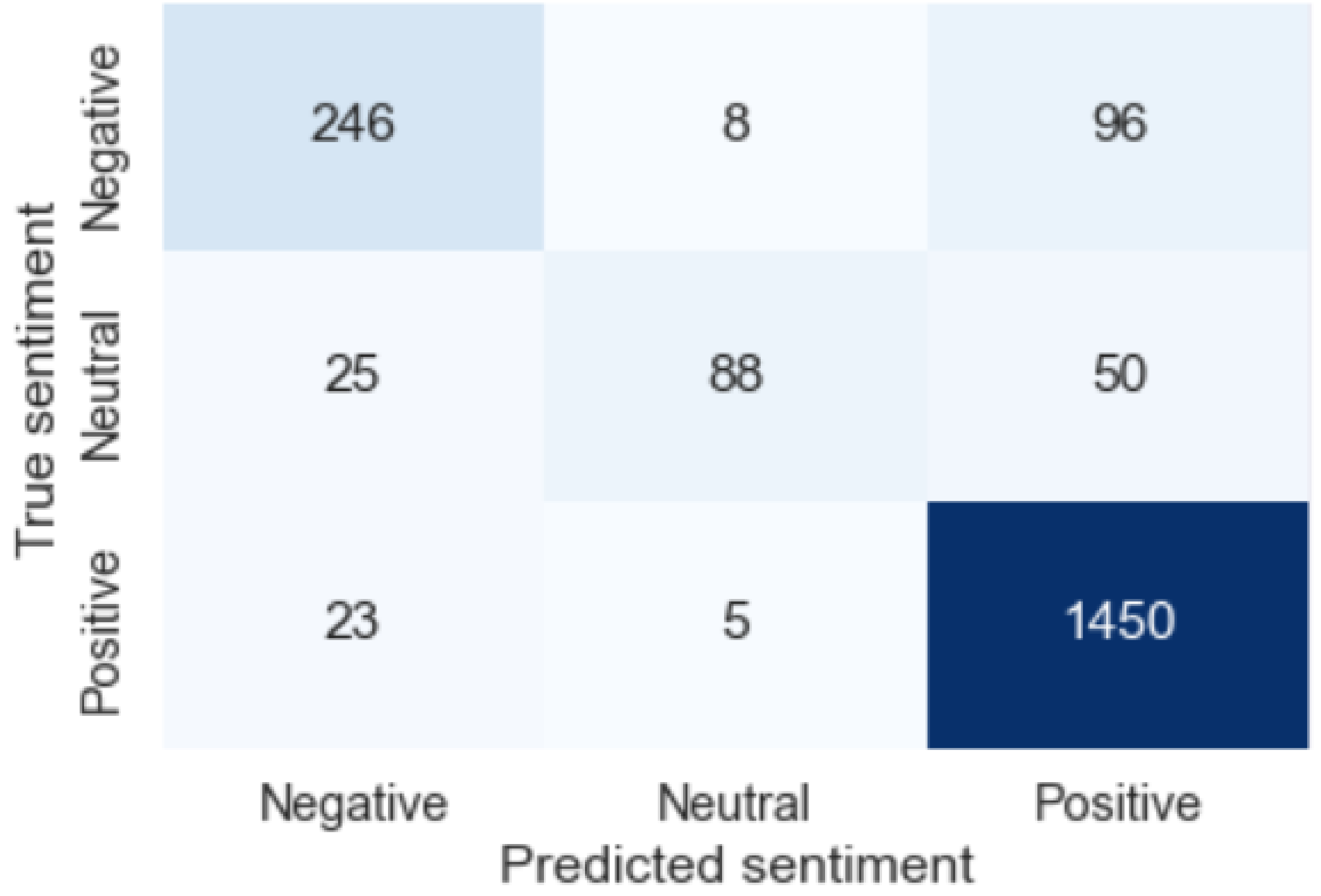
CLASSIFICATION

1. Multinomial Logistic Regression

Accuracy: 0.8960321446509292

Classification Report:

	precision	recall	f1-score	support
negative	0.84	0.70	0.76	350
neutral	0.87	0.54	0.67	163
positive	0.91	0.98	0.94	1478
accuracy			0.90	1991
macro avg	0.87	0.74	0.79	1991
weighted avg	0.89	0.90	0.89	1991



