Capstone Project - 4

Zomato Clustering & Sentiment Analysis

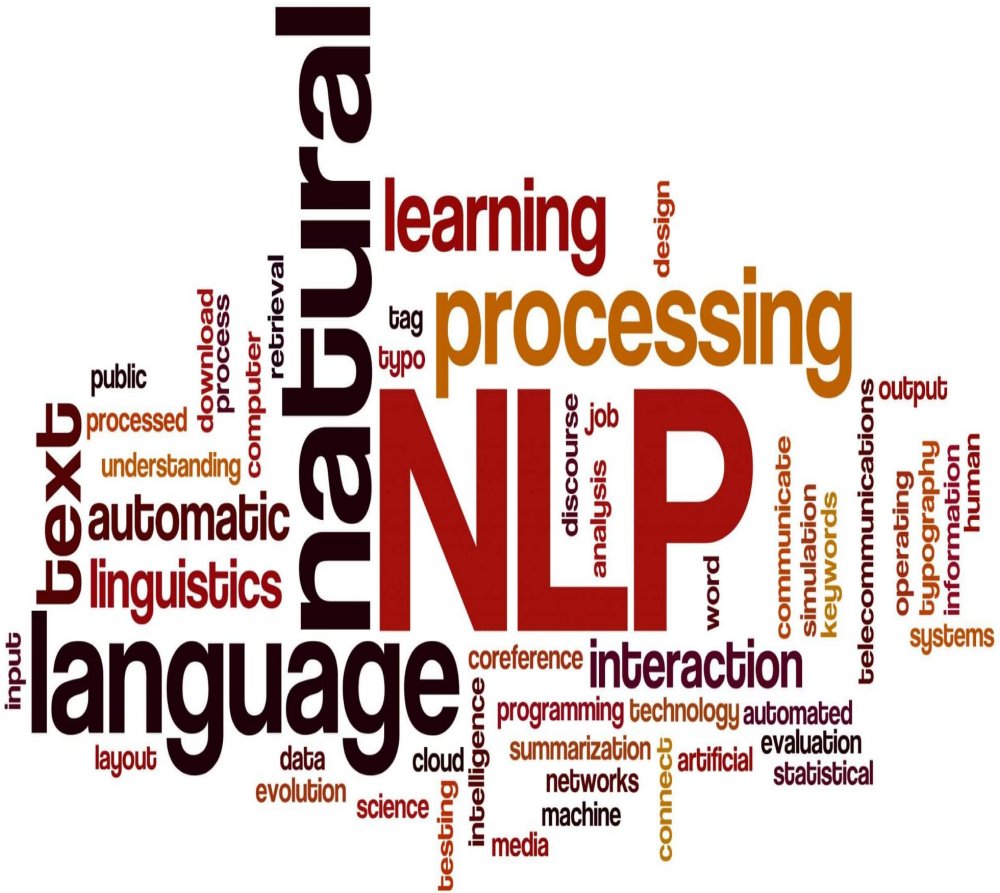
Team Members:

Sharad Tawade

Sagar Malik

Vinay Kumar

Contents:

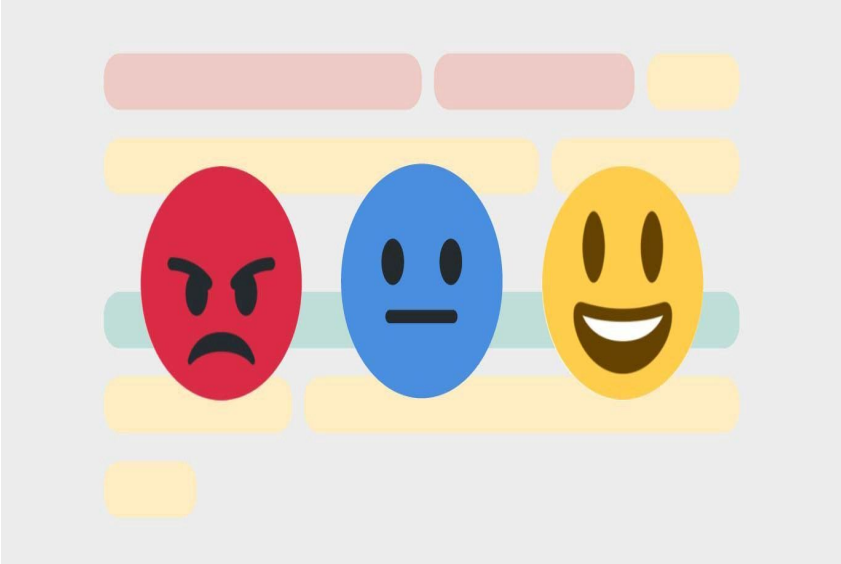


Need for Sentiment Analysis?

