

# Ring Doorbell

Semantic analysis of amazon reviews

Petr Andreev [pand269@aucklanduni.ac.nz](mailto:pand269@aucklanduni.ac.nz)



# A little about myself

- Second year 2 UoA student majoring in Data Science
- Always was fascinated by AI&ML

**Future goals:** learn from professionals in the field, possibly become an MSA student assistant

## My hobbies:

(that look good on presentation slides)

- Diving & Drone photography

## Best part of MSA:

Getting descent score on Kaggle

## Possible Improvements:

Releasing the topic of project in advance, e.g. NLP, or RNN.



**Heights with respect to sea level**

# The problem

## Understanding customers and Gaining market share

- Global smart doorbell market has potential to grow **\$1.2Billion** during 2020-2024. With estimated **year-over-year** growth of **20.5%**
- Ring aims to gain potential market share of Doorbells by addressing **key customer pains** and **gains**
- **Ring's** parent company is Amazon. Thus **Ring** completes large portion of its sales on Amazon E-Commerce website.
- Thus customer reviews on Amazon are representative of **Ring's** general **customer base**.
- Therefore, Ring's rating on Amazon can directly impact its **potential to gain new customers**
- We analyse customer reviews to find **key areas** that **Ring** has to focus on improving.



# Data Extraction

Data was scraped from **Amazon website** using **Scrapy**.

For each scraped comment the **following data was scraped**:

- Title
- Content
- Rating of Ring product
- Number of customers who found comment helpful
- Date and country of author

A typical comment on Amazon

logchief

★☆☆☆☆ Misleading on battery!!!

Reviewed in the United States on June 9, 2020

Color: Venetian Bronze | Configuration: Doorbell only | Verified Purchase

Misleading and very displeased...I expected a battery removal but now real gen meaning not the ring 2 with removal battery...RIDICULOUS!!!! Now I ha bought that I can't use...pissed off!

205 people found this helpful

Helpful | Comment | Report abuse

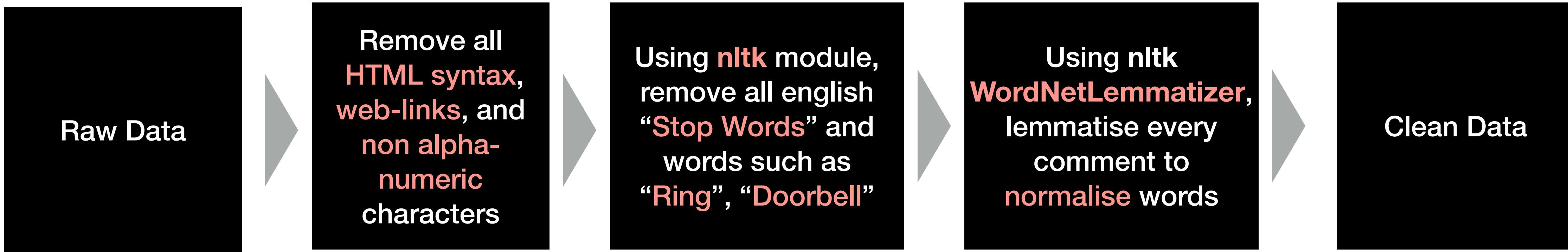
## First rows from scraped data

Score	Title	Helpful	Date	Comment
1.0 out of 5 stars	I tried, I really tried	4,360 people found this helpful	Reviewed in the United States on September 27, 2018	<span>Well, I tried – I really Tried. But it turns out my expectations w</span>
1.0 out of 5 stars	Piece of junk	2,494 people found this helpful	Reviewed in the United States on October 28, 2017	<span>When I first got the device in July we immediately installed the p</span>

# Data Cleaning

## First rows of raw scraped data

Score	Title	Helpful	Date	Comment
1.0 out of 5 stars	I tried, I really tried	4,360 people found this helpful	Reviewed in the United States on September 27, 2018	<span> Well, I tried – I really Tried. But it turns out my expectations w</span>
1.0 out of 5 stars	Piece of junk	2,494 people found this helpful	Reviewed in the United States on October 28, 2017	<span> When I first got the device in July we immediately installed the p</span>

