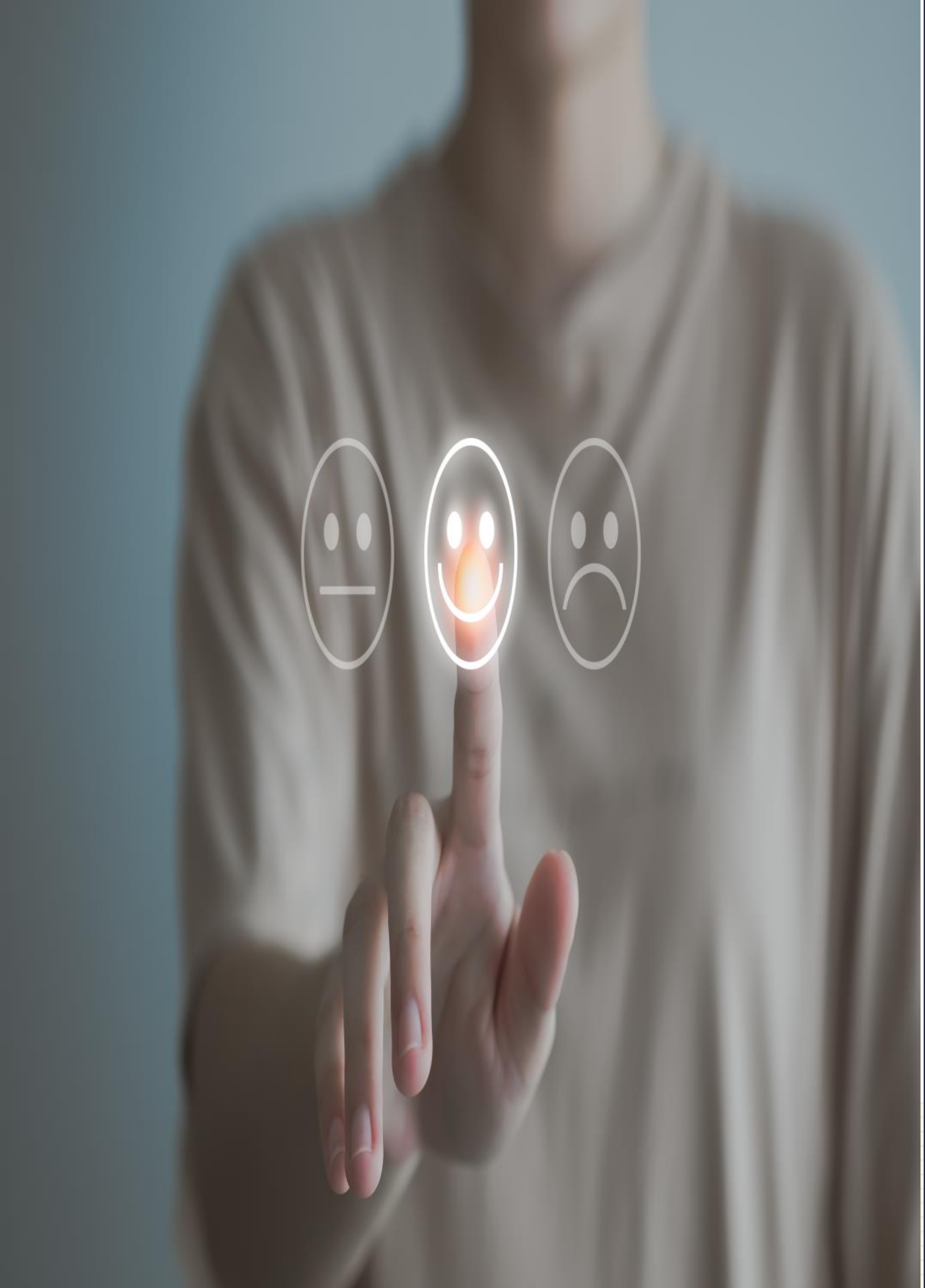
SENTIMENT ANALYSIS

CUSTOMER FEEDBACK ANALYSIS FOR IMPROVED SERVICE DELIVERY

TABLE OF CONTENT



Business Overview

- **Sentiment Analysis is a** technique to determine emotions (positive, negative, neutral) in text.
- In today's competitive market, analyzing customer feedback is vital but challenging due to its unstructured nature. Manual methods are error-prone, and existing tools lack depth. This project offers an automated system for sentiment analysis, trend tracking, and thematic discovery, providing real-time actionable insights to improve business strategies and service delivery.