- Sentiment analysis is a powerful marketing tool that enables product managers to understand customer emotions in their marketing campaigns.

- The term emotion-based marketing is a broad umbrella phrase that encompasses emotional customer responses, such as "positive," "negative," "neutral," "negative," "uptight," "disgust," "frustration," and others.

- Understanding the psychology of customer responses can also increase product and brand recall.



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.

- 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 analysing the Zomato restaurant data for each city in India.

Problem Statement

- 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 analyse 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.

Description of the dataset - 1

-> **Zomato Restaurant names and Metadata**

1. **Name** : Name of Restaurants
2. **Links** : URL Links of Restaurants
3. **Cost** : Per person estimated Cost of dining
4. **Collection** : Tagging of Restaurants w.r.t. Zomato categories
5. **Cuisines** : Cuisines served by Restaurants
6. **Timings** : Restaurant Timings

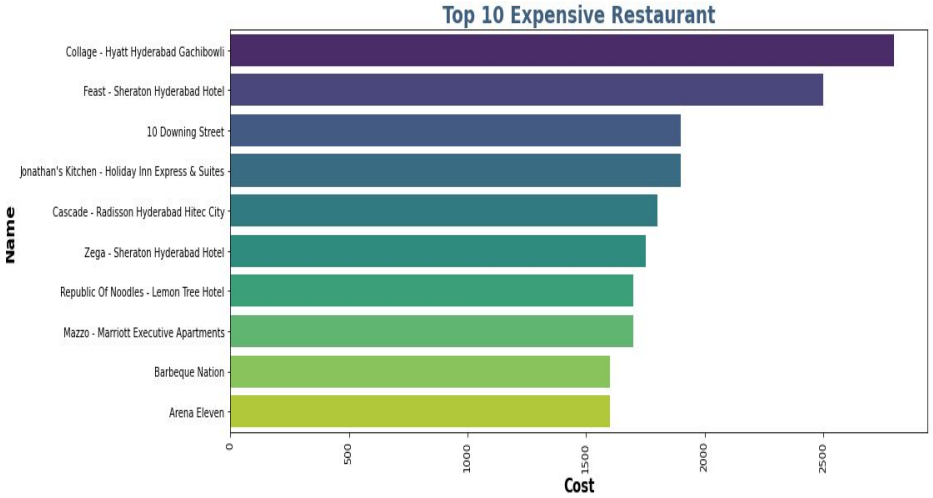
Description of the dataset - 2

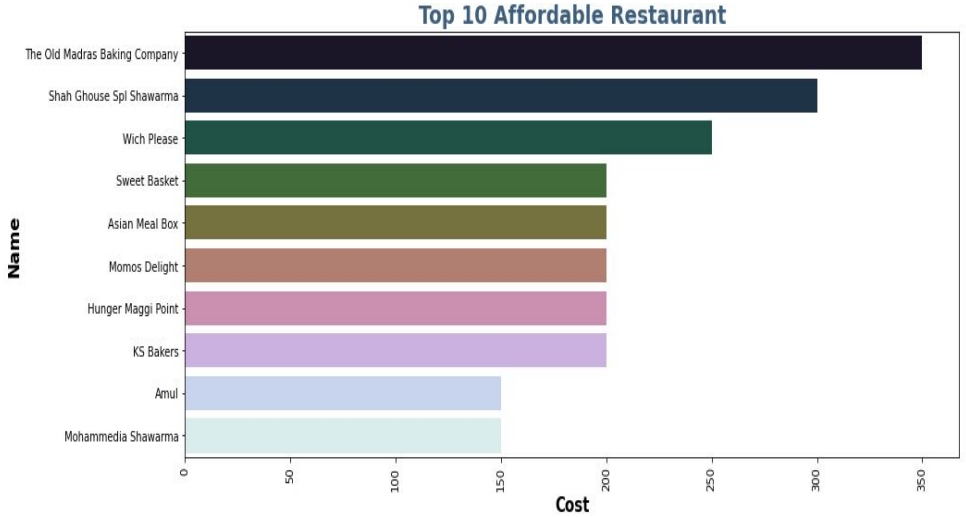
-> **Zomato Restaurant reviews**

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

EDA (Meta Data)