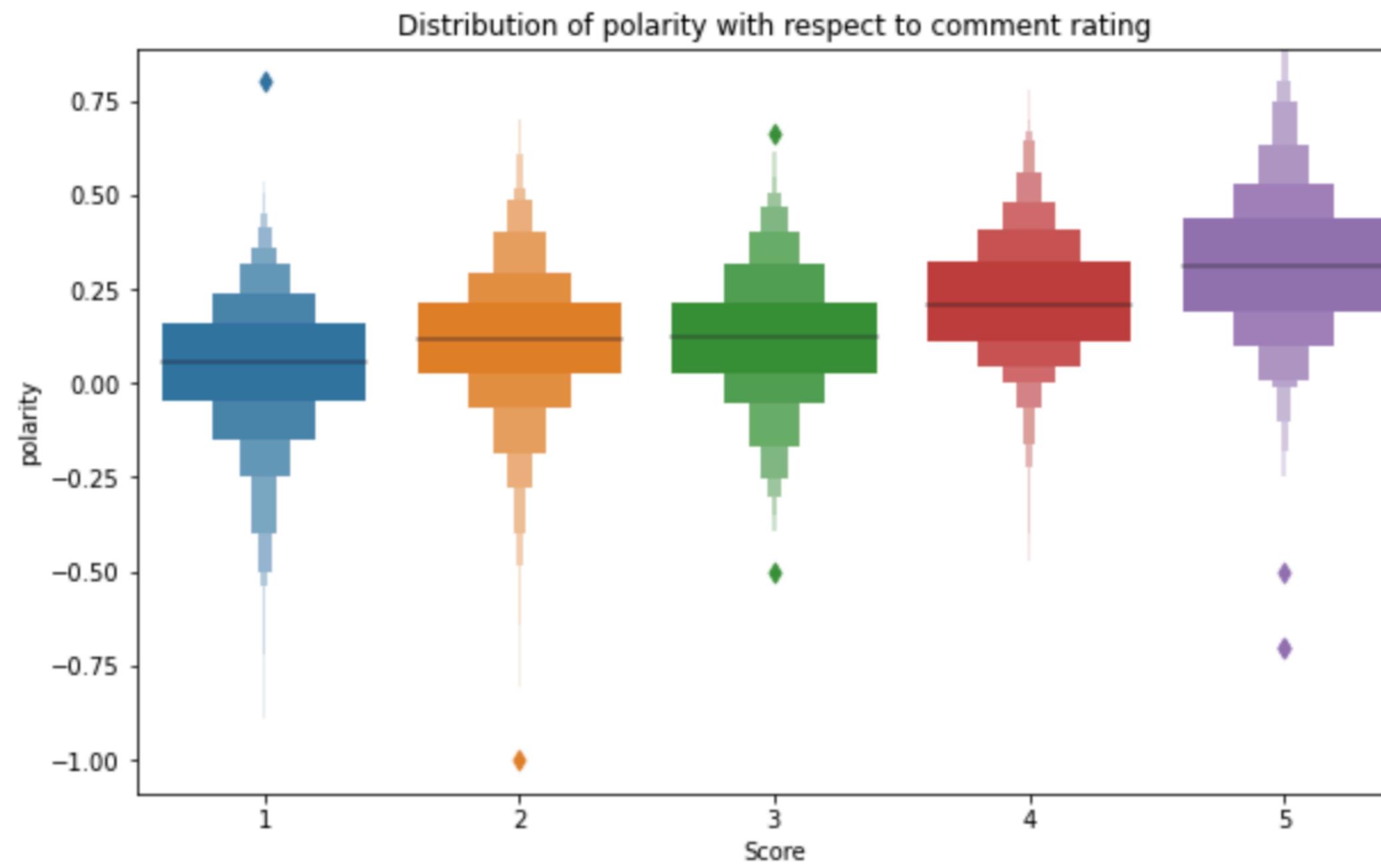


## First rows of clean data

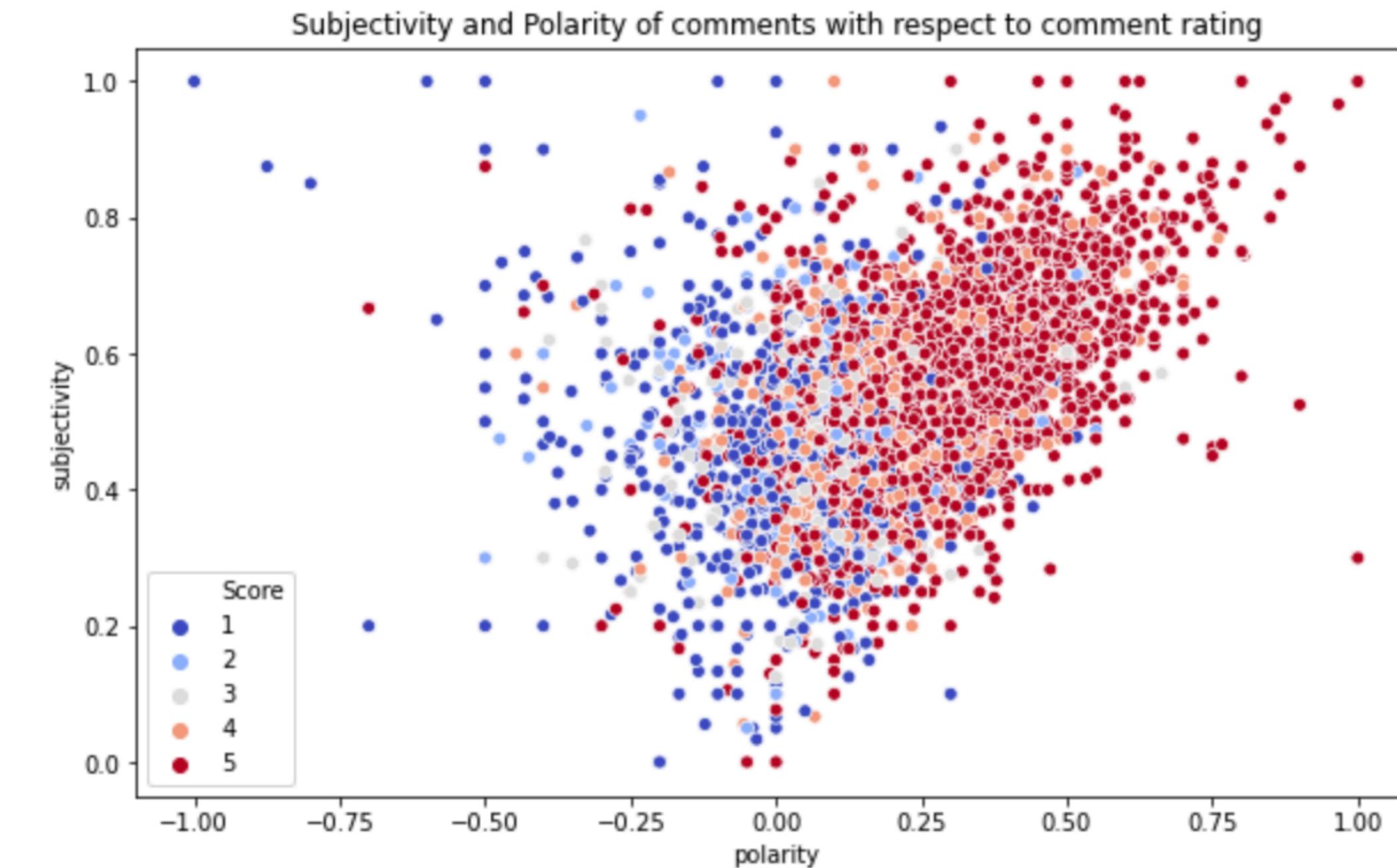
Score	Title	Helpful	Date	Country	StrippedComment
1	tried really tried	4360	2018-09-27	United States	well tried really tried turn expectation way high mesmerized com
1	piece junk	2494	2017-10-28	United States	first got device july immediately installed product install went gre
1	look great hidden cost gotcha	3385	2018-01-24	United States	work well install easy one major complaint blurb amazon explain
4	paid le 6 month updated october 2018	4507	2017-02-22	United States	october 2018 update revisiting review making 4 star would rande