PROBLEM STATEMENT

- Businesses face challenges in analyzing vast customer data, hindering their ability to deliver personalized experiences and understand sentiment, leading to missed opportunities in engagement and loyalty. The proposed system addresses this by:
- Enhancing Satisfaction: Tackling negative sentiments.
- Improving Decisions: Revealing key trends for strategy
- .Boosting Advantage: Strengthening loyalty and reducing churn.

Data Understanding

- The dataset combines internal and public repositories to analyze customer feedback.
- **Internal Sources**: Company databases with post-purchase reviews, customer service feedback, and survey responses. **Public Sources**: Datasets from Yelp, Amazon reviews, and Kaggle, offering diverse customer insights.
- **Data Features**:

1. **Review Text**: Unstructured feedback for sentiment and topic analysis.
2. **Ratings**: Quantifies satisfaction, validates sentiment classification, and offers a quick satisfaction overview.
3. **Product/Service Category**: Labels for segmentation, enabling targeted analysis of sentiment and themes specific to offerings.

DATA PREPARATTION

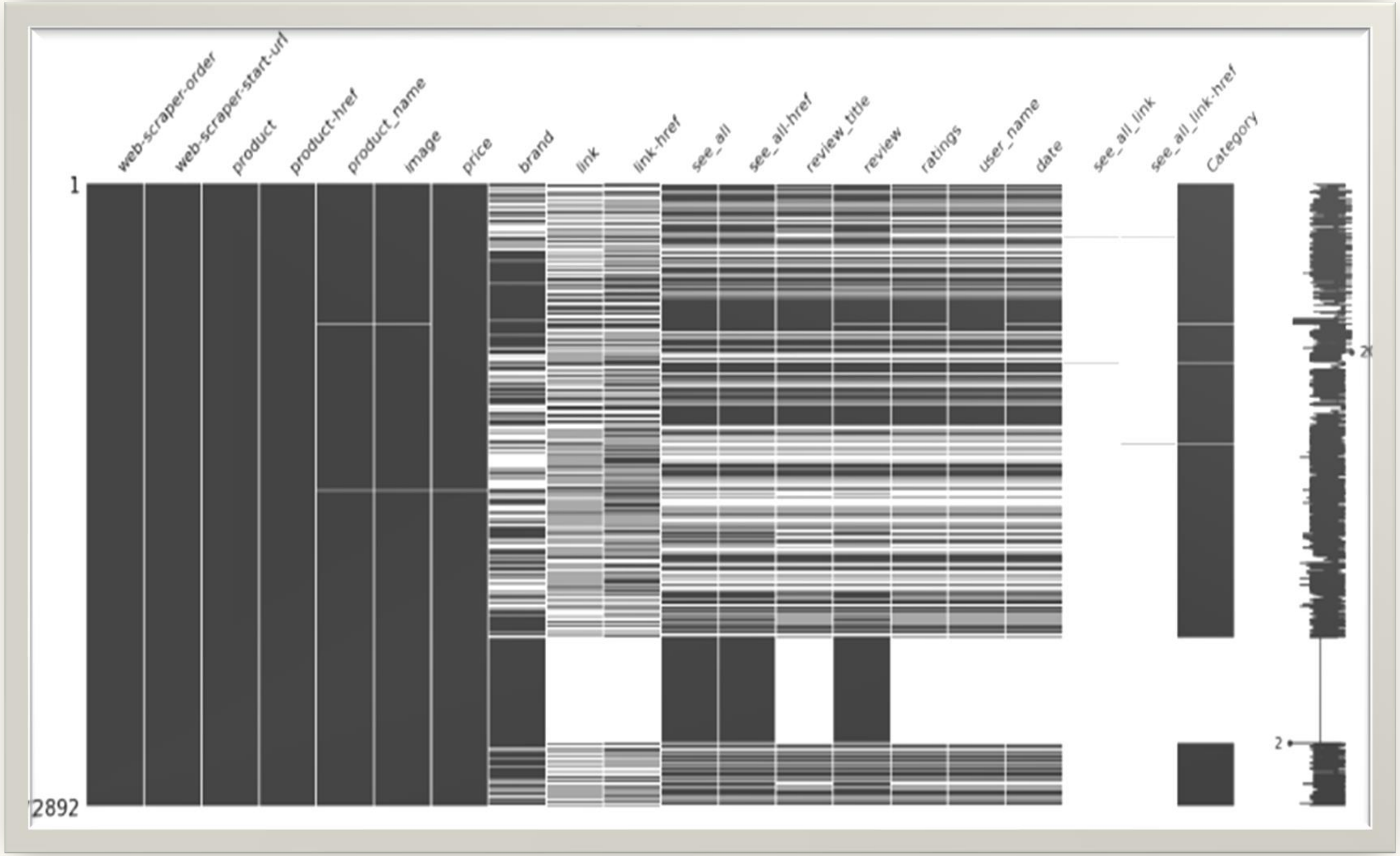
To prepare the data for modelling, these are the steps we took:

- Handling missing values: Identify and address gaps in the data by imputing missing values and removing incomplete records to ensure consistency

- Feature selection and engineering: Select the most relevant variables and created new ones to improve model performance and highlight underlying patterns.

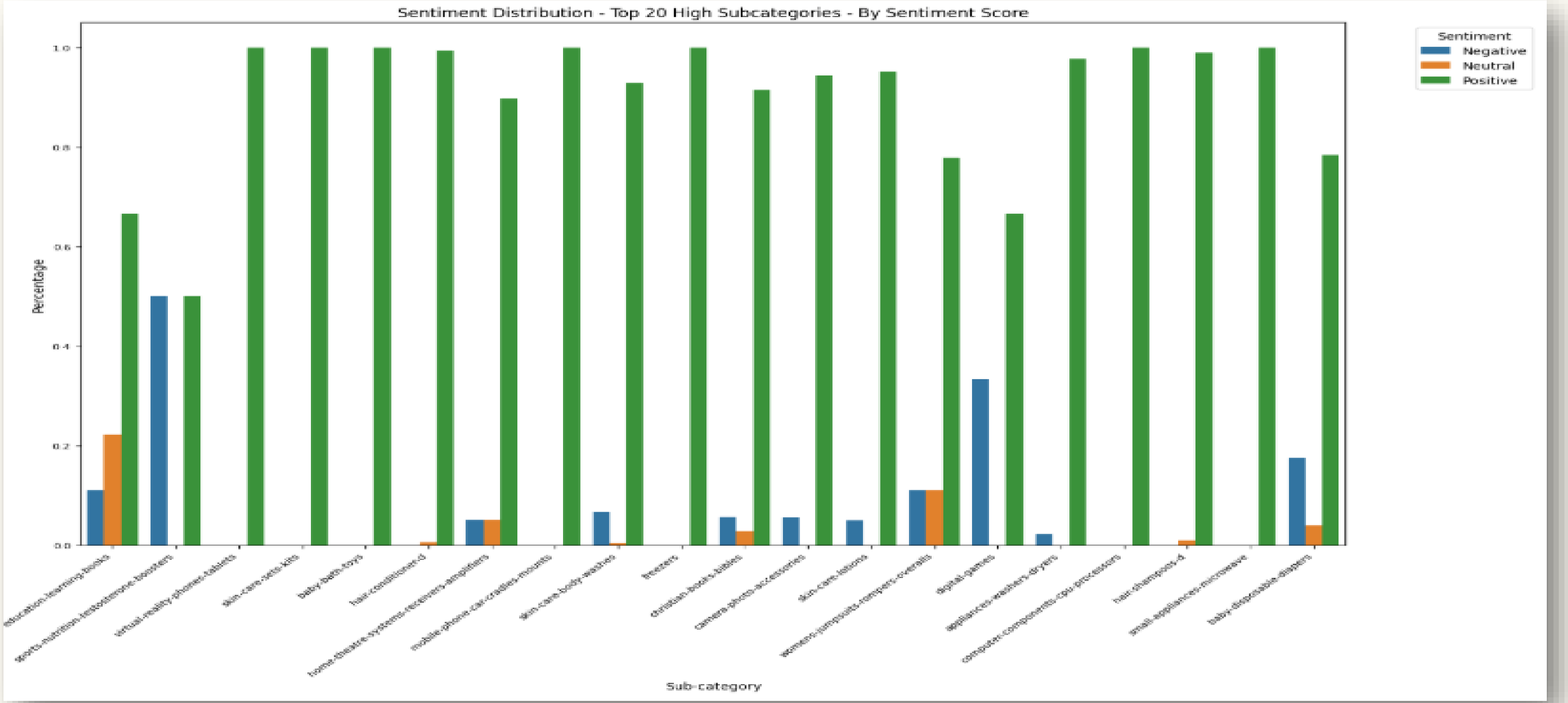
- One-hot encoding and standardization: Convert categorical variables into binary columns and scale numerical features to a uniform range for better compatibility with our machine learning algorithms