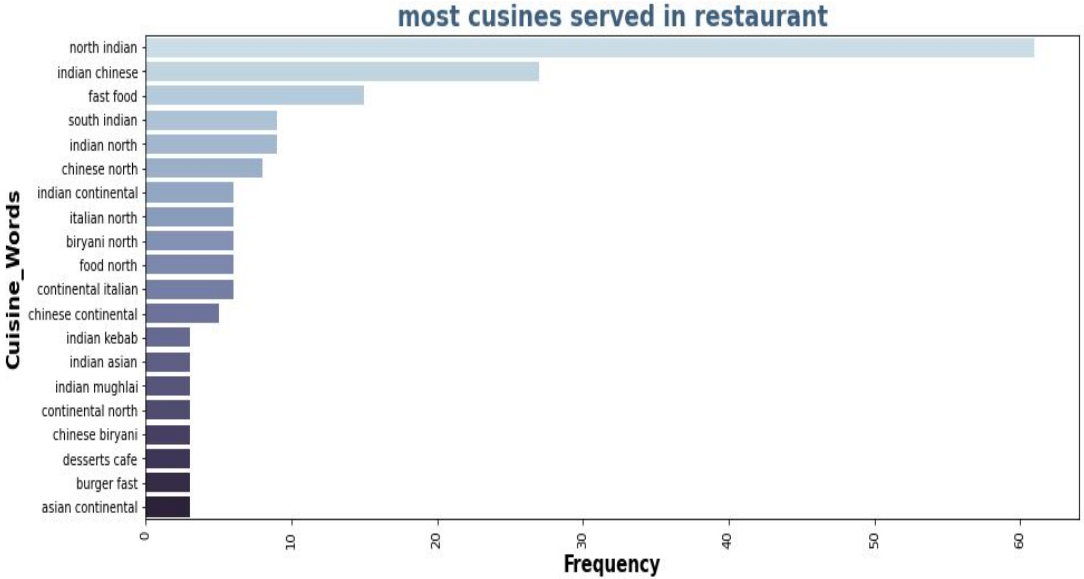
- Finding out the most expensive and most affordable restaurants can help a lot according to different pocket sizes in india.

- **Word Clouds :** A word cloud is a collection, or cluster, of words depicted in different sizes. The bigger and bolder the word appears, the more often it's mentioned within a given text and the more important it is.
  - Word cloud for expensive restaurants