# Data Exploration

Negative reviews don't have strong negative polarity



Positive comments seem more subjective



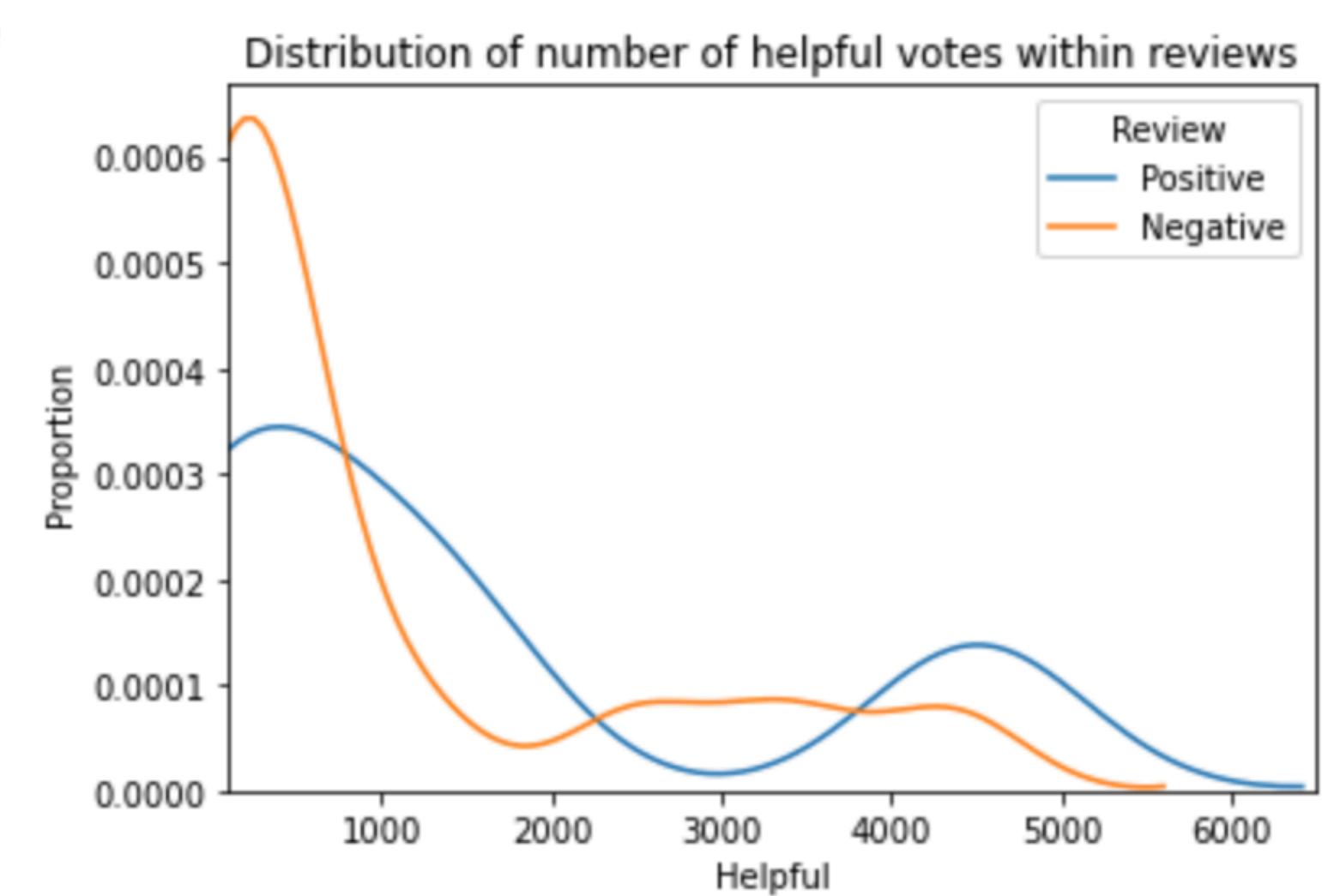
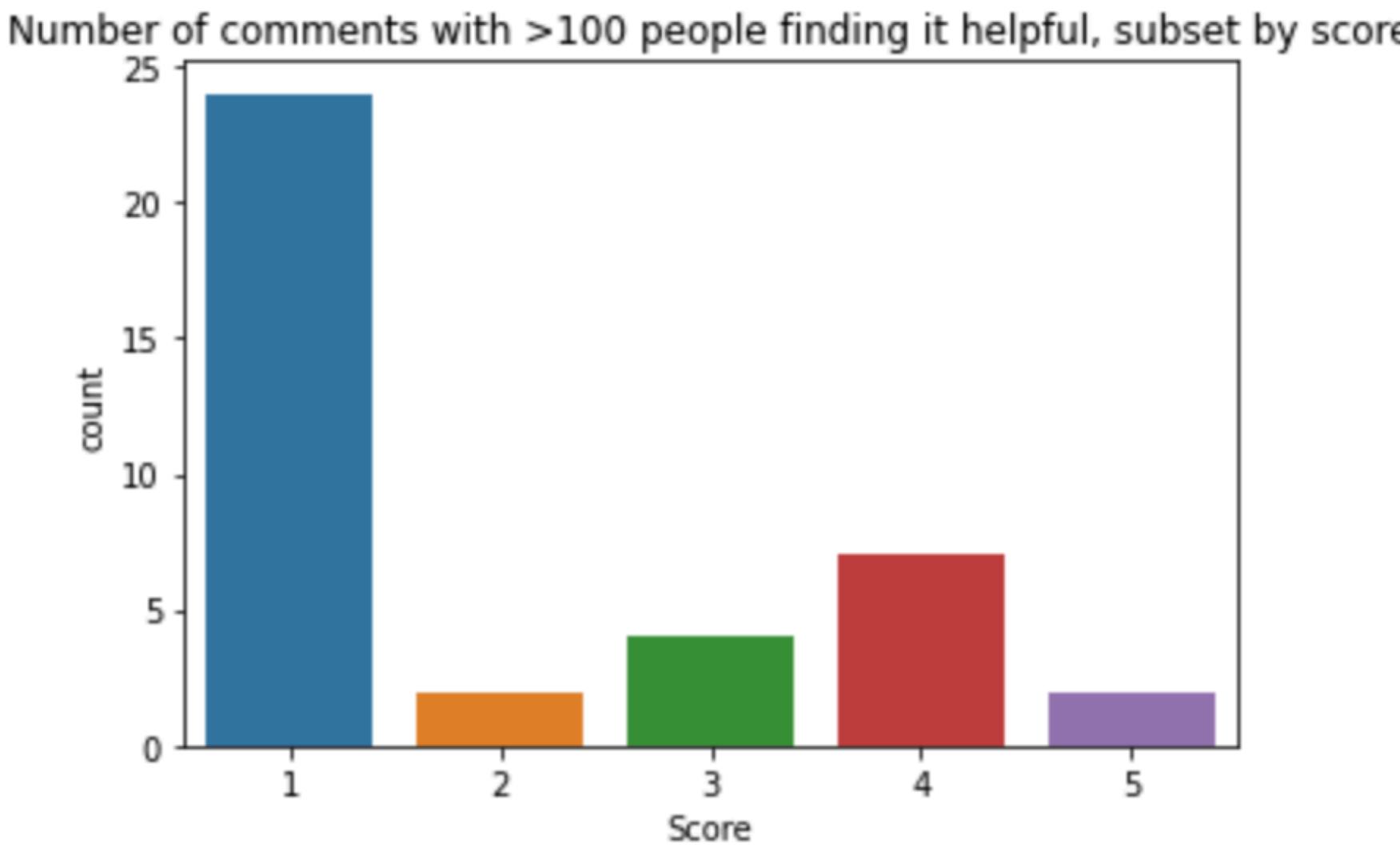