CLASSIFICATION

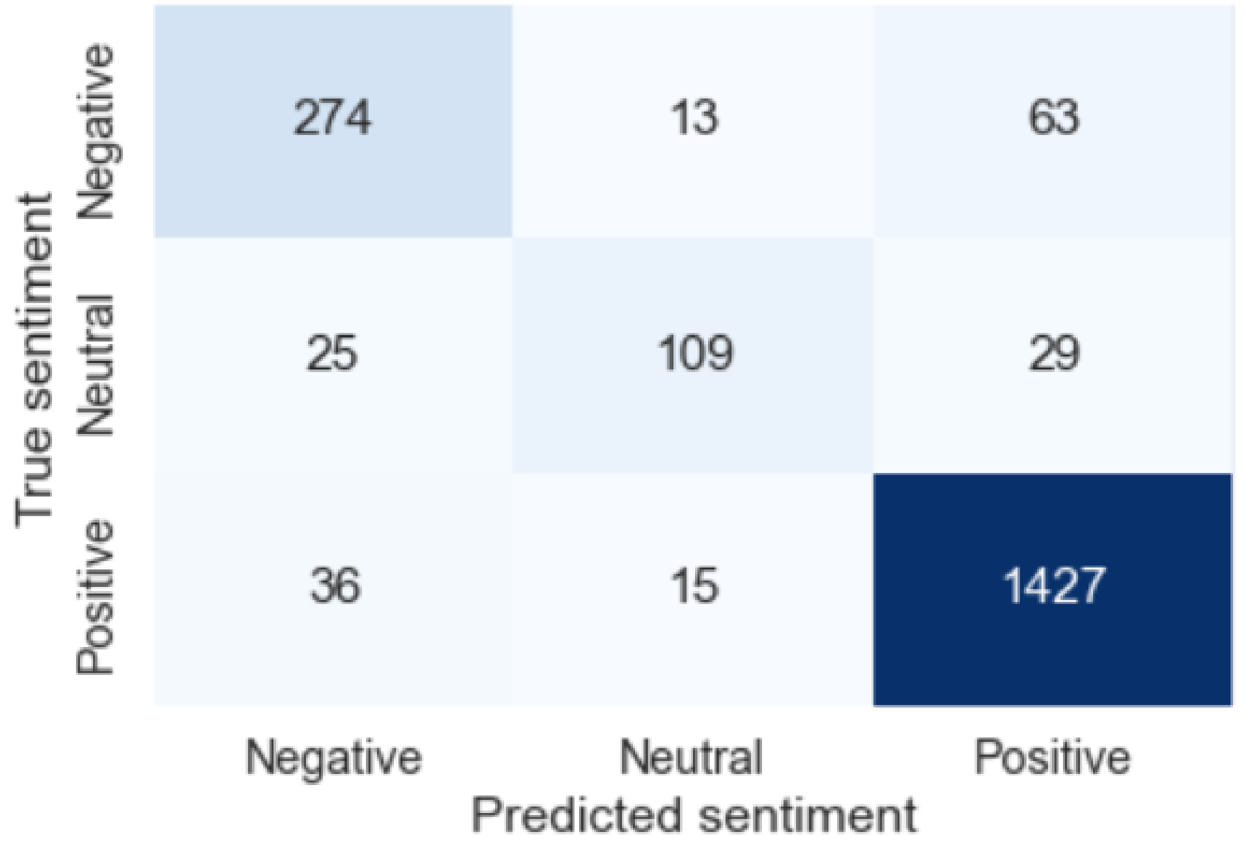
2. Linear Support Vector Classification

```
Accuracy: 0.9090909090909091


Classification Report:
              precision    recall  f1-score   support

    negative       0.82      0.78      0.80       350
     neutral       0.80      0.67      0.73       163
    positive       0.94      0.97      0.95      1478

    accuracy                           0.91      1991
   macro avg       0.85      0.81      0.83      1991
weighted avg       0.91      0.91      0.91      1991
```



