Analyzing Product Reviews

Uncovering Topics and Sentiment

By: Wesley Yu

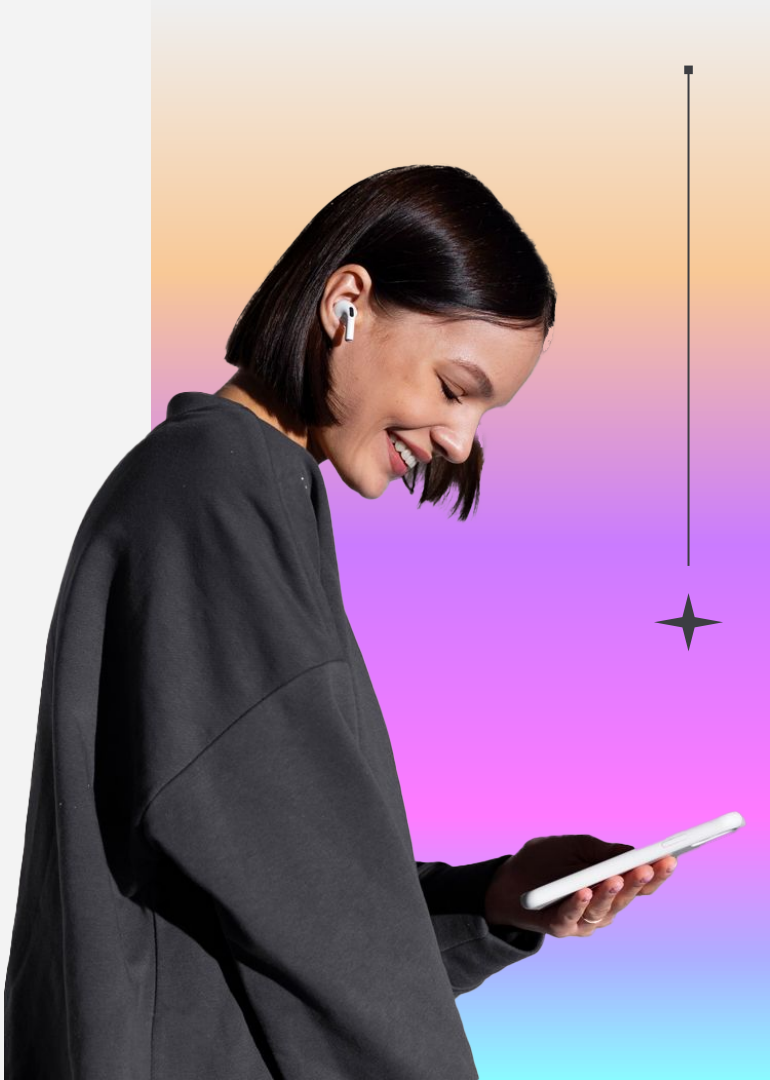


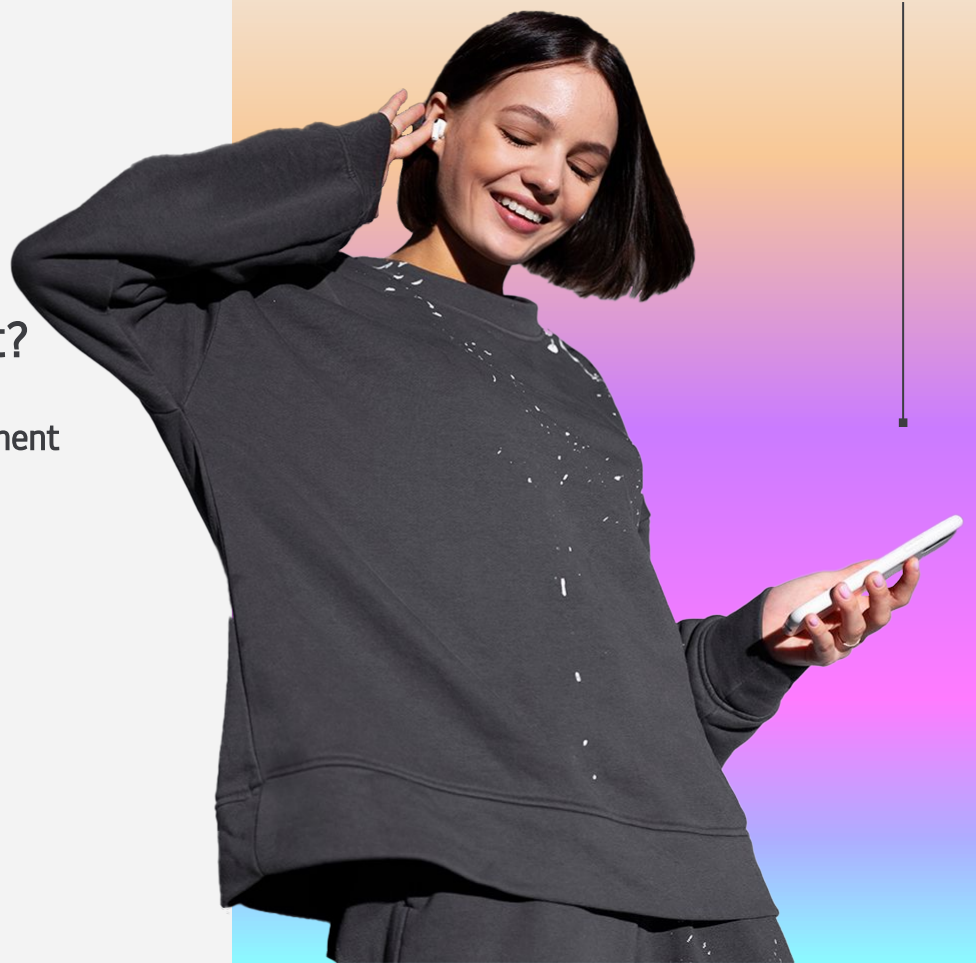
TABLE OF CONTENTS

BUSINESS OVERVIEW

New to the headphone market?

Be prepared with Topic Modeling and Sentiment Analysis

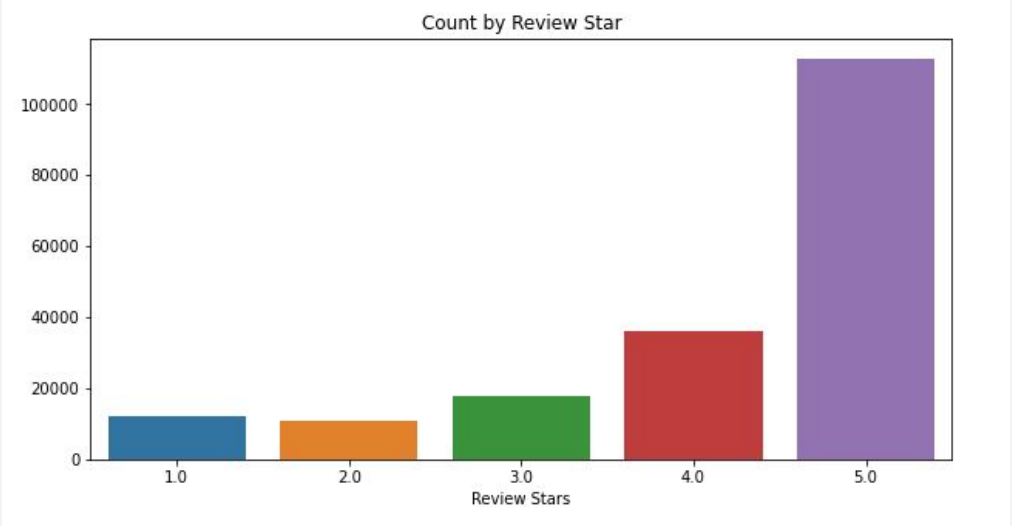
Product reviews can provide better understanding of customer feedback and preferences



DATA UNDERSTANDING

AMAZON PRODUCT REVIEWS

- Around 130,000 reviews
- Across 395 different products and 169 different brands
- Spanning from June 2000 to October 2018
- Around 60% of total reviews have 5 star ratings
- Time consuming to process manually



