



Web Scrapping and Analysis of Online Product Reviews

May 15, 2024

Aleika Chery and Tahsina Khan





Presentation **Agenda**

01 Introduction

04 Visualizations

02 Data
Collection

05 Next Steps

03 Data Processing





01

Introduction

Motivation, Problem Statement



Background

"Many businesses and customers rely on product reviews to improve their products and make the best purchases respectively. However, there is an immense amount of data available and analyzing each review is impractical and time-consuming. "

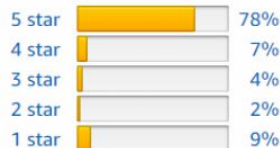
Objective

Our objective is to create a tool that classifies product reviews from Amazon and derives meaningful insights from them.

Customer Reviews

★★★★☆ 1,069

4.5 out of 5 stars



[See all 1,069 customer reviews](#)



Amelia

★★★★☆ **Not for me**

Reviewed in the United States on May 2, 2024

Size: 3.38 Fl Oz (Pack of 1) | **Verified Purchase**

I didn't feel like this product did anything for me.



AMILLER-RN

★★★★★ **LOVING MY SKIN AGAIN FEELS SO GOOD!!!** 😊

Reviewed in the United States on April 21, 2024

Size: 3.38 Fl Oz (Pack of 1) | **Verified Purchase**

I'm no beauty guru or beauty product expert. I'm just a 47 yr old woman struggling trying new products and treatments with hopes of reversing some of the unsightly sun while covered in baby oil when I was in my teens and 20's trying to get a dark tan

Problems we want to address

Motivation

Volume and Variety:

Amazon hosts an immense number of product reviews spanning a wide range of products

Subjectivity and Bias:

Reviews on Amazon can be highly subjective and biased, making it challenging to extract unbiased insights

Data Quality and Noise:

The quality of reviews can vary significantly, with some being overly positive or negative, irrelevant, or even fake

Temporal Dynamics:

Trends and sentiments around products can change over time due to various factors such as product updates, market trends, or external events.

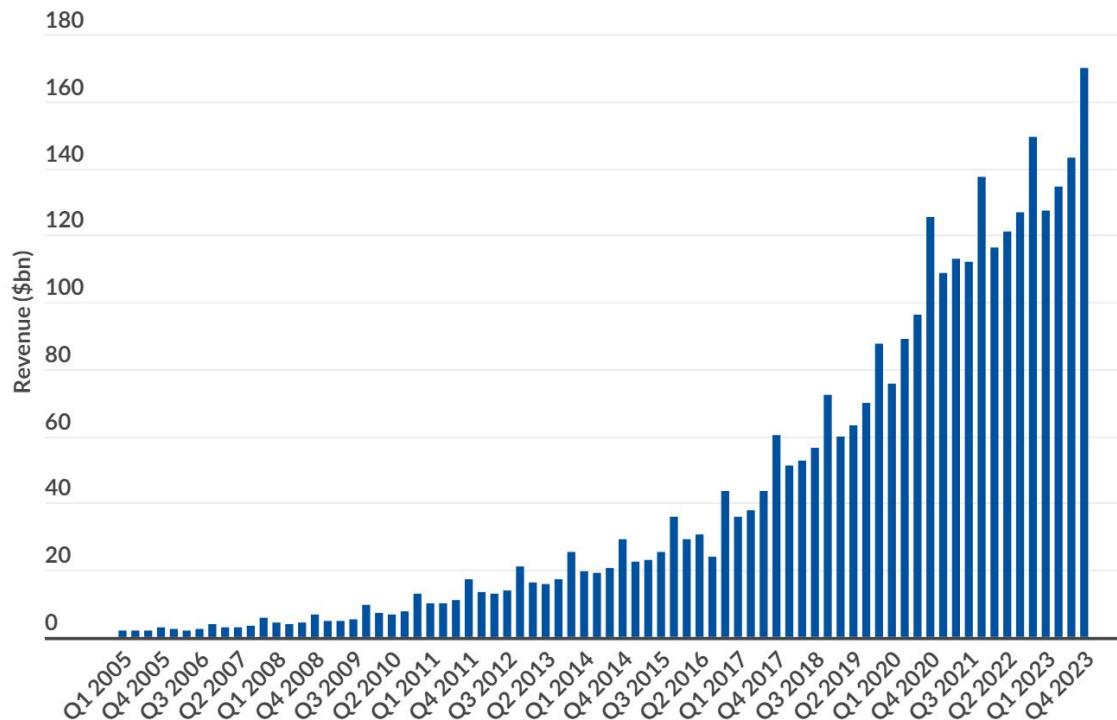


2 billion users

In March of 2023, Amazon attracted 2.4 billion visitors
(Similarweb, 2023).