CLASSIFICATION

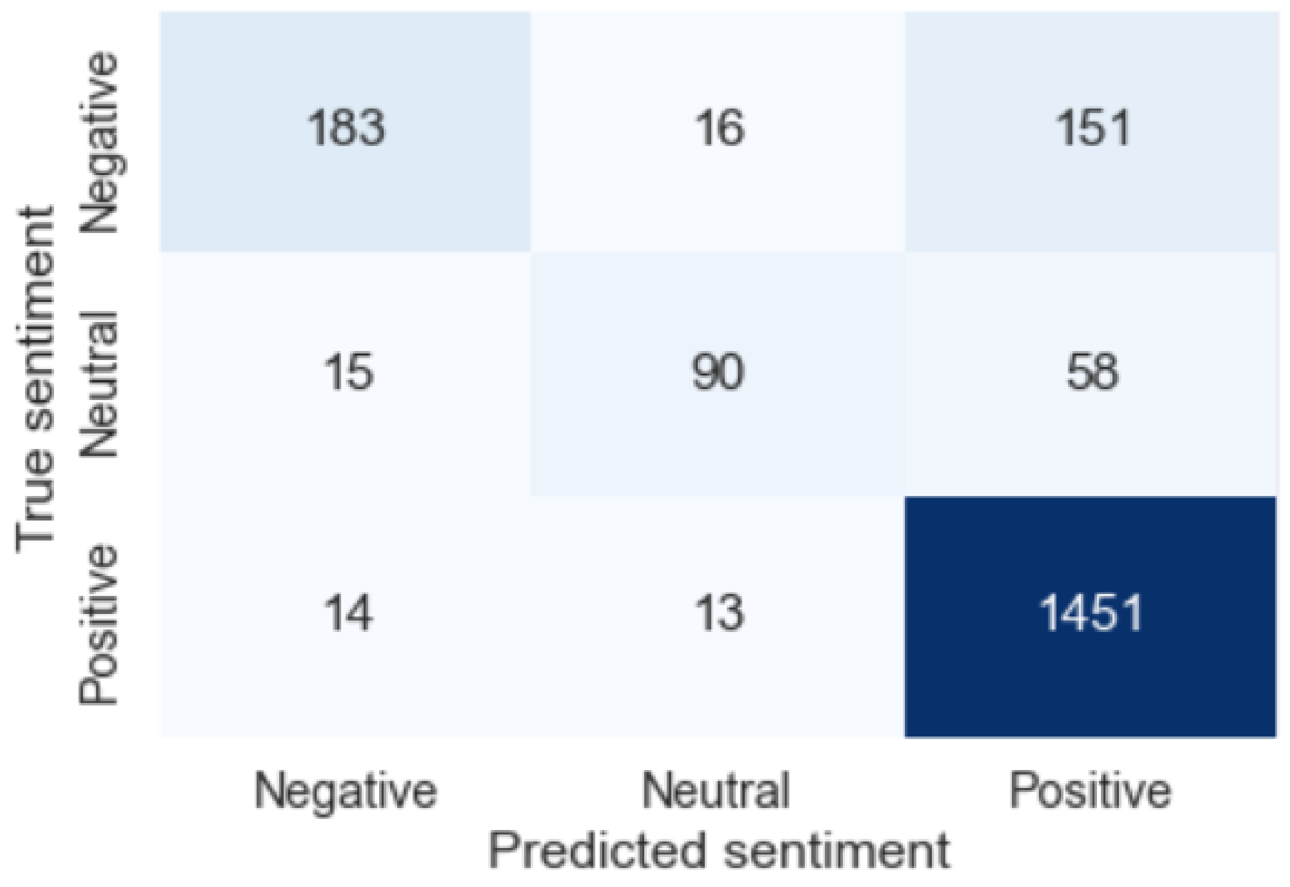
3. Random Forest

```
Accuracy: 0.8658965344048217


Classification Report:
              precision    recall  f1-score   support

    negative       0.86      0.52      0.65       350
     neutral       0.76      0.55      0.64       163
    positive       0.87      0.98      0.92      1478

    accuracy                           0.87      1991
   macro avg       0.83      0.69      0.74      1991
weighted avg       0.86      0.87      0.85      1991
```