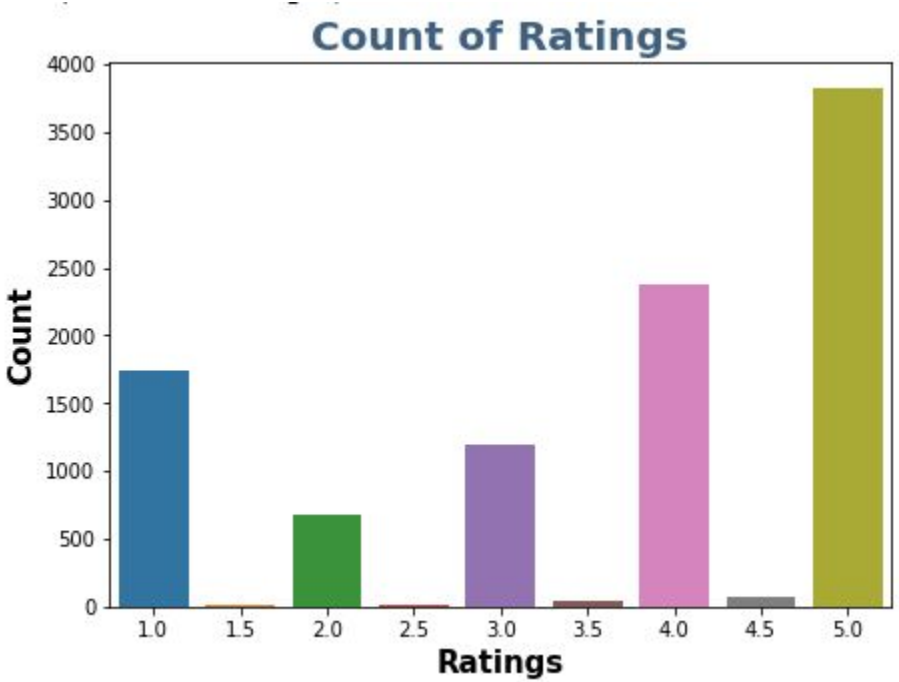
Word cloud for affordable restaurant

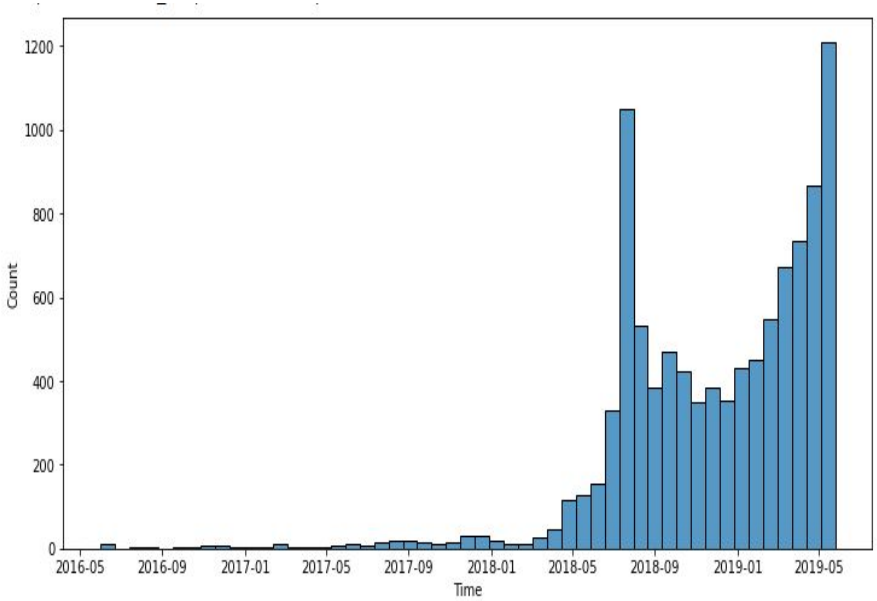
Most Cuisine Served

- North-Indian being the most served while chinese lag behind. It can be due to the data belonging to the northern regions of india.



EDA (Reviews)