**It is the fourth most popular site in the United States, and
the 13th ranked site worldwide.**

Amazon quarterly revenue 2005 to 2023 (\$bn)



Resources we leveraged



BeautifulSoup



kaggle





02

Data Collection

Gathering Amazon reviews for our dataset

Obstacles encountered:

- What methods will we use to web scrape the data?
- What category of products should we target?
- What should be the ratio of negative to positive reviews?
- How many reviews are sufficient for an effective future model?
 - Large & diverse training dataset helps the model generalize better
 - Social media/online reviews usually have noisier data



Logitech Combo Touch Detachable 10th Gen iPad Keyboard Case with Large Precision Trackpad, Full-Size Backlit Keyboard, and Smart Connector Technology - Oxford Gray

Visit the Logitech Store

4.5 ★★★★★ 942 ratings | Search this page

1 sustainability feature


1K+ bought in past month

-32% \$108⁹⁵

List Price: \$159.99

Or \$10.95 /mo (12 mo). Select from 1 plan

FREE Returns



Apple iPhone 14 Pro, 128GB, Space Black - Unlocked (Renewed)

Visit the Amazon Renewed Store

4.2 ★★★★★ 1,505 ratings | Search this page

1 sustainability feature

300+ bought in past month

Typical price: \$688.00 Details

Price: \$649.95

FREE Returns

You Save: \$38.05 (6%)

Or \$38.38 /mo (24 mo). Select from 1 plan

Pay \$54.16/month for 12 months, interest-free upon approval for Amazon Visa

Service Provider: Unlocked

AT&T GSM Carriers T-Mobile Unlocked Verizon

Color: Space Black

Size: 128GB

1TB 128GB 256GB 512GB



SAMSUNG Galaxy S24 Cell Phone, 256GB AI Smartphone, Unlocked Android, 50MP Camera, Fastest Processor, Long Battery Life, US Version, 2024, Amber Yellow

Visit the SAMSUNG Store

4.4 ★★★★★ 1,038 ratings | Search this page

Amazon's Choice in Cell Phones by SAMSUNG

1 sustainability feature

100+ bought in past month

\$859⁹⁹

Or \$50.78 /mo (24 mo). Select from 1 plan

FREE Returns

Pay \$71.67/month for 12 months, interest-free upon approval for Amazon Visa

Available at a lower price from other sellers that may not offer free Prime shipping.

Color: Amber Yellow

Size: 256GB

128GB 256GB 512GB

Amazon Review Dataset

- 530 line dataset

Features:

- Index
- Profile_name
- Rating
- Rating_date
- Title
- Review_text
- Sentiment

Sentiment:

- Positive reviews: 299
- Negative reviews: 231

index	profile_name	rating	rating_date	title	review_text	sentiment
0	Adrian	5.0 out of 5	Reviewed i	5.0 out of	least, I mean you can connect the device to a monitor, mouse and	0
1	Tatiana	5.0 out of 5	Reviewed i	5.0 out of	keyboard via some 5-in-1 Anker for example, and here you just got	0
2	Lauren	5.0 out of 5	Reviewed i	5.0 out of	yourself a laptop with the size of a phone. OneUI is really the most rich	0
3	Sarah	5.0 out of 5	Reviewed i	5.0 out of	software I've used so far, and before using it I was skeptical if I'd like it	0
4	CoryBaxter	1.0 out of 5	Reviewed i	1.0 out of	or not, but I'm really getting used to it, and the features like	1
5	Walter Ker	5.0 out of 5	Reviewed i	5.0 out of	multitasking in a true desktop way on the phone - this is something else	0
6	Maribel Lo	5.0 out of 5	Reviewed i	5.0 out of	(not the Dex, just phone multi window usage, that's AWESOME).The	0
7	Alex Beyer	4.0 out of 5	Reviewed i	4.0 out of	battery is nice, maybe I expected something more by viewing some	0
8	MAS	4.0 out of 5	Reviewed i	4.0 out of	reviews with almost 20 hours of screen time or so, but really it's fine for	1
9	frank	4.0 out of 5	Reviewed i	4.0 out of	a day of usage with QHD resolution, adaptive brightness, and app and	0
10	Michelle	3.0 out of 5	Reviewed i	3.0 out of	All usage during the whole day.The only disadvantages I want to	1
11	Rather B. F	5.0 out of 5	Reviewed i	5.0 out of	mention are the price tag that is high (let's be honest) and the box	0
12	drolson9	4.0 out of 5	Reviewed i	4.0 out of	contents, I mean why copy bad stuff from Apple and not include at	0
13	Dark Knigh	5.0 out of 5	Reviewed i	5.0 out of	least a small charging brick in the box? This is really uncool, especially	0
14	Alex	5.0 out of 5	Reviewed i	5.0 out of	within that price.I started with the display, but want to finish talking	0
15	Jazmin	5.0 out of 5	Reviewed i	5.0 out of	about it as well, it's HUGE, and I really like it, the corners are sharp-	0
16	A. Einstein	5.0 out of 5	Reviewed i	5.0 out of	ended, but that gives you the advantage of viewing a non-cropped-out	0
17	JiCar	4.0 out of 5	Reviewed i	4.0 out of	content, the resolution, the brightness, refresh rate... The display is a	0
18	Rhys E.	4.0 out of 5	Reviewed i	4.0 out of	killer really, I really don't get the people that complain about the color,	0
19	Tonya Hop	5.0 out of 5	Reviewed i	5.0 out of	or perform some tests on the device, etc.... no need.Overall I'm really	0
20	Ricky	4.0 out of 5	Reviewed i	4.0 out of	happy with my purchase, this is a great phone I can recommend to any	0
21	Kingcomp	5.0 out of 5	Reviewed i	5.0 out of	software and hardware enthusiasts like me, and maybe to iPhone	0
22	Vale Pinto	5.0 out of 5	Reviewed i	5.0 out of	users as well, to come out of the box and see that there are other great	0
23	Semper Fi	5.0 out of 5	Reviewed i	5.0 out of	phones out there in the world that can totally kill any iPhone with all the	0
24	marcello r	5.0 out of 5	Reviewed i	5.0 out of	features and stuff you get.	0
25	Florin	5.0 out of 5	Reviewed i	5.0 out of		0

index	profile_name	rating	rating_date	title	review_text	sentiment
0	0	Adrian	5.0 out of 5 stars	Reviewed in the United States on February 15, ...	5.0 out of 5 stars\nThe perfect phone that doe... \nPLEEEEEEASE people, stop complaining about th...	0
1	1	Tatiana	5.0 out of 5 stars	Reviewed in the United States on February 27, ...	5.0 out of 5 stars\nThe Pinnacle of Smartphone... \nIf there were ever a device deserving of a 1...	0
2	2	Lauren	5.0 out of 5 stars	Reviewed in the United States on May 1, 2024	5.0 out of 5 stars\nSkip the Negativity\n \nI read a lot of negative reviews about this ...	0
3	3	Sarah	5.0 out of 5 stars	Reviewed in the United States on May 9, 2024	5.0 out of 5 stars\nI love it!\n \nI loved my Galaxy S9 Plus and replaced it wi...	0
4	4	CoryBaxterWH	1.0 out of 5 stars	Reviewed in the United States on May 10, 2024	1.0 out of 5 stars\nDo NOT purchase the Samsun... \nSince my S22 Ultra died, I began phone shopp...	1



- Unnecessary features (i.e, profile name, rating_date)
- Redundant/disorganized formatting
 - Tags, repeated ratings