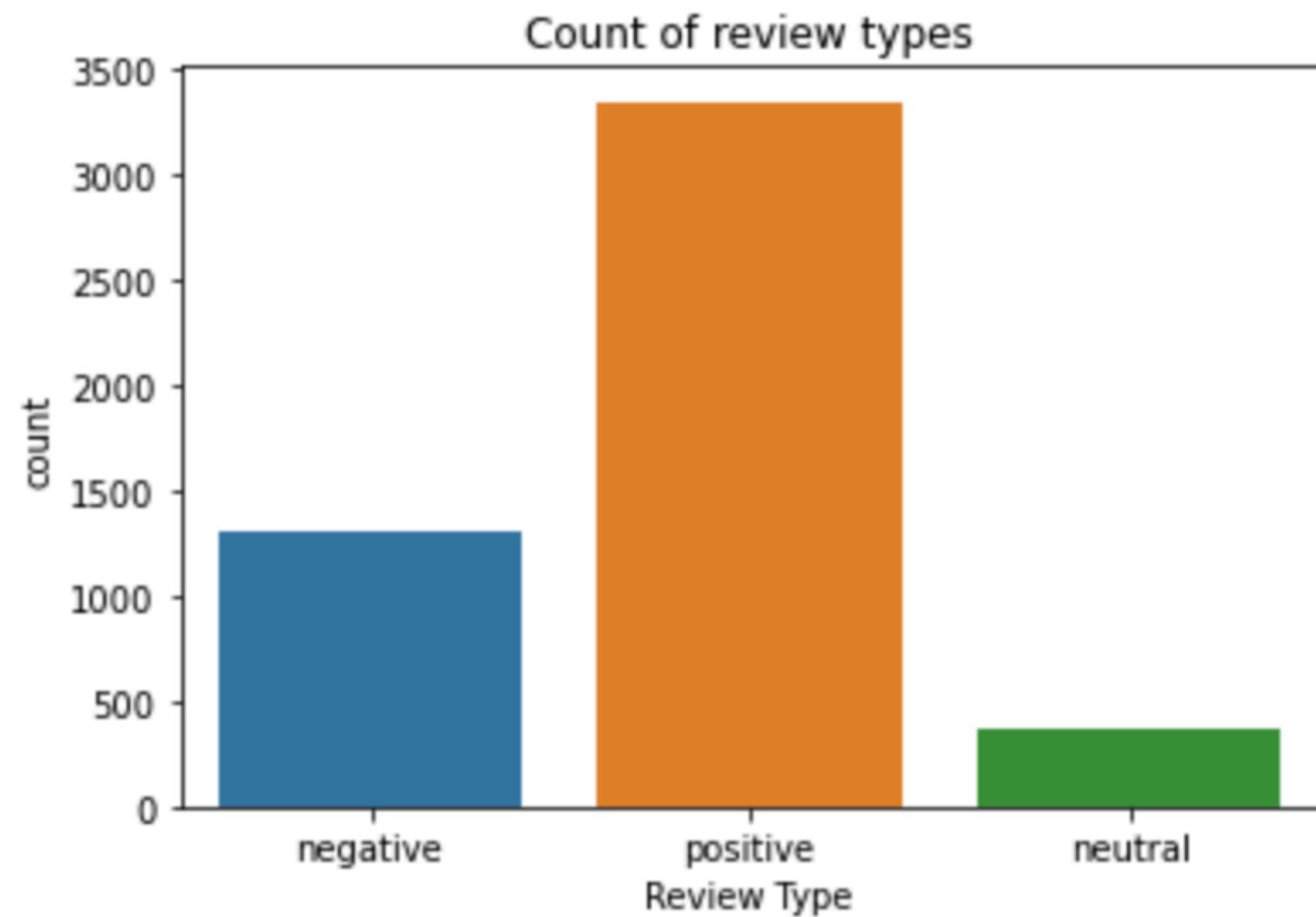
Negative reviews appear to be **less subjective** than positive reviews, thus can be found **more valuable**, since complains are **more objective** and **less emotional**.