DATA CLEANING



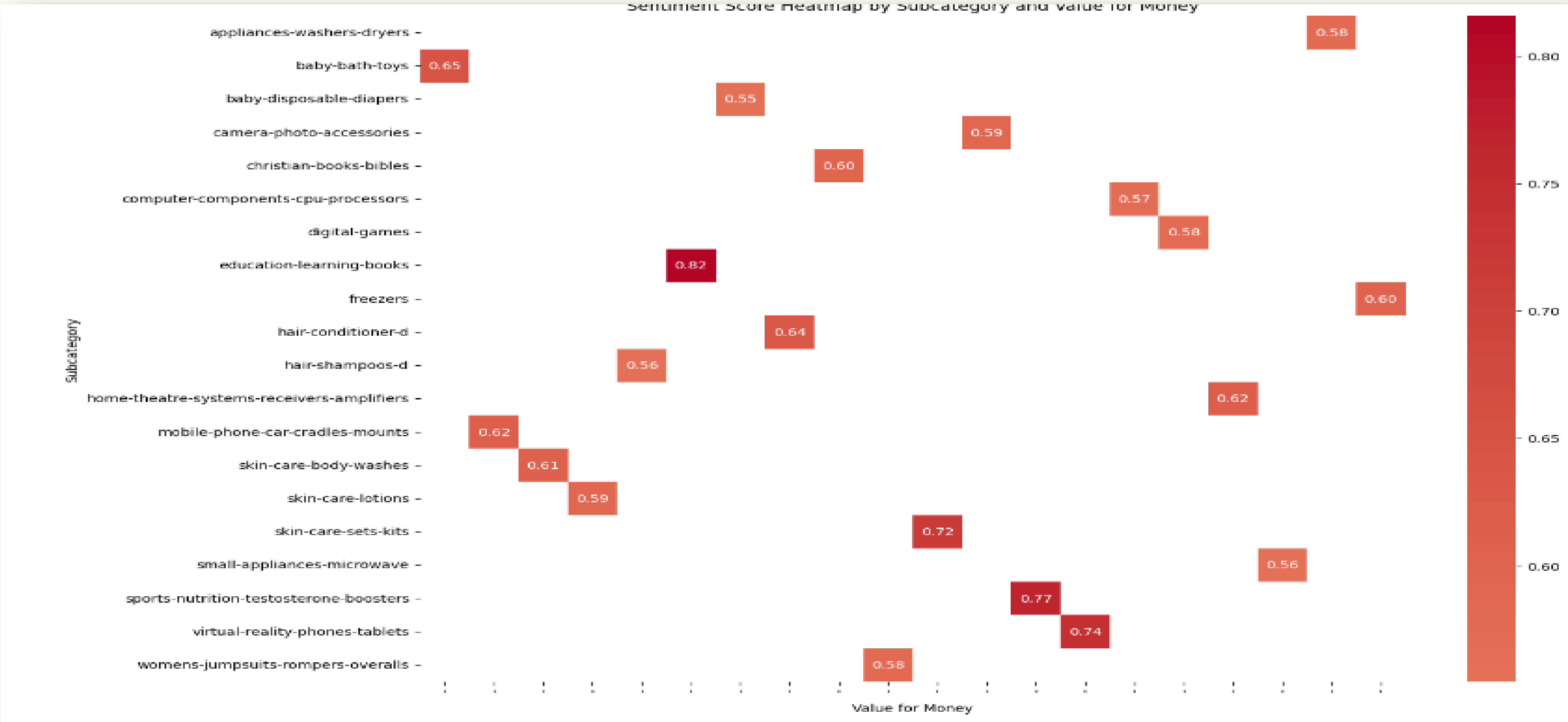
The graph identifies which columns in your dataset have missing values and how significant the missing data is in terms of percentage. It helps in cleaning and preprocessing the data.