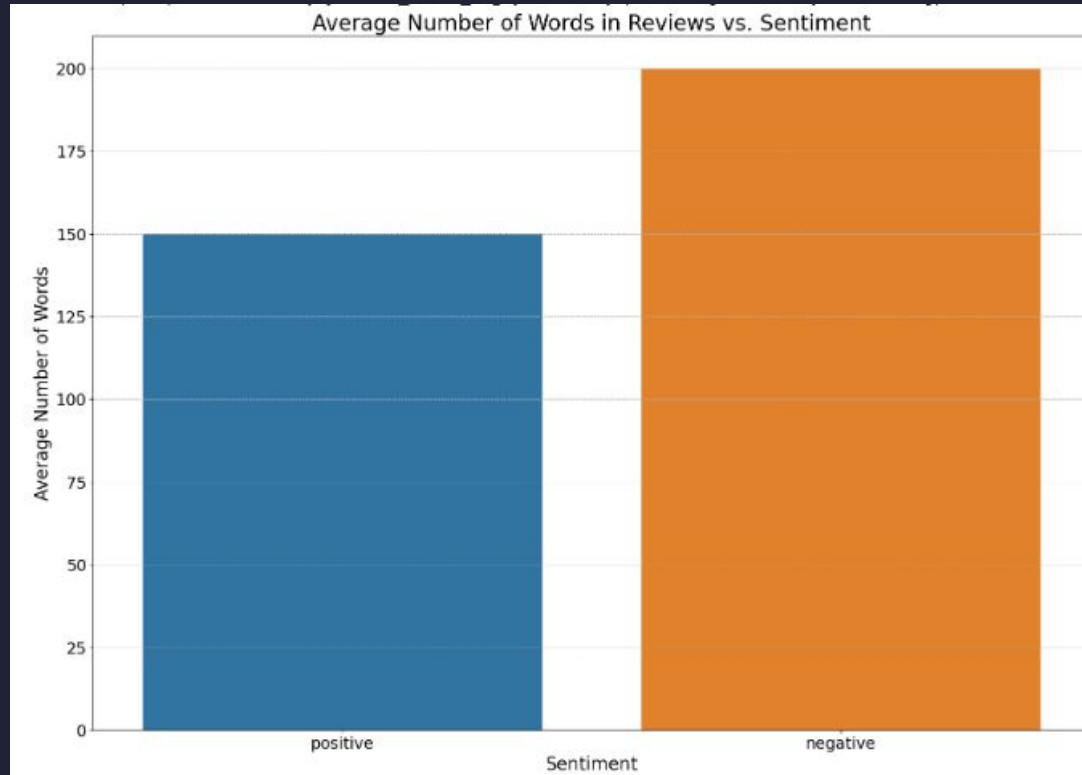


Amazon Review Dataset

- Resources used: BeautifulSoup, selenium, chrome webdriver, RainforestAPI

Issues with the data:

- Very noisy data
 - Duplicate reviews
 - Unnecessary features
 - Spam/fake reviews
 - Language variations
 - Inconsistent formatting
- 
- 





03

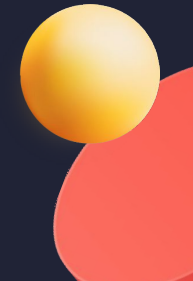
Data Processing

Exploring and cleaning our dataset



Data Preprocessing/Cleaning



- Removed noise using NLTK, spaCy, regex:
 - Punctuation
 - Numbers
 - HTML tags
 - Emojis
 - Converted all words to lowercase to reduce complexity of text
 - Filtered out non-english reviews
 - Removed stopwords
 - Lemmatized words
 - Tokenized & vectorized text to further machine comprehension
- 

Non-english reviews

```
df['language'].value_counts()
```

```
language
en      487
es       36
so        1
it        1
de        1
ro        1
da        1
cy        1
sl        1
Name: count, dtype: int64
```

```
df['language'].value_counts()
```

```
language
en      487
Name: count, dtype: int64
```