We now test the following hypothesis using two sample t-test:

```
ttest_ind(reviews[(reviews['Score'] > 3)].subjectivity, reviews[(reviews['Score'] < 3)].subjectivity)  
Ttest_indResult(statistic=19.399089937388634, pvalue=1.1350511380663544e-80)
```

With **P-value is <0.05** thus it is clear that the difference is **statistically significant** and feedback from negative comments is more objective than from positive. Therefore, negative reviews will contain objective feedback.

# Data Exploration



Most “helpful” votes are given by potential customers who **haven’t bought** the product but were looking at reviews. Thus it is clear that Ring is **loosing a large potential customer base** due to small number of very “helpful” negative reviews.

Therefore, addressing the needs of unsatisfied customers is **a priority**, if Ring wants to **gain market share**.

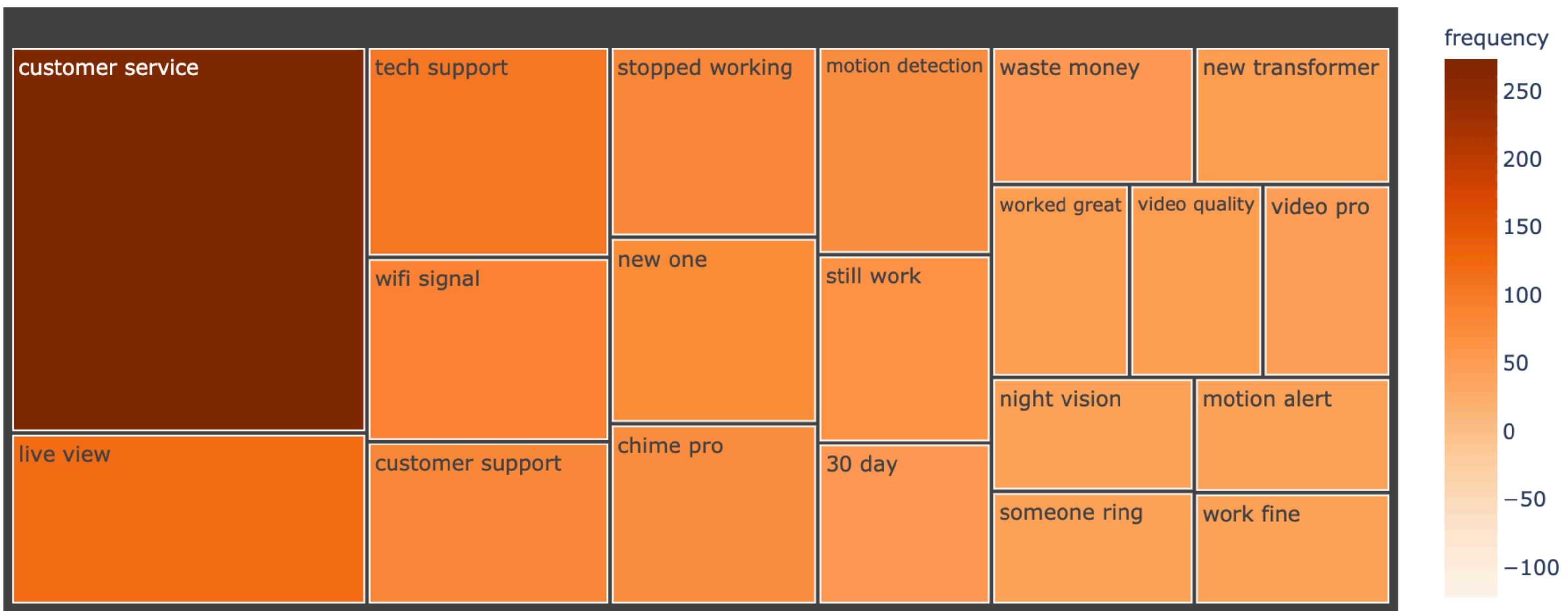
# Data Analysis

Most **common bigrams** show that both satisfied and dissatisfied customers are experiencing **similar issues**, showing that **Ring** has to focus on improving the following **top 5 features**:

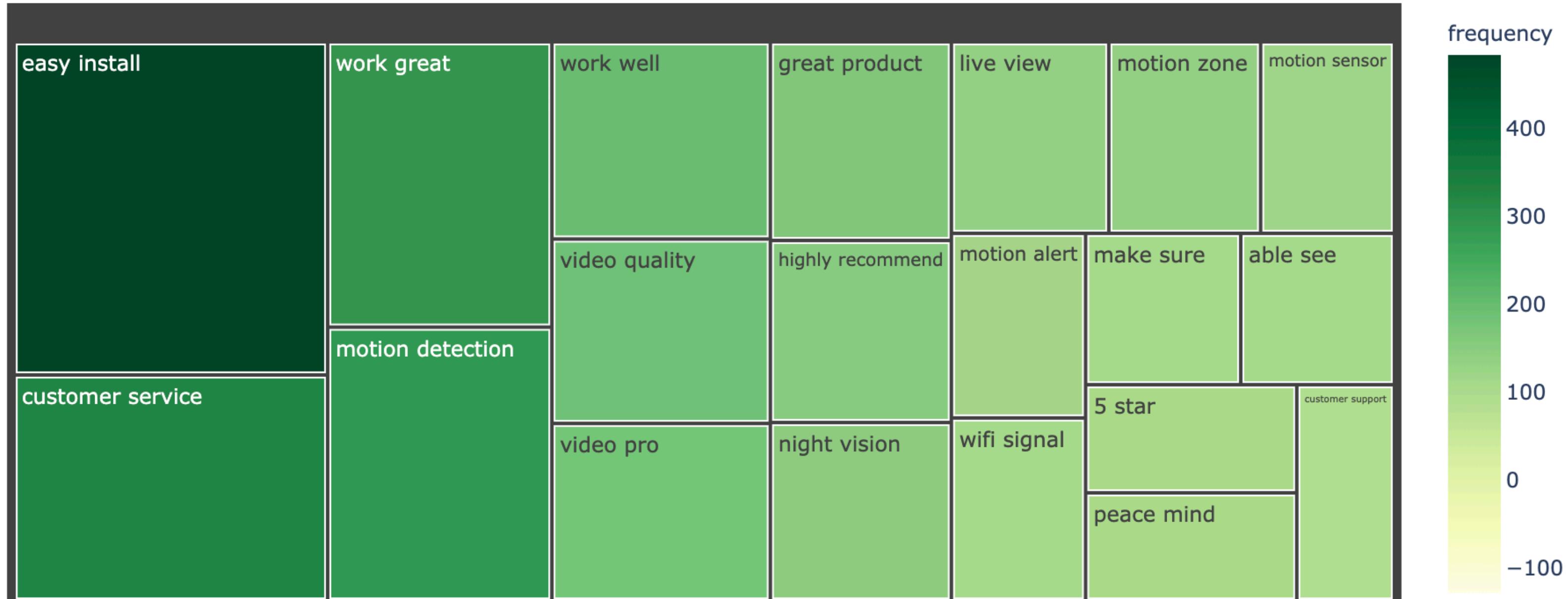
- Customer Service / Tech Support
- WiFi Signal reception
- Motion Detection
- Night Vision
- Pro subscription for Video

Credits for visualisation ideas to Tanul Singh:  
<https://www.kaggle.com/tanulsingh077/twitter-sentiment-extraction-analysis-eda-and-model>