CLASSIFICATION

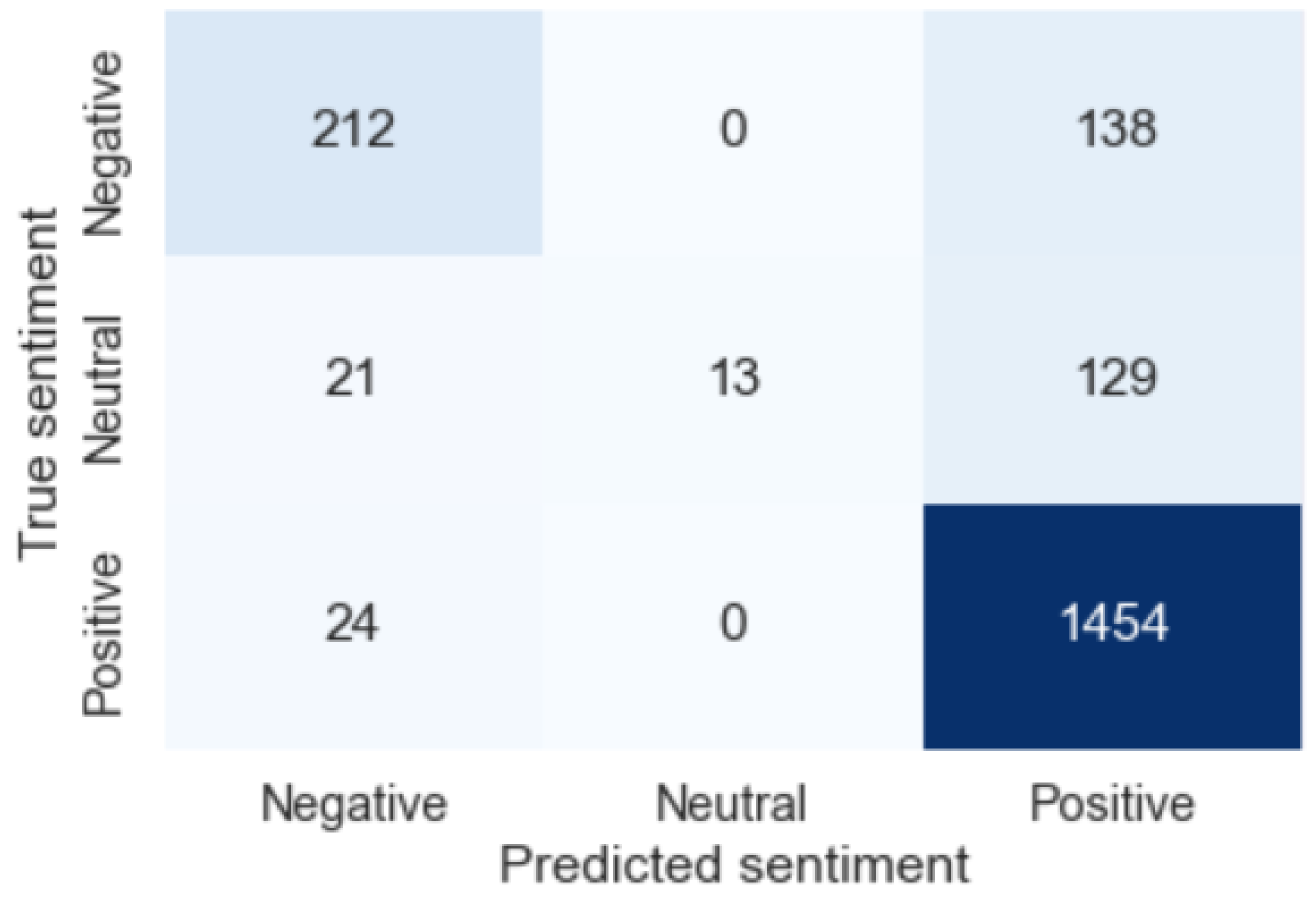
4. Multinomial Naive Bayers

```
Accuracy: 0.843294826720241


Classification Report:
              precision    recall  f1-score   support

    negative       0.82      0.61      0.70       350
     neutral       1.00      0.08      0.15       163
    positive       0.84      0.98      0.91      1478

    accuracy                           0.84      1991
   macro avg       0.89      0.56      0.59      1991
weighted avg       0.85      0.84      0.81      1991
```