Word count of reviews

word_count_org	word_count_new
125	67
1218	743
1055	593
2788	1943
1949	1117

Reviews before and after

review_text	sentiment	new_reviews
iPhone delivered with a screen protector that ...	1	iphone deliver screen protector cover scratch ...
Whit out this damage it will be 5 stars	1	whit damage star
I was very excited to have found a "regular" w...	1	excited find `` regular `` waffle maker . ever...
Received item with a crack on the edge of the ...	1	receive item crack edge screen advertise `` ex...
So I use this everyday and managed to get cof...	0	use everyday manage get coffee day one still w...
I purchase this iPad just as transient iPad wh...	0	purchase ipad transient ipad recover ipad pro ...
Switching from my Note 10 Plus to the Samsung ...	0	switch note plus samsung ultra s big move I
This fit my iPad perfectly and it is compact! ...	0	fit ipad perfectly compact ! sturdy durable
I had an old iPad and no problems transferring...	0	old ipad problem transfer datum set up . love ...

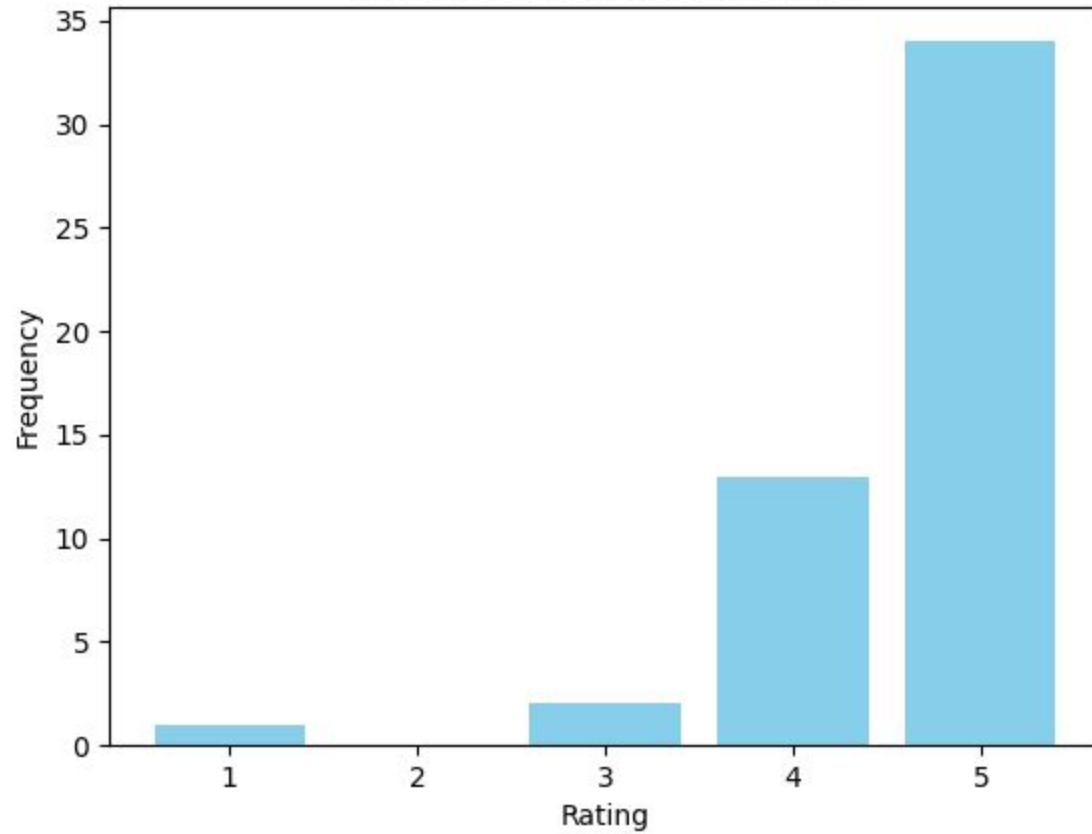
- Positive sentiment: 0
- Negative sentiment: 1