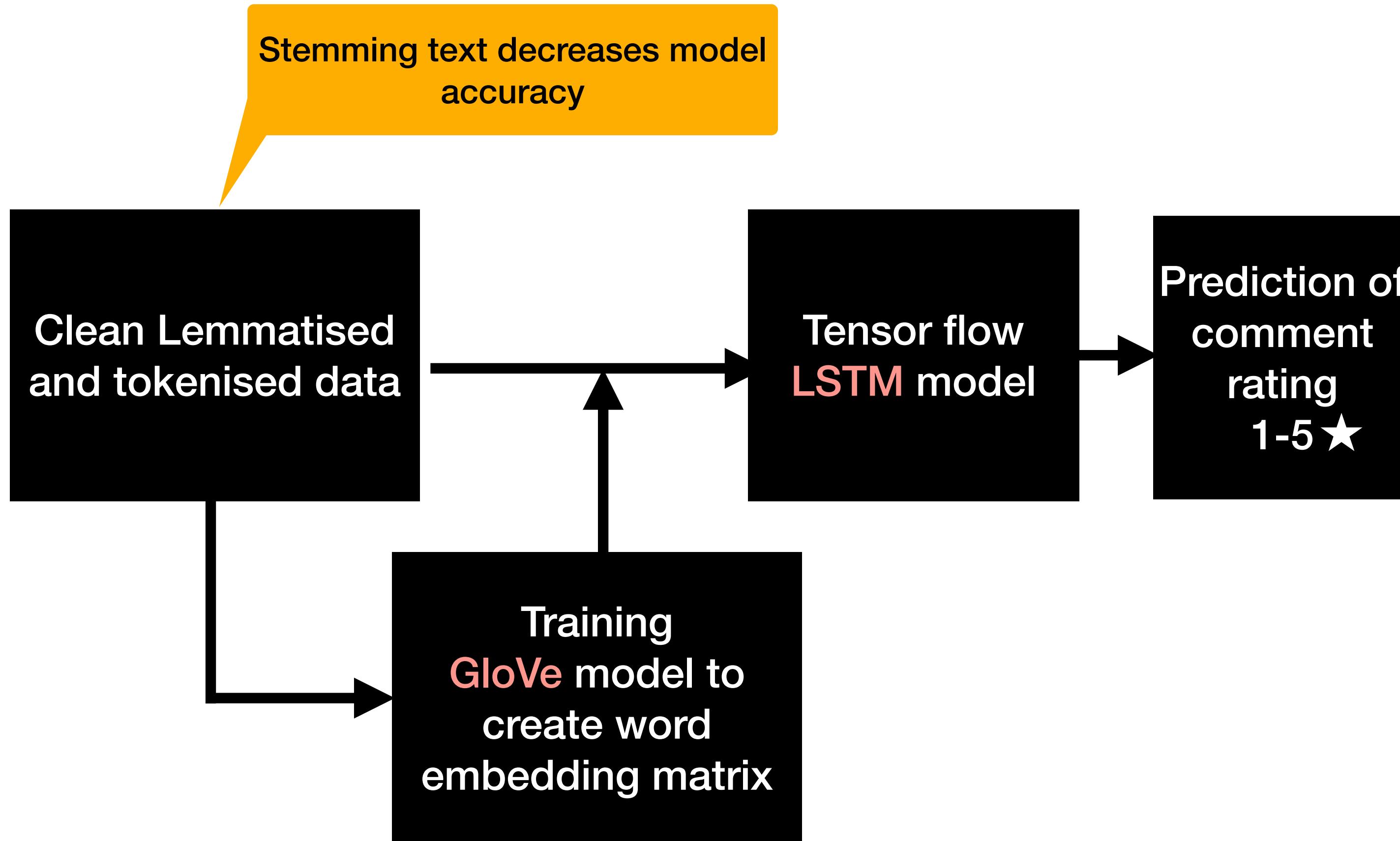
Tree Of Most Common Negative Bigrams



Tree Of Most Common Positive Bigrams



# Sentiment Analysis



## TF Model Structure:

- **Embedding Layer** with weights set to GloVe embedding matrix
- **Dropout Layer**
- **LTSM Layer** with 50 neurones
- **Dense Layer** of 3 neurones for prediction output and sigmoid activation function.

Model loss is **categorical cross-entropy**, with **adam** optimiser.

Trained in 5 epochs, with early stopping callbacks on **validation accuracy** with 0 patience.

Credits to Anshulrai for reading GloVe vectors into Python: <https://www.kaggle.com/anshulrai/cudnnlstm-implementation-93-7-accuracy>

Credits to Paolo Ripamonti for Sentiment analysis model structure: <https://www.kaggle.com/paoloripamonti/twitter-sentiment-analysis>