Data Analysis: Univariate Analysis

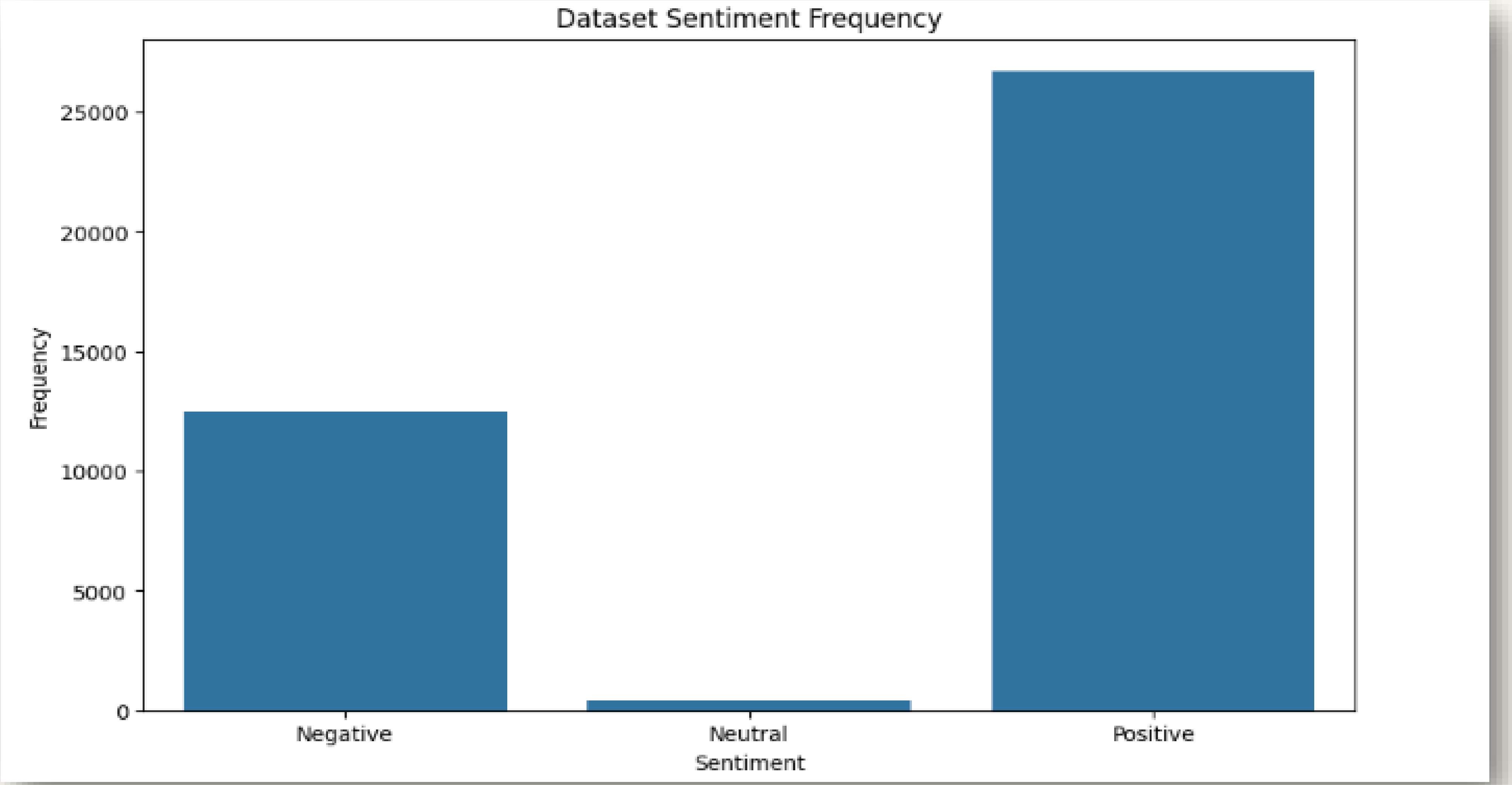


Data Analysis: Bivariate Analysis

A heat map on the value for money by sub categories



Data Analysis



Sentiment Analysis

Positive Reviews

The words shown represent the positive sentiments that were observed in out data set.

Sentiment Analysis

Negative Reviews

The words shown represent the negative sentiments that were observed in out data set.

Sentiment Analysis

Neutral Reviews

The words shown represent the neural sentiments that were observed in out data set.

Modeling

After training the dataset, SVM, TFIDF, Naive Bayes, and BOW were tested. Naive Bayes performed best, compared to SVM, TFIDF and BOW. we deployed a sentiment analysis app that classifies sentiments according to Negative positive and neutral sentiments.

Model 1 : SVM

	precision	recall	f1-score	support
Negative	0.61	0.23	0.33	2590
Neutral	0.00	0.00	0.00	79
Positive	0.70	0.93	0.80	5259
accuracy			0.69	7928
macro avg	0.44	0.39	0.38	7928
weighted avg	0.67	0.69	0.64	7928

Model 2: Naïve Bayes

	precision	recall	f1-score	support
Negative	0.61	0.24	0.34	2590
Neutral	0.00	0.00	0.00	79
Positive	0.71	0.93	0.80	5259
accuracy			0.69	7928
macro avg	0.44	0.39	0.38	7928
weighted avg	0.67	0.69	0.64	7928

Modelling

Model 3: BOW

	precision	recall	f1-score	support
Negative	0.61	0.23	0.33	2590
Neutral	0.00	0.00	0.00	79
Positive	0.70	0.93	0.80	5259
accuracy			0.69	7928
macro avg	0.44	0.39	0.38	7928
weighted avg	0.67	0.69	0.64	7928

Based on the provided metrics,

Model 2 Naïve Bayes appears to be the best overall performer, offering good precision, recall, and F1-score for the positive class.