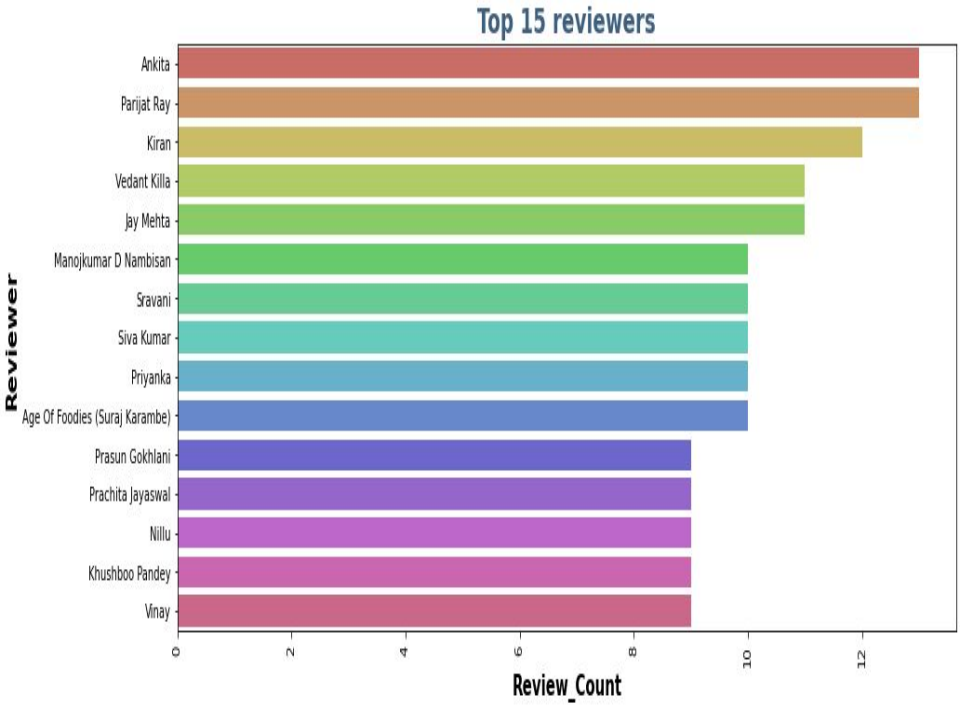


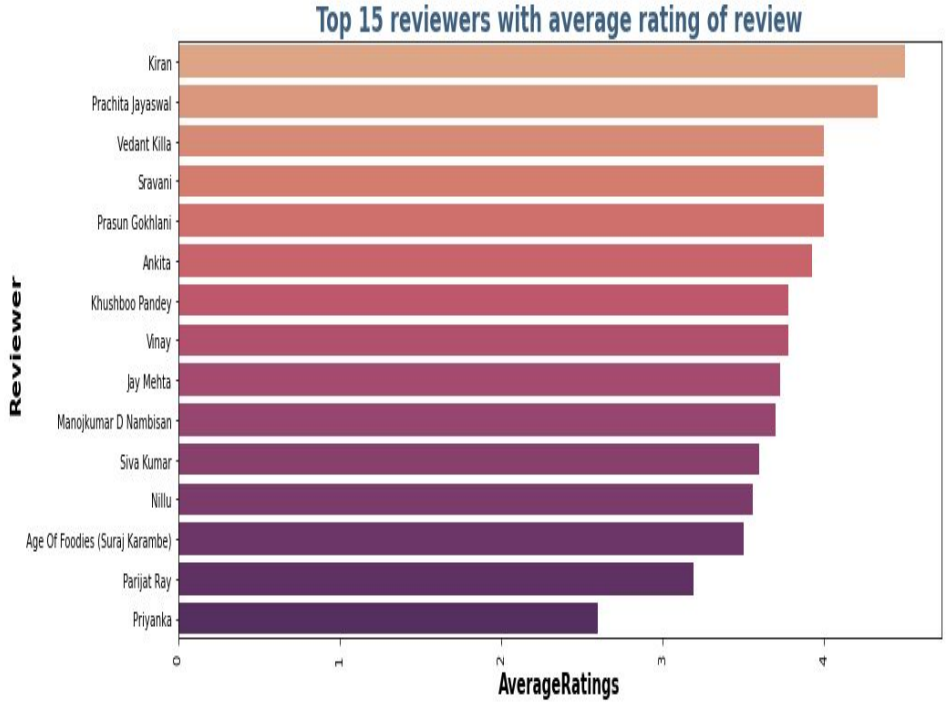


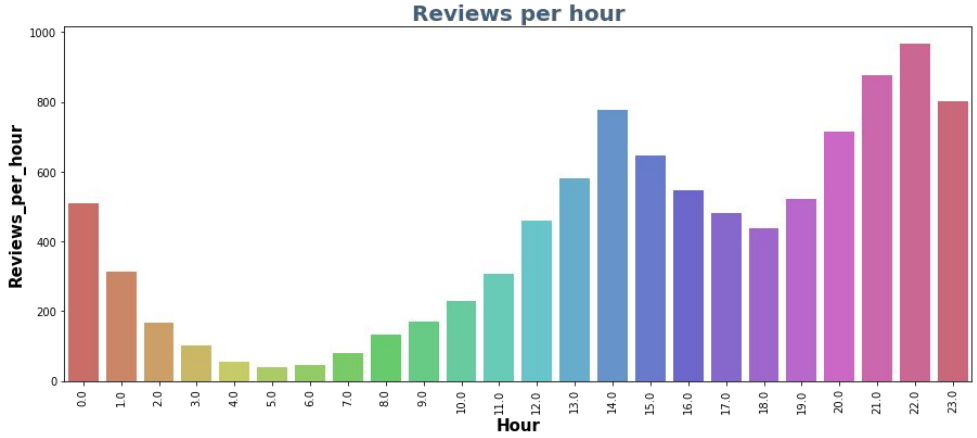
- Count of ratings and the time at which the ratings were received are displayed above.

TOP reviewers & the TOP average rating by the reviewers

- Ankita & Parijat Ray tops the list with 12+ reviews
- Kiran is the most satisfied customer it seems as she has nearly 5 star rating average







- We can spot a pattern here as most people order's start from the afternoon and slows down midnight, and hence most reviews are received in that time period.

- On the right, we have the word cloud for reviews and gives a overall positive sentiment.

Sentiment Analysis

- After completing the necessary text processing part, which contained removing punctuation, Removing stopwords & Lemmatization, we move towards Sentiment Analysis.