CLASSIFICATION

Comparison of all models

Models	Accuracy_score
SVC	0.909091
Logistic Regression	0.896032
Random Forest	0.865897
Naive Bayes Multinominal	0.843295

Among all the models, Multinomial Support vector classification has the highest accuracy score/Hit Ratio. The best performer is SVC in terms of overall sentiment and also individual sentiments.
Among all the 4 models, NB is computationally efficient but has the lowest accuracy score.

TOPIC MODELLING

Data Preparation Before Topic Modeling

- Regular expressions
- Stop words Removal
- Tokenization
- Text Representation/Vectorization

CountVectorizer(ngram_range = (1,2), max_features=1000, max_df=0.5)

TOPIC MODELLING

```
Topic 1: food, order, ordered, bad, time, worst, delivery, restaurant, even, service
Topic 2: place, food, good, service, one, ambience, staff, would, menu, go
Topic 3: chicken, good, biryani, ordered, taste, rice, food, quantity, spicy, paneer
Topic 4: veg, starters, good, main, course, place, main course, one, cream, ice
Topic 5: good, food, place, service, great, ambience, nice, visit, staff, really
```

Topic 1 - Negative Customer Feedback
Business Use Case: Restaurants can use this topic to identify specific issues that lead to negative reviews and take corrective actions. They can analyze reviews in this topic to pinpoint common problems and work on improving food quality, delivery efficiency, and overall service to enhance customer satisfaction and reduce negative feedback.

Topic 2 - Dining Experience
Business Use Case: Restaurants can utilize this topic to understand what contributes to positive dining experiences and maintain those aspects. Identify the most frequently mentioned positive attributes (e.g., good food, friendly staff, appealing ambiance) and highlight them in marketing efforts to attract more customers.

Topic 3 - Spicy Dishes
Business Use Case: Restaurants can use this topic to identify which menu items are popular or need improvement. Analyze feedback related to specific dishes in this topic to fine-tune recipes, portion sizes, and spice levels to better meet customer preferences.

Topic 4 - Menu Exploration
Business Use Case Restaurants can use this topic to understand customers' preferences for different parts of their menu. Analyze feedback on menu items in this topic to optimize the menu offerings, ensure variety, and create appealing vegetarian options to cater to diverse customer tastes.

Topic 5 - Positive Dining Experience
Business Use Case: Restaurants can leverage this topic to identify and celebrate their strengths. Use the positive feedback in this topic for marketing purposes, showcasing the restaurant's strengths in advertising and promotions to attract more customers and enhance its reputation.