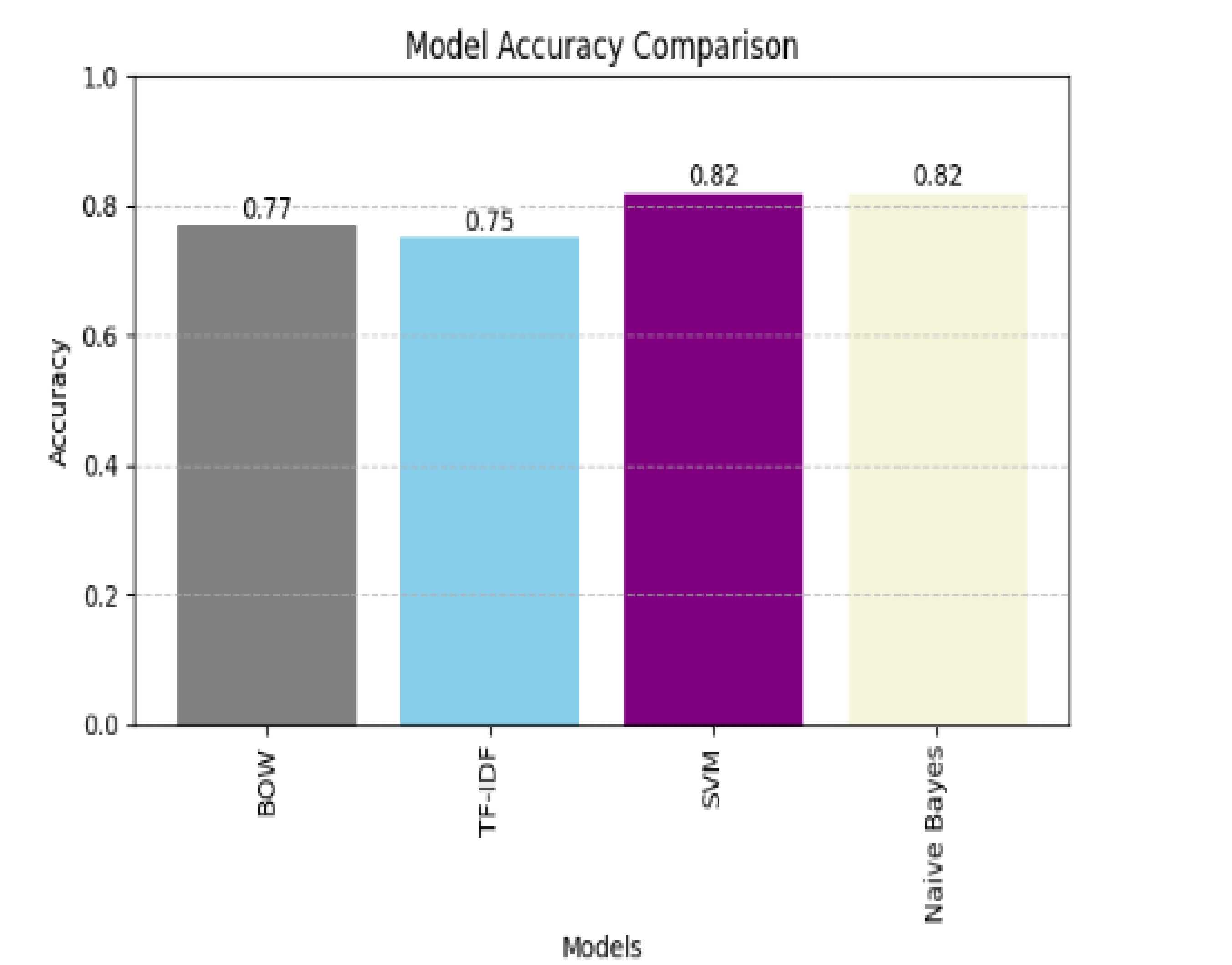
It is also likely to be computationally efficient and interpretable.

Model Evaluation

- **SVM and Naive Bayes Lead** with 82% Accuracy: Both SVM and Naive Bayes achieve the highest accuracy of 82%, indicating strong performance for this sentiment analysis task.
- **TF-IDF Outperforms BOW**: The TF-IDF model demonstrates improved accuracy (75%) compared to the BOW model (77%). This suggests that incorporating term frequency and inverse document frequency can enhance sentiment analysis.

- Consider Model Complexity and Interpretability: SVM and Naive Bayes have similar accuracy, SVM might be more complex and less interpretable.

- Naive Bayes, on the other hand, offers a simpler and more interpretable approach, which can be valuable in certain applications.



Conclusion

- Through the analysis of reviews and sentiments using Naive Bayes, I was able to classify sentiments into Positive, Neutral and Negative. The results indicate that:
- - some categories like Educational books category offers the highest value for money while some like baby disposable dippers offer the least value for money. This information can be valuable for various purposes, such as product recommendations, marketing strategies, and business decision-making.
- - Overall, The sentiment count shows that the positive sentiments have a higher count showing that customers are generally satisfied with the products.
- - By analyzing the sentiment count against price of goods we discovered that goods with medium prices get more positive sentiments compared to those with low or high prices. This shows that prices of products affect the reviews and sentiments they get.
- These insights can be valuable for: Business Owners, Customers, Product Managers, Marketing Teams, Development Teams and Data Analysts.

Recommendations

- 1. Data Quality: Invest in improving data quality and consistency to enhance model performance.
- 2. Model Selection: Consider exploring other machine learning algorithms or ensemble methods to potentially improve accuracy and robustness.
- 3. Hyperparameter Tuning: Conduct rigorous hyperparameter tuning to optimize model performance.
- 4. Feature Engineering: Experiment with different feature engineering techniques to extract more informative features from the data.
- 5. Domain Expertise: Collaborate with domain experts to gain deeper insights and refine the analysis.
- 6. Ethical Considerations: Ensure that the model is fair, unbiased, and adheres to ethical guidelines.

THANK YOU