- The subjectivity column that showcases the sentiment is visualized above, where lite purple being ***Positive***, red being ***Negative*** and green being ***Neutral***.

LDA top 15 words in each topic

THE TOP 15 WORDS FOR TOPIC #0
['particular', 'cleanliness', 'strawberry', 'speciality', 'unfortunately', 'extraordinary', 'exceptional', 'cheesecake', 'personally', 'improvement', 'presentation', 'comfortable', 'completely', 'expectation', 'restaurant']

THE TOP 15 WORDS FOR TOPIC #1
['unhygienic', 'satisfactory', 'traditional', 'flavourful', 'understand', 'ingredient', 'complimentary', 'environment', 'combination', 'interesting', 'management', 'absolutely', 'vegetarian', 'especially', 'experience']

THE TOP 15 WORDS FOR TOPIC #2
['undoubtedly', 'background', 'surprisingly', 'satisfaction', 'instruction', 'disgusting', 'unprofessional', 'collection', 'reservation', 'impressive', 'atmosphere', 'hyderabadi', 'reasonable', 'disappoint', 'definitely']

THE TOP 15 WORDS FOR TOPIC #3
['decoration', 'undercooke', 'accommodate', 'thoroughly', 'mayonnaise', 'outstanding', 'celebration', 'perfection', 'sufficient', 'affordable', 'professional', 'appreciate', 'suggestion', 'disappointment', 'disappointed']

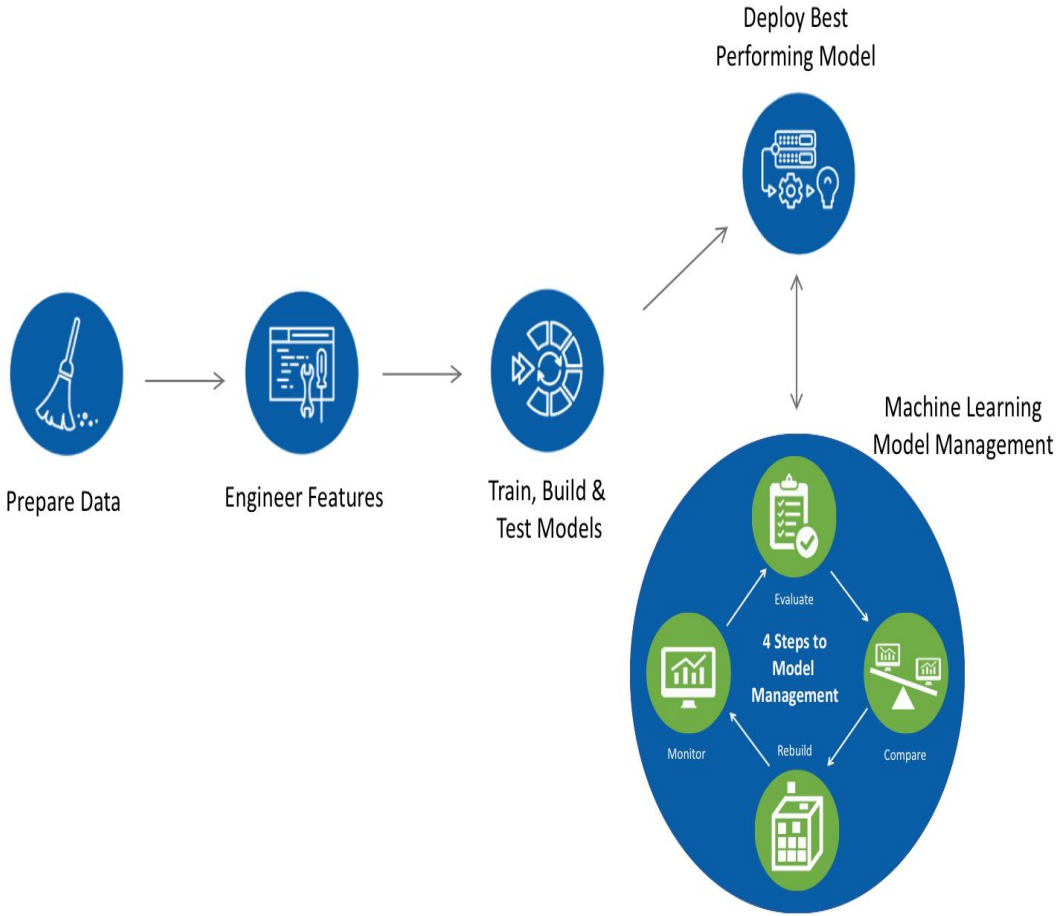
THE TOP 15 WORDS FOR TOPIC #4
['compensate', 'membership', 'specifically', 'delectable', 'conversation', 'consistency', 'ulavacharu', 'preparation', 'recommendation', 'continental', 'arrangement', 'manchurian', 'disappointing', 'hospitality', 'gachibowli']