Topic 1 - Word Cloud

Topic 2 - Word Cloud

Topic 3 - Word Cloud

4 - Word Cloud

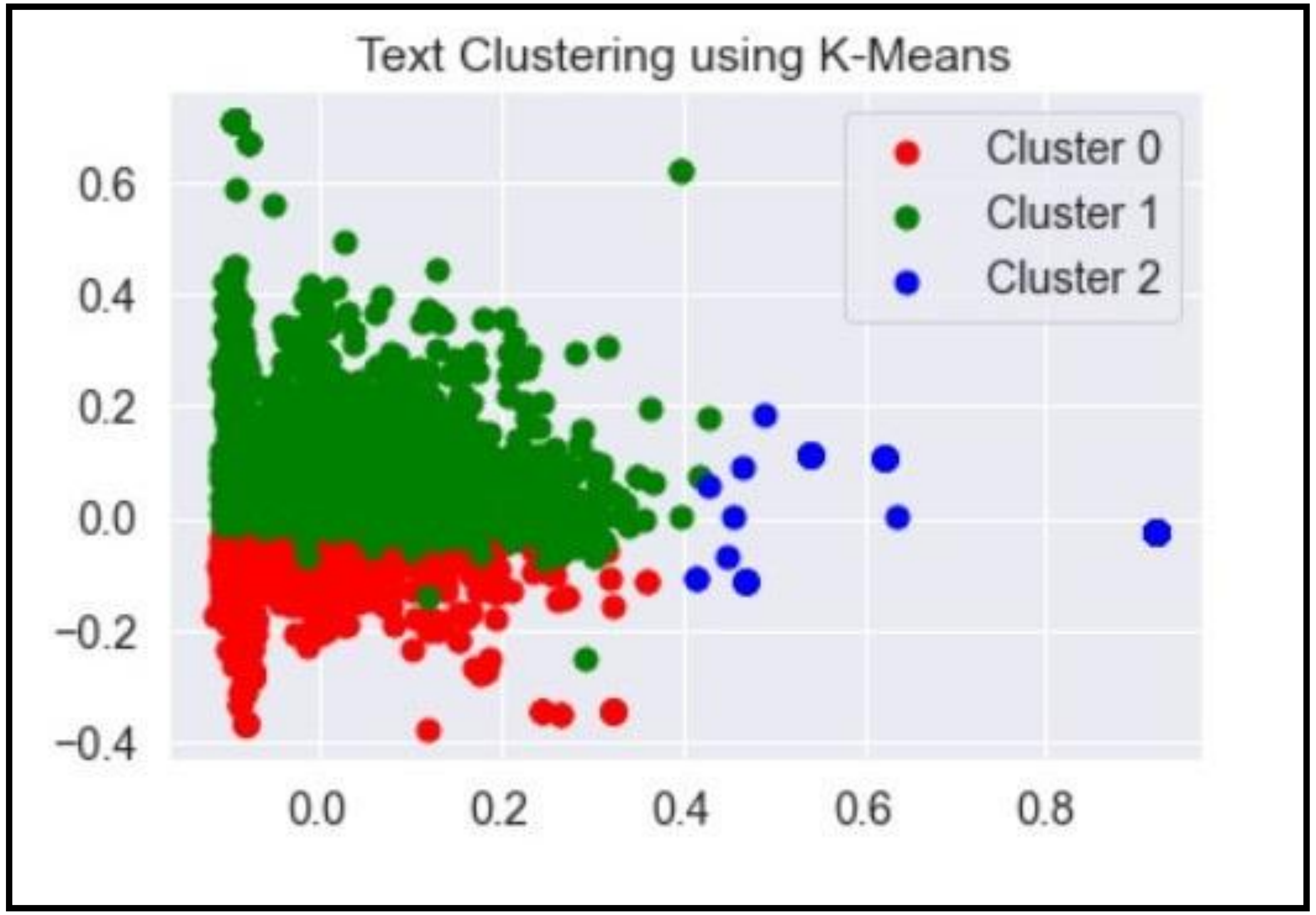
Topic 5 - Word Cloud

CLUSTERING

Text clustering algorithm(K-means) is used to group similar reviews into clusters based on the similarity of their contents.

The goal of clustering is to identify underlying patterns or structures in the data, which can be useful for recommendation systems, for example improving customer service

The **ARI of 0.0195** indicated a very weak positive correlation between the true labels (the actual ratings of the reviews) and the predicted labels (the cluster labels assigned by the K-Means algorithm).