TOPIC MODELING

IDENTIFYING TOPICS FROM WORD CLOUDS

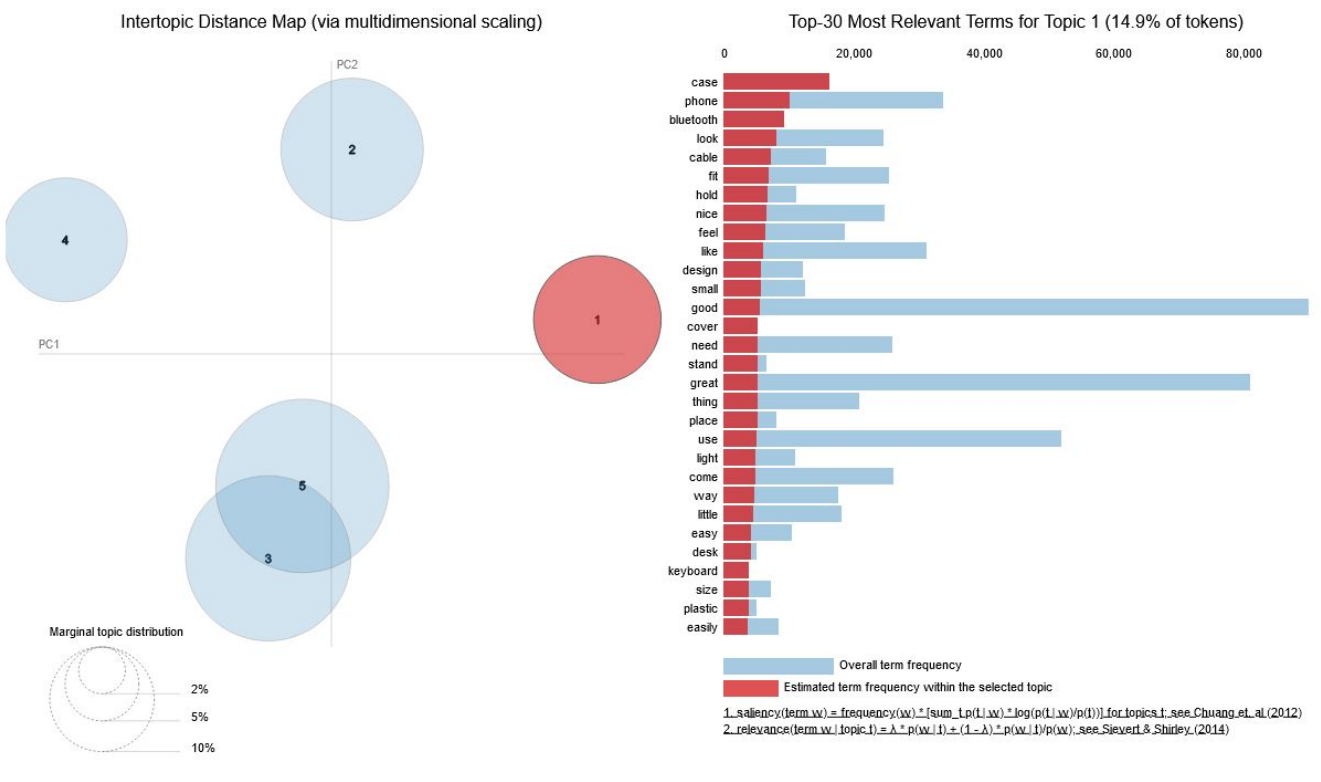
- Provides quick summary of most frequent words
- Limited insights available
- Difficult to associate context with words

TOPIC MODELING WITH LDA

- Statistical approach to automatically find underlying topics
- Assumes each review is made up of different topics and each topic is made up of words that frequently appear together in all reviews

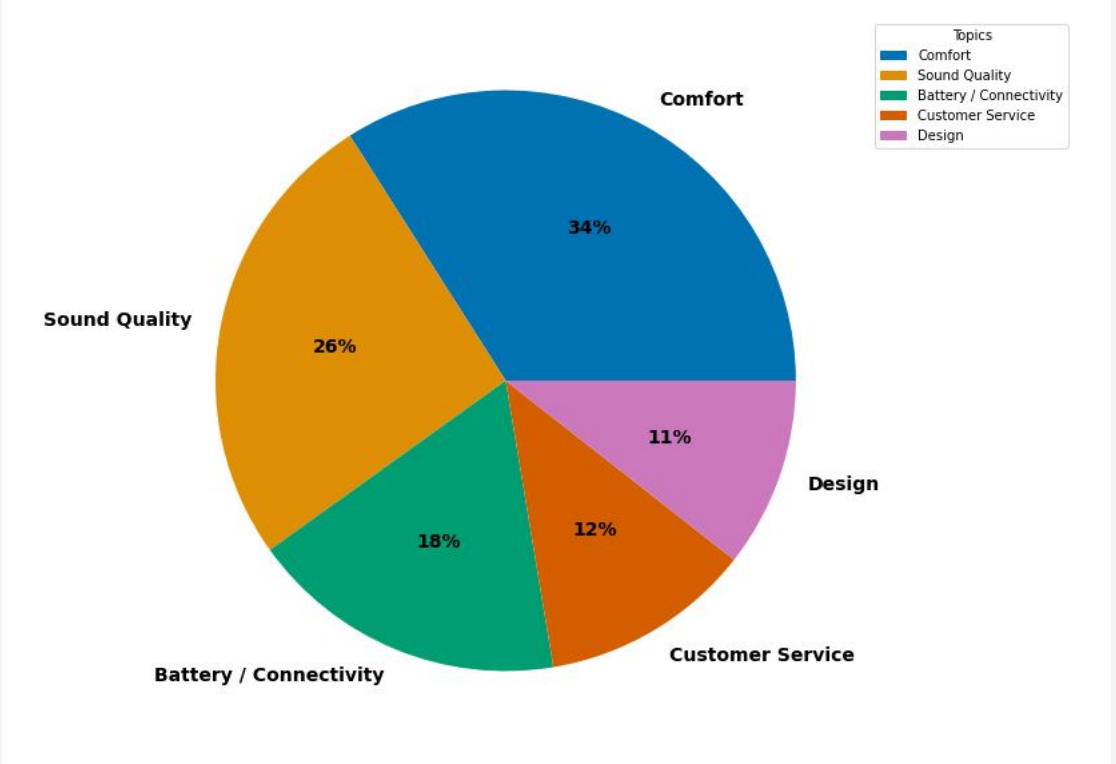
Discovered 5 topics:

- Design
- Battery / Connectivity
- Sound Quality
- Customer Service
- Comfort



DISTRIBUTION OF TOPICS

- Customer preferences can be discovered by grouping reviews according to their topics
- Comfort was the most discussed topic making up 34% of total reviews
- Sound Quality is the next most discussed topic at 26% of total reviews



SENTIMENT MODEL

TF-IDF VECTORIZER

Algorithm to transform text data into numerical vectors

LOGISTIC REGRESSION

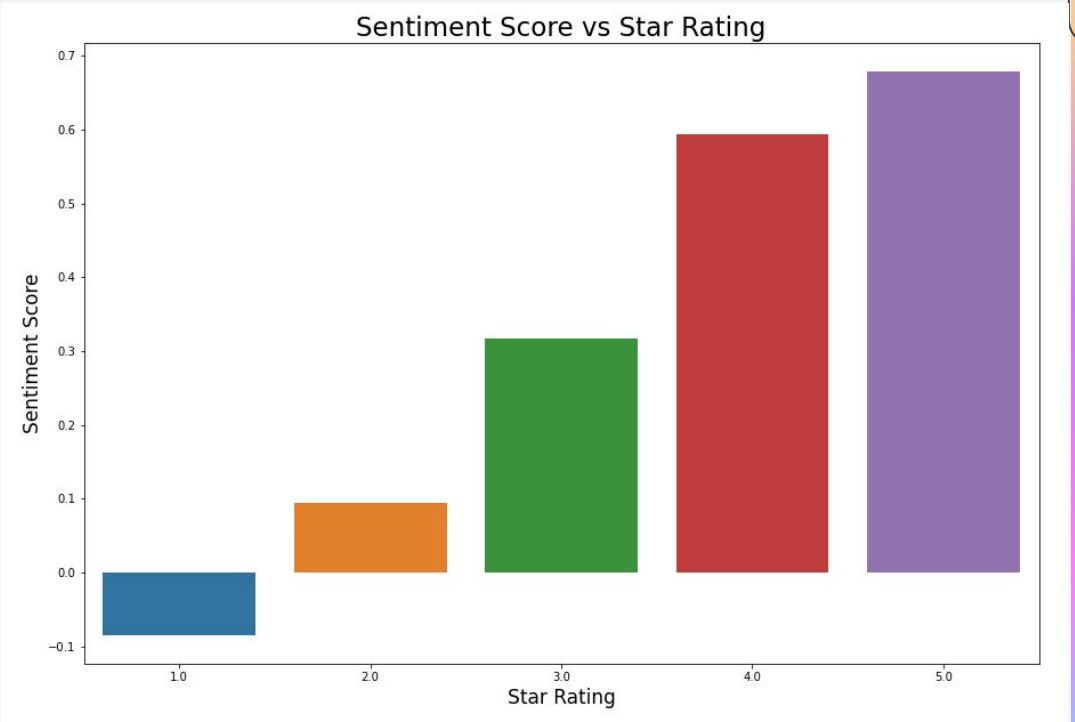
Classification algorithm to predict positive or negative sentiment