Top 3 Cuisines in 5 clusters K-Means

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Northindian	Northindian	Northindian	Asian	Northindian
Chinese	Continental	Chinese	Italian	Chinese
Fast food	Asian	Biryani	Continental	Italian

Model’s performed

- Multinomial Naive Bayes
- Random Forest Classifier
- XGB Classifier
- Support Vector Classifier



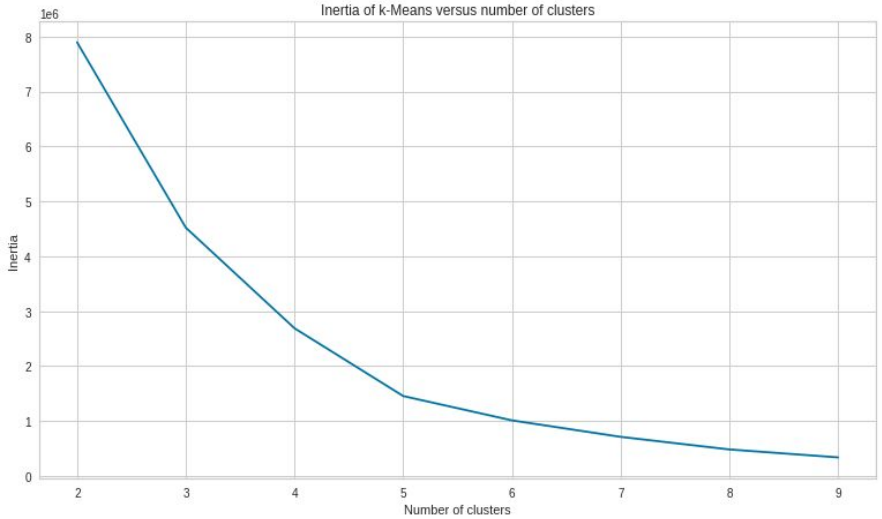
Model Validation

	Model_Name	Training_accuracy	Test_accuracy
0	MultinomialNB	0.8397	0.8264
1	Random Forest	0.8176	0.8123
2	XGB	0.9880	0.9280
3	Support Vector Machine	0.9900	0.9212

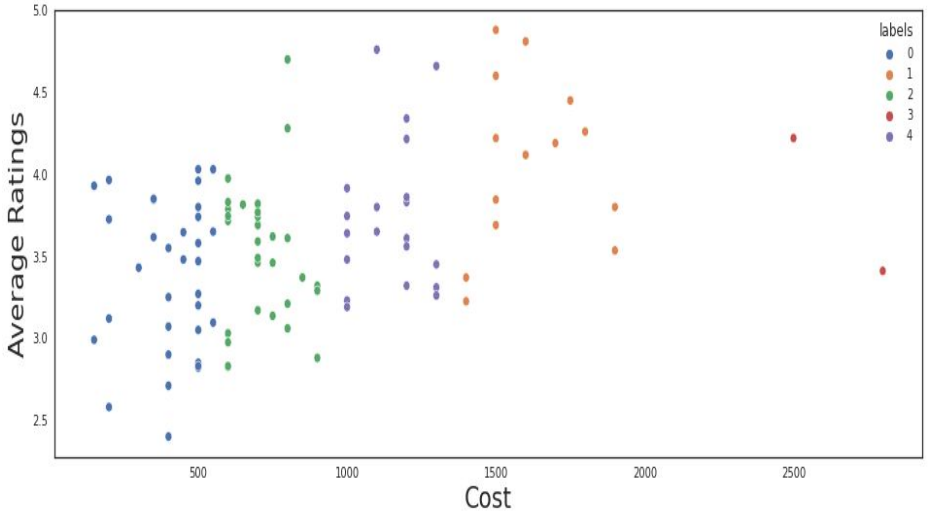
- As it is clear form the validation table that both XGB and SVM (Classifier) are working exceptionally well then compared to the other model.
- Thus; we can choose between any one of them for the production.

