

Text Analysis

Report

COURSE PRESENTER: (Dr. Omaima Fallatah)

|  |  |
| --- | --- |
| Name | Id |
| Yara abodayeh | 444015039 |
| Shahad hasan | 444006091 |

Report on Review Analysis Using Python

# Introduction

This report summarizes the findings derived from the code used to analyze product reviews. Various Python libraries were employed for data processing and sentiment analysis, providing valuable insights into customer preferences.

### Objective

* **Understanding Customer Behavior**
* **Improving Store Layout**
* **Developing Effective Marketing Strategies**
* **Increasing Sales**

### Data Overview

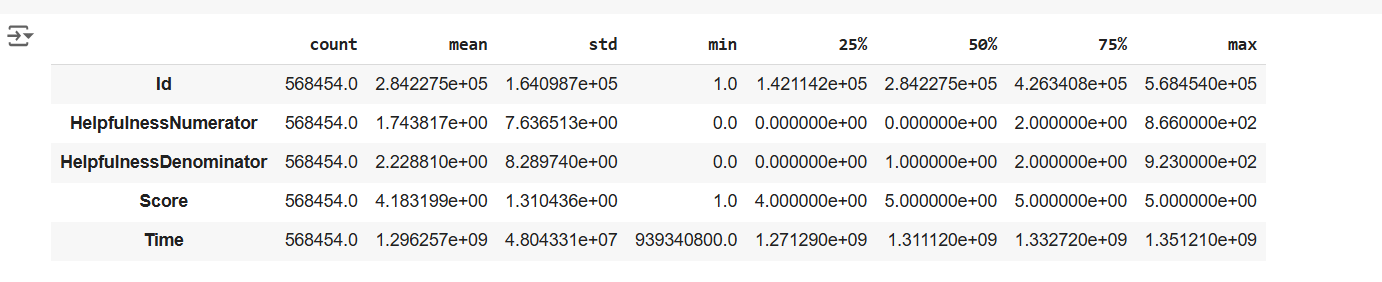
This report analyzes product reviews from users to evaluate product quality and the helpfulness of the reviews. The dataset includes the following columns:

1. **Id**: Unique identifier for each review.
2. **ProductId**: Identifier for the product being reviewed.
3. **UserId**: Identifier for the user who wrote the review.
4. **ProfileName**: Name or profile of the user who submitted the review.
5. **HelpfulnessNumerator**: Number of users who found the review helpful.
6. **HelpfulnessDenominator**: Total number of users who rated the review.
7. **Score**: Rating given by the user (typically from 1 to 5).
8. **Time**: Timestamp when the review was published (in Unix time).
9. **Summary**: Short summary or title of the review.
10. **Text**: Full text of the review.

# 1. Data Statistics

-The dataset was successfully loaded, containing information about product reviews and their ratings.

-The describe() function was used to obtain descriptive statistics, helping to understand the data distribution.



# 3. Word Cloud

A word cloud was generated for the 5-star reviews, allowing for visualization of the most frequently used words in positive reviews.

A close-up of words

Description automatically generated

# 4. Text Cleaning

-The cleaning() function was applied to remove links, special characters, numbers, and punctuation from the text.

-This process improved the quality of the data used for analysis.

# 5. Analysis of Most Common Words

-The Counter library was utilized to count the frequency of words in the reviews after removing stopwords.

-The top 10 most common words in the reviews were identified, providing insights into recurring themes.

A screenshot of a computer

Description automatically generated

# 6.Word Frequency Analysis in Reviews

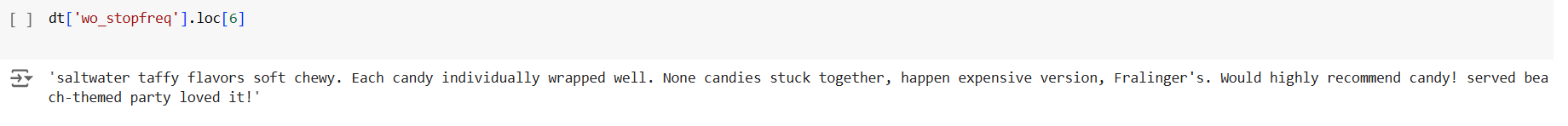
# Objective: Identify the most common words in reviews after removing stop words to understand user descriptions of products.

Count words in the no\_sw column using Counter.

Extract and display the top 10 most frequent words.

# 

# Remove frequent words using the remove\_freqwords function.



# In the project, we used the process of tokenization to split texts into individual words. This process is a fundamental step in text processing.



# 7. Model Evaluation

Several classification models were used, with the following results:

**7.1 Complement Naive Bayes (CNB) Model**

A screenshot of a computer code

Description automatically generated

**-Model Accuracy**: 87.49%

-Demonstrated strong performance in classifying reviews as positive or negative.

**7.2 Multinomial Naive Bayes (MNB) Model**

**A screen shot of a computer code

Description automatically generated**

**-Model Accuracy**: 90.89%

-Provided reliable results in sentiment analysis.

# 8. ROC Curve Analysis

The models were evaluated using the ROC curve, showing good performance for both the Complement Naive Bayes and Multinomial Naive Bayes models, with closely matched results.

A graph of a curve

Description automatically generated

# 9. N-grams Analysis

-Count Vectorizer was used with n-grams (bigrams) to enhance model performance.

-The Multinomial Naive Bayes model achieved an accuracy of 90.89% when using bigrams.

# 10. TF-IDF Analysis

-TF-IDF was applied to convert reviews into numerical representations.

-The MNB model achieved an accuracy of 90.89%

# Conclusion

The results indicate that employing natural language processing techniques and machine learning models can significantly contribute to the sentiment analysis of reviews. This aids companies in better understanding customer preferences, ultimately enhancing products and services.