**HACKATHON**

***3.2 Sentiment Analysis***

**Problem statement**

* Building a sentiment analysis model to analyse customer reviews and feedback for a hospitality company.
* The model should be able to classify reviews as positive, negative or neutral.

**Dataset Information:**

* The dataset contains columns such as Review and Rating
* 'Rating' is the sentiment label (i.e target label)

**Key Observations:**

* The dataset has a substantial number of reviews, providing a rich source of data for sentiment analysis.
* 'Rating' distribution shows an imbalance

**Data Cleaning:**

* Checked for missing values, duplicates, and irrelevant information.
* Ensured data quality by addressing anomalies and inconsistencies.

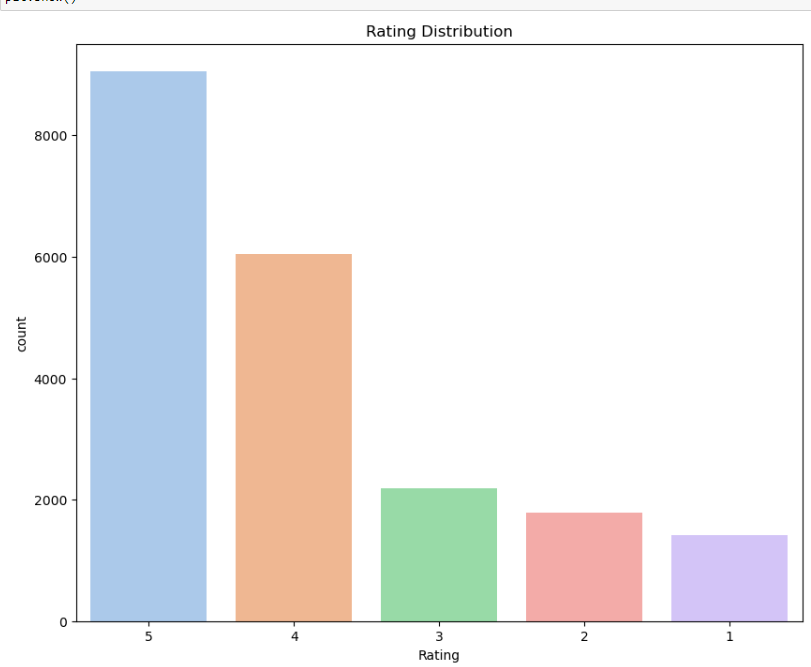
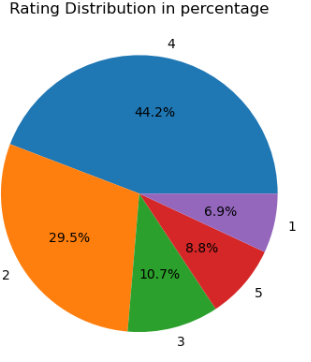
**Data Preprocessing:**

* Text preprocessing involved removing stop words, special characters and etc.
* Tokenization and lemmatization were applied to prepare the text for sentiment analysis.

**Exploratory Data Analysis (EDA):**

**Sentiment Distribution:**

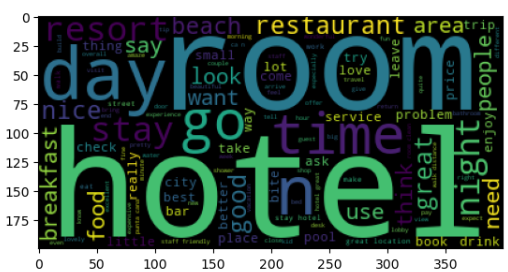
* Visualized the distribution of sentiment labels using a count plot.

* Observed an imbalance between target labels.

**Word Frequency Analysis:**

* Analysed the frequency of words in reviews to identify common terms and themes.
* Created word clouds for positive, negative and neutral sentiments to visualize most frequent words.



**Model Selection:**

• Deployed Bernoulli’s Naïve Bayes and Multinomial Naïve Bayes models using Count Vectorizer for Machine Learning models to predict the sentiment of the reviews.

• Deployed a Deep Learning model using LSTM for sentiment analysis prediction

**Model Evaluation:**

* Evaluated model performance on both training and test sets.
* Calculated accuracy and generated a classification report.

**Results:**

* Achieved an accuracy of 72% using Deep learning model (i.e LSTM)
* Achieved an accuracy of ------- using Machine learning models (i.e Bernoulli’s Naïve Bayes and Multinomial Naïve Bayes)

**Conclusion:**

The sentiment analysis model successfully classifies customer reviews into positive, negative, or neutral categories. This can provide valuable insights into customer satisfaction and areas for improvement.

**Future Improvements:**

• The accuracy can be increased by performing **hyperparameter** tuning like GridSearchCV and balancing the data more.