

Sentiment Analysis using NLP on Hotel Reviews

Presented by : Group 2

- 1.Mr.Madhuresh Kumar
- 2.Ms.Varsha Reddy
- 3.Mr.Hemant Shinde
- 4.Mr.Karan Patil
- 5.Ms.Madhuri Varpe
- 6.Mr.Omprasad Kolhal
- 7.Ms.Shraddha Hukkeri

Business Objective:

To understand the
attributes that travelers
consider while selecting
a hotel.



Thereby understanding
the elements to
be added to improve
the hotel's brand value

Exploratory Data Analysis [EDA]



BASIC DATA
UNDERSTANDING



TEXT PROCESSING



DATA VISUALIZATIONS

Basic Data Understanding



Data contains an overall of 20491 reviews with respective ratings.



Data doesn't contain any Null Entries.

Text Processing

The maximum count of review by the rating is , 5 star rating with 9054 reviews.

The least count of review by the rating is , 1 star with 1421 reviews.

The higher the rating , longer the review.

After generalizing the reviews :

- 73.7% of reviews were positive
- 15.7% of reviews were negative
- 10.7 % were neutral.

Data Visualization



Common words & Stop words



Bi and Tri Grams

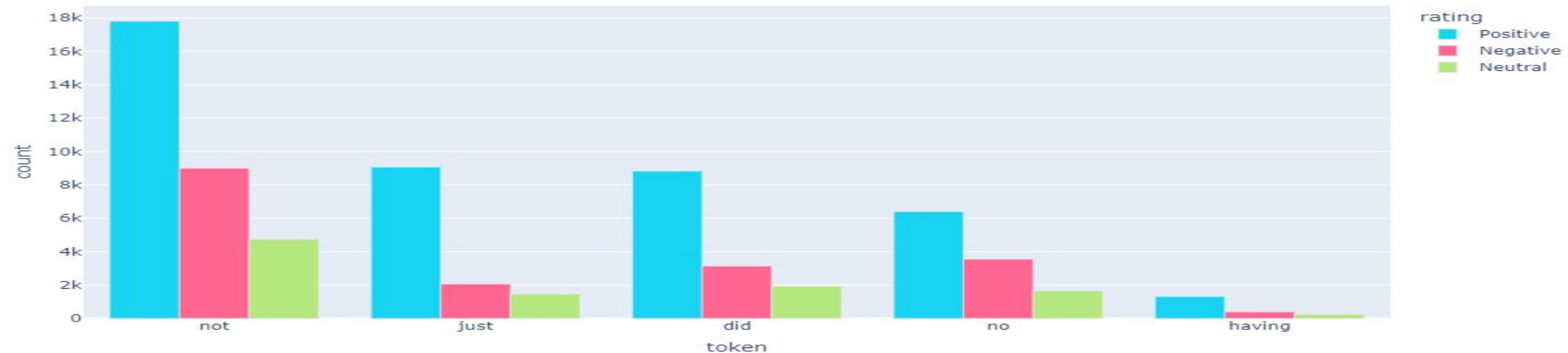


Word Clouds

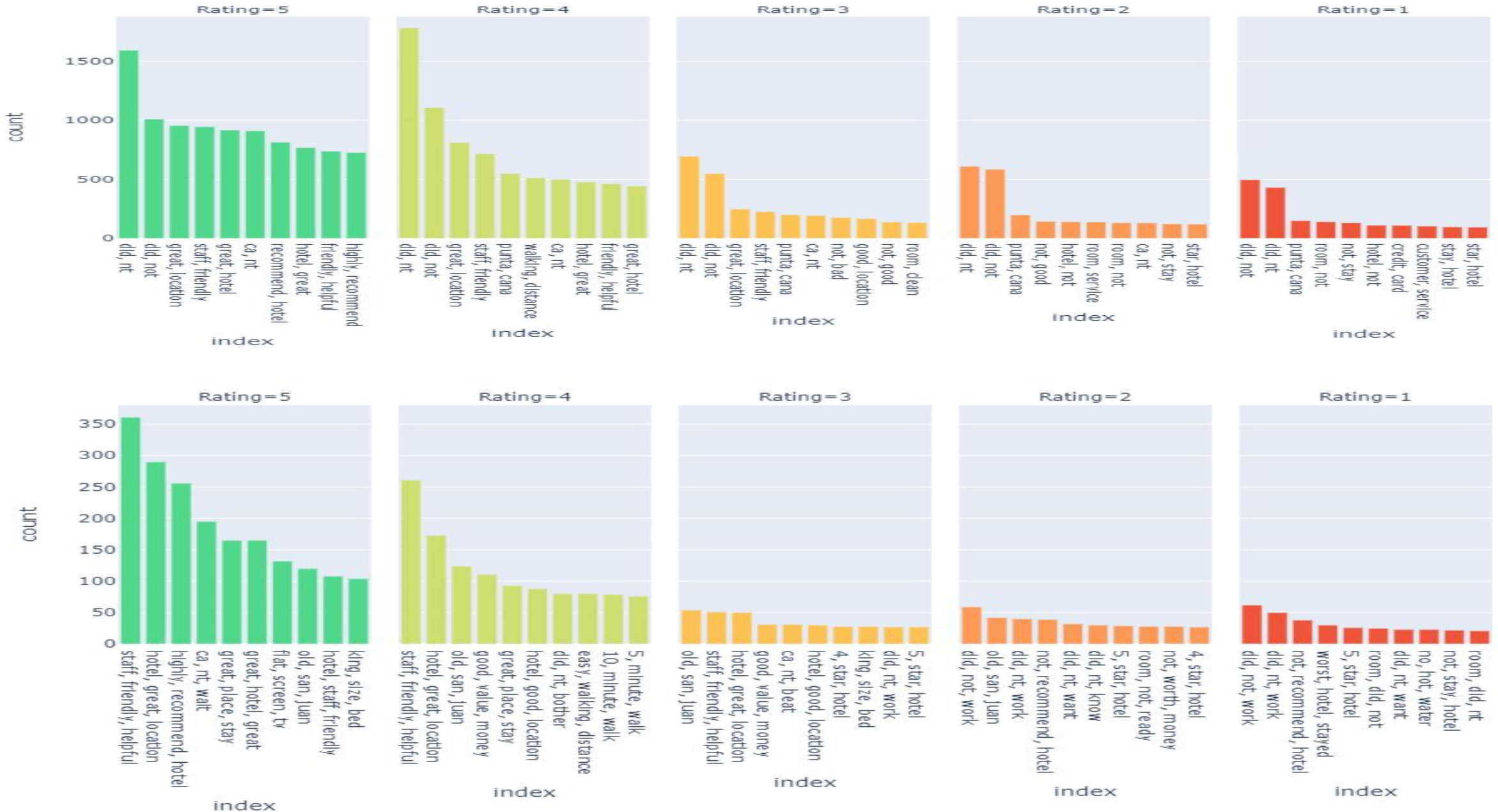
Common words & Stop words



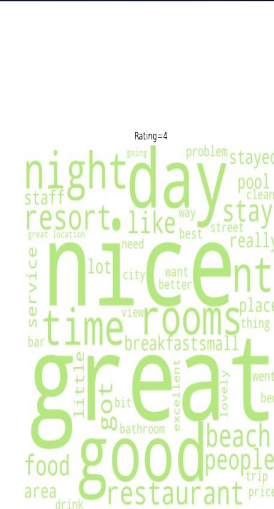
Top common stopwords



Bi and Tri Grams

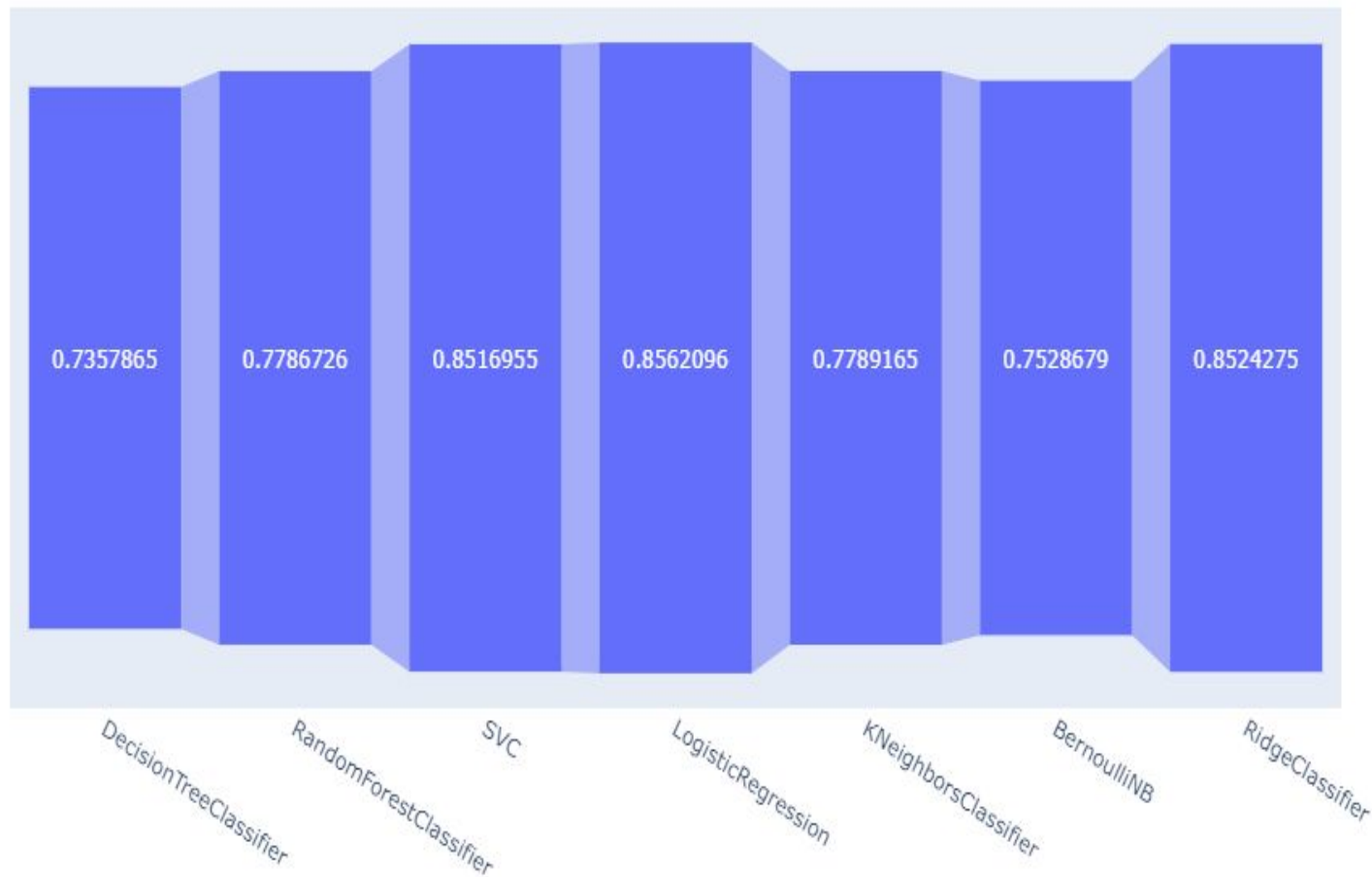


Word Clouds



Model Building with SKLEARN

Data was split into 80 : 20 as for training and testing purposes .



