

1. Polarity Distribution

- **Purpose:** To understand the distribution of the target variable (Polarity). Are the reviews mostly positive or negative?
- **Reason:** Knowing the class distribution is important for understanding dataset balance. For instance, a heavily imbalanced dataset might require special handling in model training (e.g., using techniques like SMOTE to handle class imbalance).

2. Title Length Analysis

- **Purpose:** To analyze the length of review titles across different sentiments.
- **Reason:** Titles often summarize the key point of a review. Shorter, concise titles may indicate strong opinions (either positive or negative), whereas longer titles might reflect indecision or detailed feedback. This analysis can provide insight into customer communication patterns.

3. Text Length Analysis

- **Purpose:** To examine the distribution of review text lengths.
- **Reason:** Text length might correlate with the sentiment. For example, negative reviews could be longer if the user is describing problems in detail, while positive reviews might be shorter if the user is simply expressing satisfaction.

4. Polarity vs Text Length

- **Purpose:** A boxplot that shows the relationship between review text length and polarity (positive/negative).
- **Reason:** This helps check whether there's a significant difference in the length of positive and negative reviews. Often, users who are dissatisfied might write longer, more detailed reviews. This can indicate how much effort users put into describing their experience based on their sentiment.

5. Word Clouds for Positive and Negative Reviews

- **Purpose:** To visualize the most frequently used words in both positive and negative reviews.
- **Reason:** Word clouds highlight the most prominent words in the dataset. By separating positive and negative reviews, you can see which terms are common in each sentiment. This gives a quick overview of the main themes and patterns in user feedback.

6. Most Common Words in Positive and Negative Reviews

- **Purpose:** To extract the most frequent words from both positive and negative reviews.
- **Reason:** This is more detailed than a word cloud. By identifying the top words used in positive and negative reviews, you can understand common themes or issues. For example, positive reviews might frequently mention words like "quality" or "great," while negative reviews could have words like "poor" or "disappointed."

7. Bigrams and Trigrams Analysis

- **Purpose:** To identify the most common bigrams (two-word combinations) and trigrams (three-word combinations) in positive and negative reviews.
- **Reason:** Single words provide insight, but phrases often capture more meaningful patterns. For example, "poor quality" or "excellent product" are more informative than individual words like "poor" or "excellent." This analysis can highlight key issues or praises in reviews.

8. Sentiment by Title Keywords

- **Purpose:** To find the most common words in review titles for both positive and negative sentiments.
- **Reason:** The title often acts as a summary of the review, so identifying frequently used words in titles helps reveal key customer concerns or satisfaction. Positive review titles might use words like "love" or "great," while negative ones could have words like "bad" or "terrible."

9. Top Positive and Negative Reviews by Text Length

- **Purpose:** To retrieve and analyze the longest and shortest reviews for both positive and negative sentiments.
- **Reason:** This helps you explore the extremes in the dataset. Long reviews could give you detailed insights into customer experiences, while short reviews might provide quick, impactful feedback. This can be useful for deeper qualitative analysis of specific reviews.