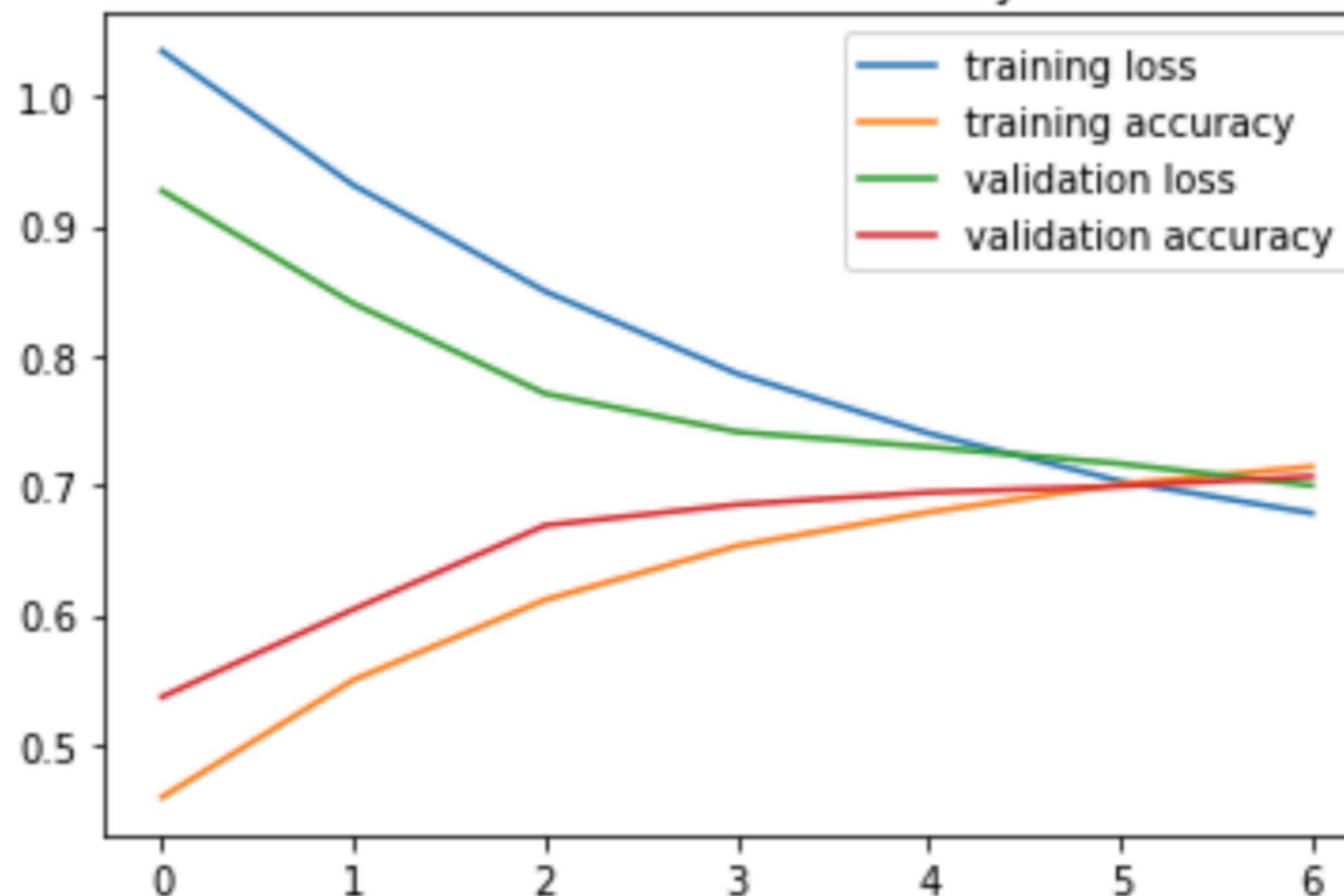
# Kaggle Challenge

## Model Development Stages



Seemed to be preferred by Kaggle contributors for sentiment analysis

## Model Loss and accuracy



# Advanced Requirement

- 1. The use of web scraping to gather data:* Used **Scrapy** to gather data from Amazon Website
- 2. Using model trained from Kaggle, apply it to your business case. Compare and comment on differences in predictions between model and prebuilt package:*

Recall and precision of model is **0.0 on neutral set.**

The sample of neutral comments is only 11% of the dataset, thus **resampling** has to be used to improve model predictions.

Recall and precision of model is much better than TextBlob, which on opposite mainly recalls neutral comments with very poor precision (0.08), leading to much better precision for negative and positive since they are only chosen when model is very sure about it.

Overall my model can be considered to give better predictions. Possibly due to **poor performance** of text blob on **multi-class datasets**

	My model predictions	precision	recall	f1-score	support
		negative	0.72	0.67	0.69
		neutral	0.00	0.00	0.00
		positive	0.82	0.94	0.87
		accuracy		0.79	1657
		macro avg	0.51	0.53	0.52
		weighted avg	0.73	0.79	0.76
	Text blob predictions	precision	recall	f1-score	support
		negative	0.66	0.28	0.39
		neutral	0.08	0.78	0.15
		positive	0.94	0.27	0.42
		accuracy		0.31	1657
		macro avg	0.56	0.44	0.32
		weighted avg	0.80	0.31	0.39

- 3. Achieve Kaggle score of 64% or higher:* Score **68.9%**

# Final Comments

**Thank you!**

## Bibliography

Credits to Prateek Joshi for implementation of Word2Vec model: <https://www.analyticsvidhya.com/blog/2019/07/how-to-build-recommendation-system-word2vec-python/>

Credits to Anshulrai for reading GloVe vectors into Python: <https://www.kaggle.com/anshulrai/cudnnlstm-implementation-93-7-accuracy>

Credits to Paolo Ripamonti for Sentiment analysis model structure: <https://www.kaggle.com/paoloripamonti/twitter-sentiment-analysis>

Credits for visualisation ideas to Tanul Singh: <https://www.kaggle.com/tanulsingh077/twitter-sentiment-extaction-analysis-eda-and-model>

Data about doorbell industry: <https://www.technavio.com/report/smart-doorbell-market-industry-analysis>