Clustering (KMeans)

- According to the elbow curve we should have 5 clusters for the best results

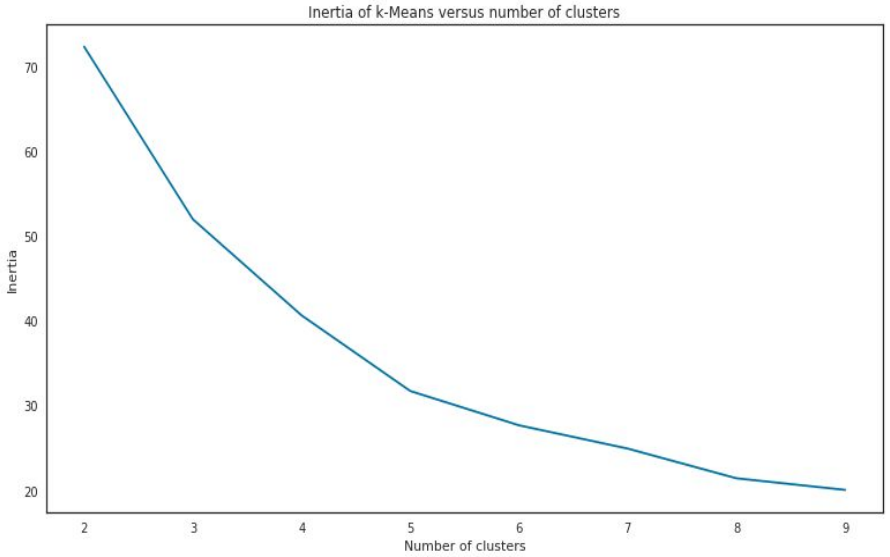


- 5 clusters on the average rating and the cost

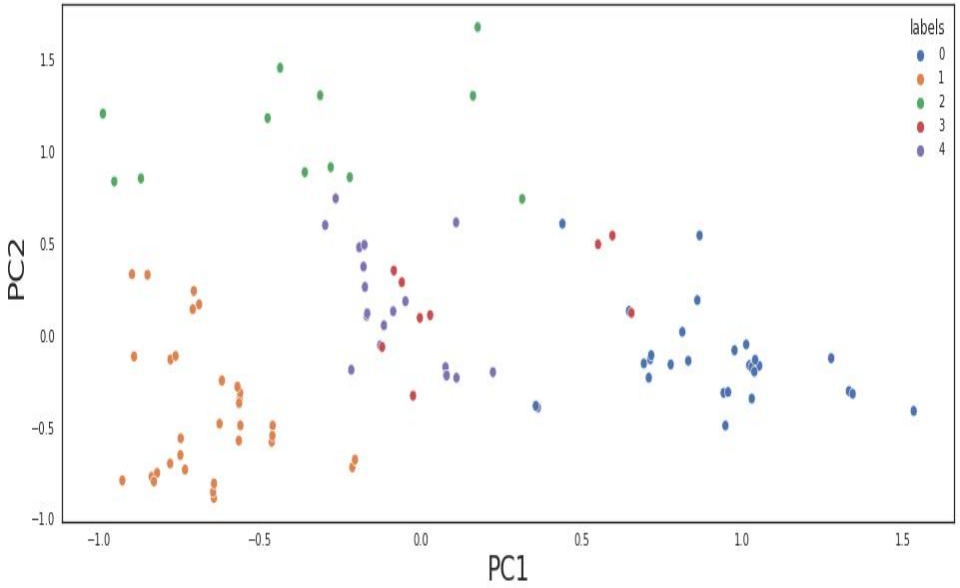


Clustering (PCA - Principal Component Analysis)

- According to the elbow curve we should have 5 clusters for the best results using PCA.



- 5 clusters on the average rating and the cost using PCA.



Conclusion

- We got best cluster as 5 in K-Means and Principal Component Analysis(PCA).
- We plot polarity and subjectivity plot for sentiment analysis, Polarity tells how positive or negative the text is. The subjectivity tells how subjective or opinionated the text is
- From the above mentioned plot, positive feedbacks are more.
- For sentiment analysis we used supervised techniques.
- We got the best model as SVM (Support Vector Machine) classifier & XGBoost


Thank You!