Model	Accuracy
Decision Tree Classifier	0.7401181 ~ 74.01%
Random Forest Classifier	0.7785512 ~ 77.85%
SVC	0.8547473 ~ 85.47%
Logistic Regression	0.8585904 ~ 85.85%
KNeighbors Classifier	0.7775745 ~ 77.75%
Bernoulli N B	0.7512816 ~ 75.12%
Ridge Classifier	0.8559050 ~ 85.59%

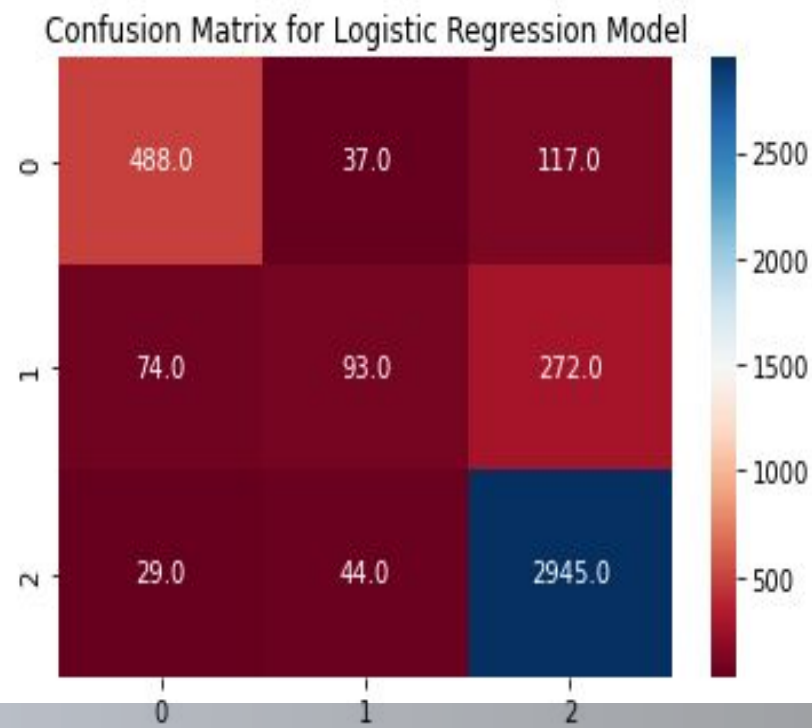
Model Finalization

Since Logistic Regression performs better than other models in this case, we will use Logistic Regression to train our model

Classification report and Confusion matrix

	precision	recall	f1-score	support
Negative	0.83	0.76	0.79	642
Neutral	0.53	0.21	0.30	439
Positive	0.88	0.98	0.93	3018
accuracy			0.86	4099
macro avg	0.75	0.65	0.67	4099
weighted avg	0.84	0.86	0.84	4099

Text(0.5, 1.0, 'Confusion Matrix for Logistic Regression Model')



SELECT THE OPTION:-

- ☐ About
- ☐ Dataset
- ☒ Sentiment of your Sentence
- ☐ Dataset Reviews

Enter your Sentence/Review :

Hotel room was neat and clean , service was actually better.

Check spellings and Analyse

After Spelling Correction:-

Hotel room was neat and clean , service was actually better.

Sentiment of your review:-

Positive

Subjectivity:- 0.6

Deployment using Streamlit

SELECT THE OPTION:-

- ☐ About
- ☐ Dataset
- ☐ Sentiment of your Sentence
- ☒ Dataset Reviews

Enter the row number of review:-

Choose between 0 to 20490

125

- +

Analyse

wonderful hotel, just spent nights mayflower park wonderful experience, neat old hotel lot charm, staff friendly concierge folks helpful, hardly noise rooms, good beds not rock-hard like hotels, oliver bar bit pricey convenient better nonsmoking like rest hotel, upper floor room great view space needle, location ca n't beat, walked, concierge helped rent car day explored olympic peninsula, husband preferred staying cheaper n't deny nice experience mayflower park,

Sentiment of your review:-

Positive

Polarity:- 0.43

Subjectivity:- 0.58



THANK
YOU