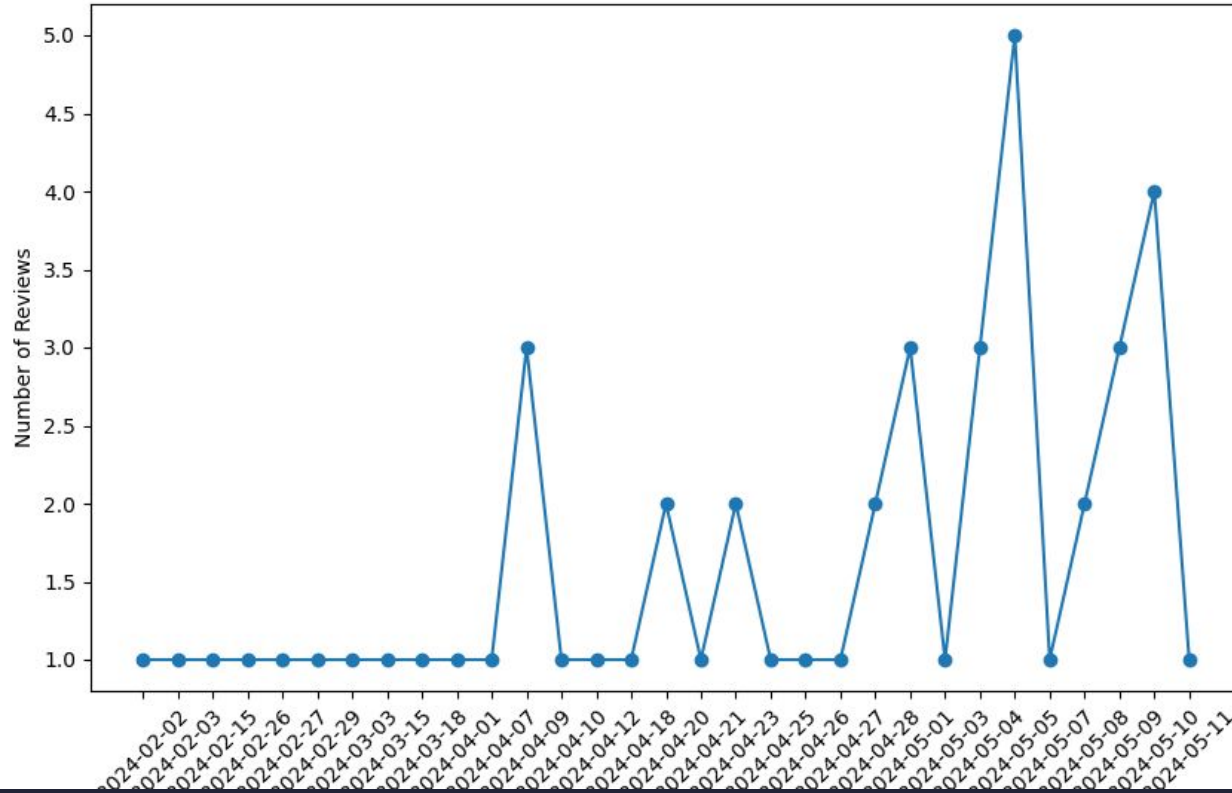
Data Visualizations

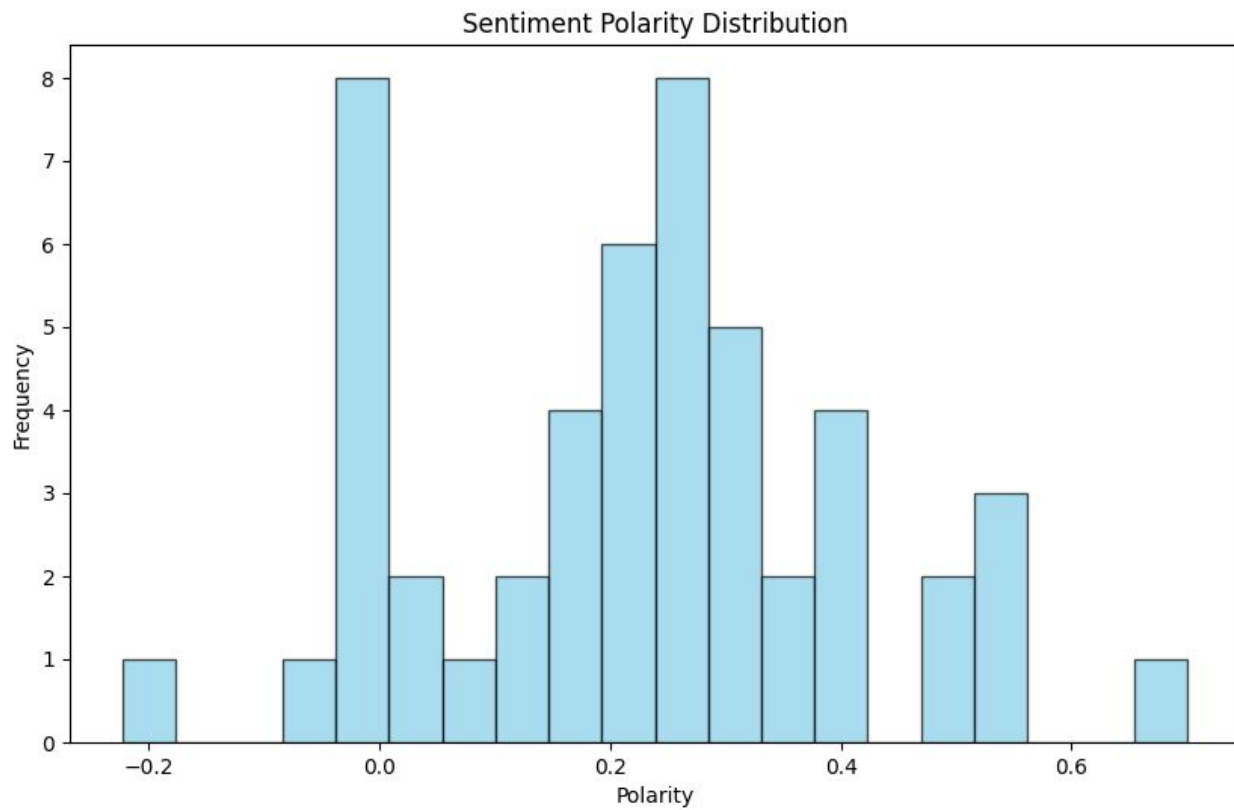
Highlighting important trends in the review
data

Rating Distribution of Reviews



Number of Reviews Over Time





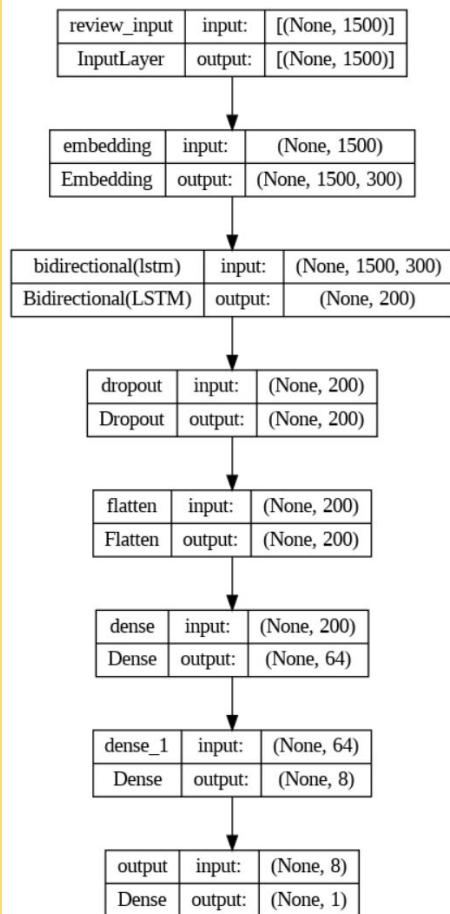


05

Next Steps

Our current status and future goals for the project



Model selection



- LSTM Model
 - Long short term memory network that uses artificial neural networks (ANN) and deep learning
 - Utilized tensorflow & keras
 - GloVe Vector
- Convolution Neural Networks (CNN)
- Transformers



Things We'd Change + Next Steps

- Enhance data quality + quantity
 - Improve model performance
 - Break down reviews to analyze sentiment at a more granular level (e.g., sentiment about product quality, customer service)
 - Create a website for users to attain real-time sentiment analysis, view visualizations and summaries of reviews
- 
- 



Thank you!
Any Questions