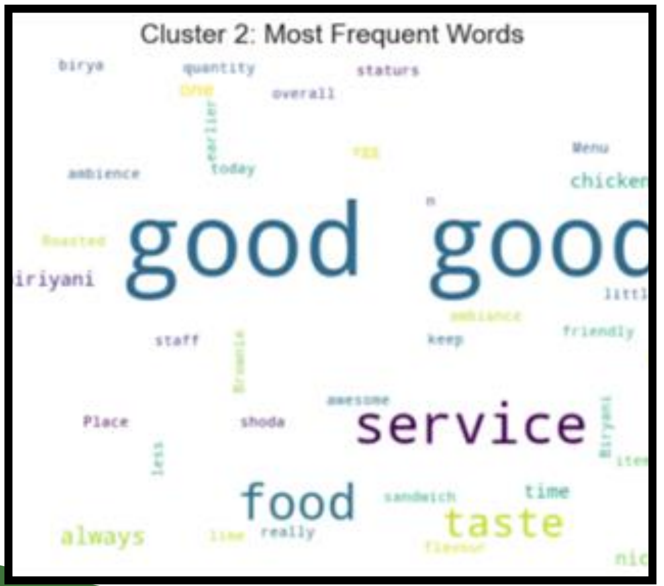



```
Cluster 1 top terms: ['chicken', 'good', 'food', 'taste', 'ordered', 'delivery', 'biryani', 'time', 'order', 'bad']
-------
Cluster 2 top terms: ['good', 'food', 'place', 'nice', 'service', 'great', 'ambience', 'staff', 'visit', 'really']
-------
Cluster 3 top terms: ['good', 'service', 'food', 'taste', 'time', 'earlier', 'little', 'quantity', 'biriyani', 'today']
```

CLUSTERING

Cluster 1 - Food Quality and Delivery:
Top Terms: ['chicken', 'good', 'food', 'taste', 'ordered', 'delivery', 'biryani', 'time', 'order', 'bad']
Centred around aspects related to food quality & delivery service
Business Use Case: To gain insights into customer sentiments about food quality & efficiency of their delivery service

Cluster 2 - Overall Dining Experience:
Top Terms: ['good', 'food', 'place', 'nice', 'service', 'great', 'ambience', 'staff', 'visit', 'really']
Revolve around the overall dining experience at the restaurant.
Business Use Case: To understand what factors contribute to positive dining experiences and reinforce them. Use for for marketing purposes



Cluster 3 - Service and Quantity Concerns:
Top Terms: ['good', 'service', 'food', 'taste', 'time', 'earlier', 'little', 'quantity', 'biriyani', 'today']
focus on service-related aspects and quantity concerns.
·**Business Use Case:** To address specific service and quantity-related feedback, to identify areas where service improvements are needed

The recommended/optimal number of clusters for K-Means clustering is **3** (silhouette score)

CONCLUSION

- Our exploration encompassed a wide range of analytical dimensions, including sentiment analysis, feature importance assessment, time-series examination, competitive landscape evaluation, menu analysis, customer segmentation, service enhancement identification, pricing and value perception analysis, geographical insights, and online engagement assessment.

AIM

PROCESS

RESULT

- In this extensive analysis of Zomato restaurant reviews, we embarked on a multifaceted journey to extract valuable insights into the diverse dining experiences of customers.
- Our aim was to empower restaurant owners and industry stakeholders with actionable knowledge.

- We found keywords such as "Delicious" and "Must-try" signified positive sentiments, while terms like "worst" and "bad" indicated negativity.
- Key factors like "ambiance," "service," "food quality," and "value for money" significantly influenced ratings, highlighting the importance of focusing on these areas for restaurants.
- Additionally, analyzing the temporal trends in reviews and delving into competitive dynamics provided strategic advantages.
- Furthermore, insights into menu preferences, customer segmentation, and avenues for service improvement offered avenues for enhancing customer satisfaction and loyalty.

Fauget Delivery

Thank You