METRICS

89% Accuracy

SENTIMENT ANALYSIS

- Each review was analyzed and scored base on the sentiment in the text
- Score ranges from -1 (most negative) to 1 (most positive
- As star rating increases sentiment score increases from negative to positive

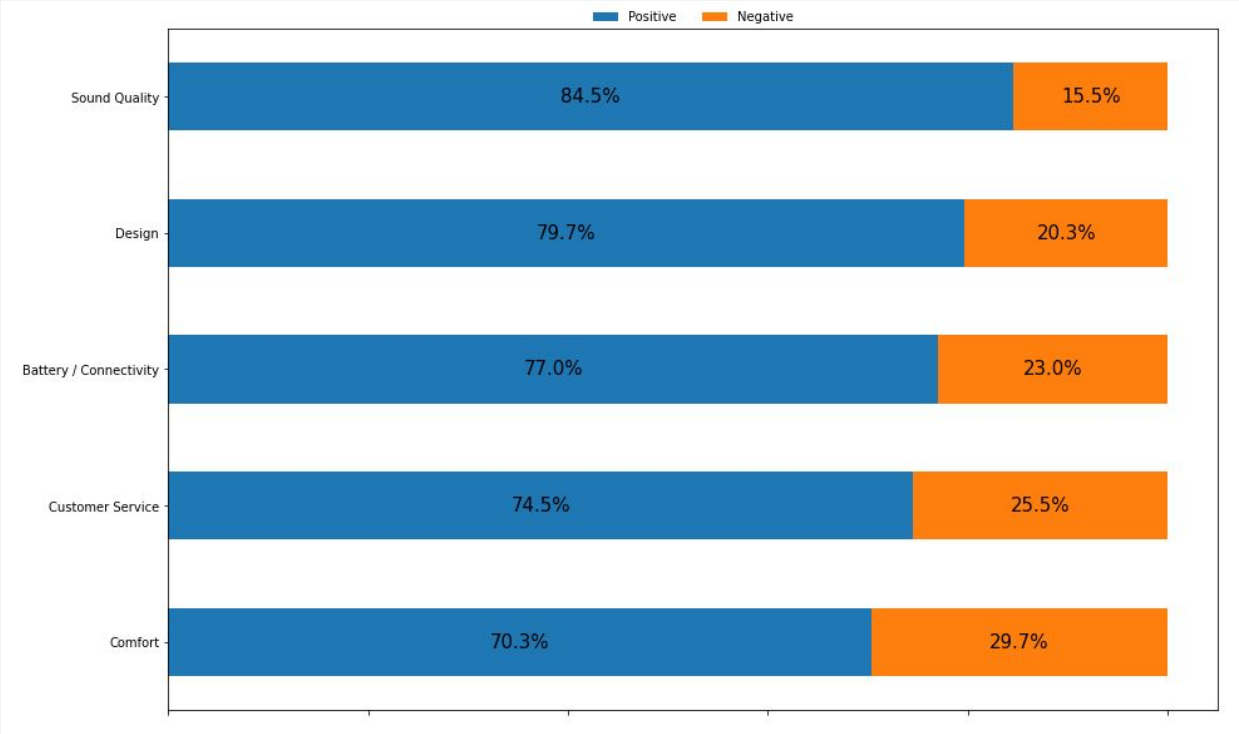


MOST POSITIVELY SCORED WORDS

MOST NEGATIVELY SCORED WORDS

SENTIMENT OF TOPICS

- Combining both methods can give us deeper insights into how customers feel about the product
- Most discussed topic comfort has the highest percent of negative reviews
- Sound Quality contained most positive reviews



RECOMMENDATIONS

Focus on Customizable Comfort

- Offer various sized earbud covers or headphone cushions for products to ensure perfect fit

Stepping up Customer Service

- Important keywords from customer service topic were found to be replacement, warranty, update review

Adjust Marketing Strategies

- Advertising longer battery life or better bluetooth connections can shift customers dissatisfied with other products to yours

CONCLUSION

Limitations

- Nature of Amazon reviews may mislead results
- Sentiment model used only takes into account frequency of words. Limits understanding of full context of text

Next Steps

- Top2Vec for topic modeling, to take advantage or pretrained embedding models
- LSTM model for classification

THANKS!

Do you have any questions?

Email:
to.wesleyyu@gmail.com

Github:
https://github.com/whyyoutoo/Sentiment_Classification_for_Amazon_Product_Reviews