TINDER: LOVE IT OR HATE IT

Eduardo Ortiz, Marci Morrell, and Leti Rinaldini

Presentation Overview

- The Data
- Project Goal
- General Trends
- Analyzation
- Model
- Solutions

The Data

Kaggle Link

- Data from Kaggle
- 24,000 rows of data
- Includes reviews, scores, dates, reply content, username and thumbs up count
- Spans from 2013-2022

Project Goal

Which Rating?

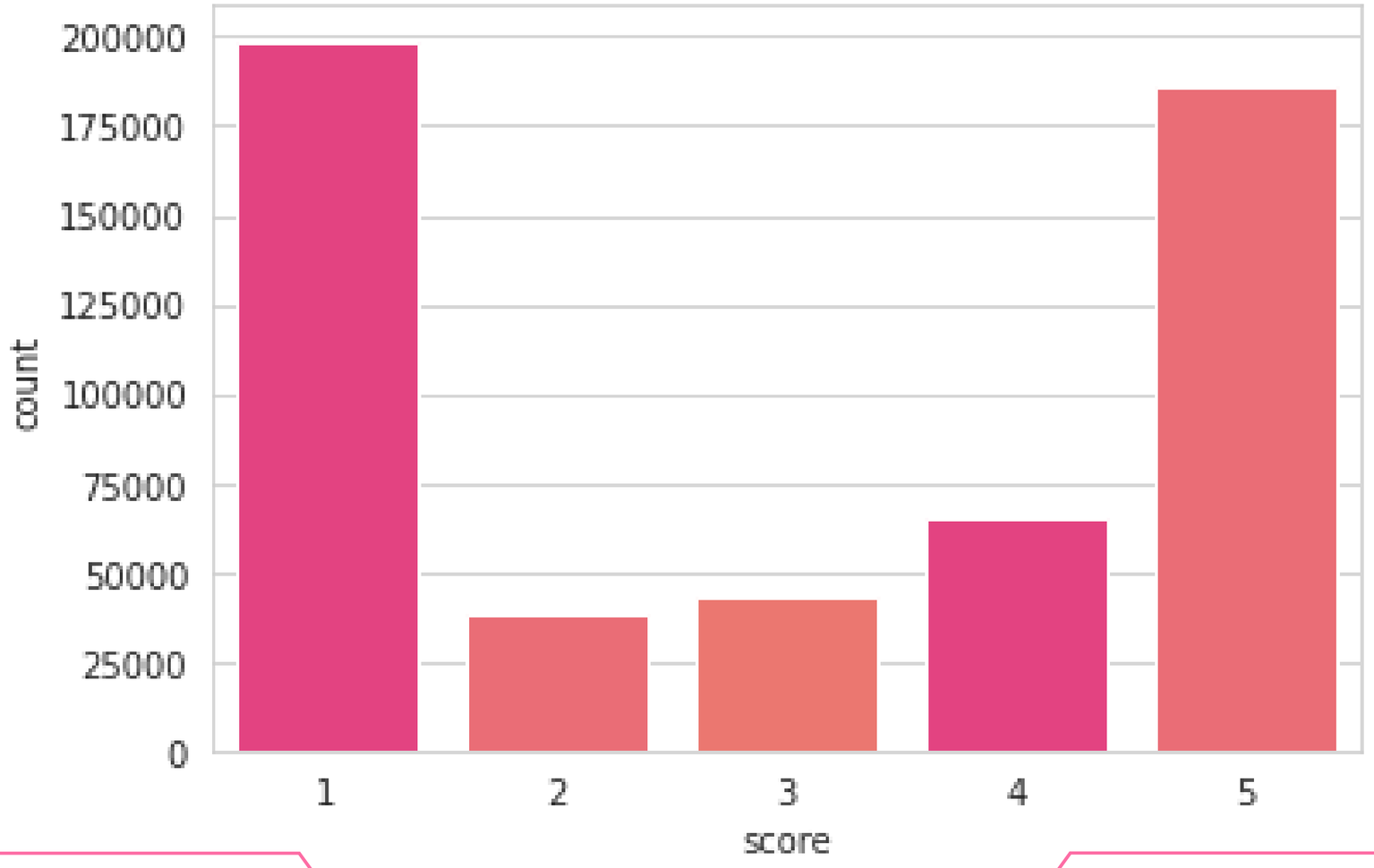
Our goal was to find trends in the data using Neural Language Processing techniques. From there we wanted to create a model that would most accurately predict the users rating based on their text review of the Tinder App.

General Trends

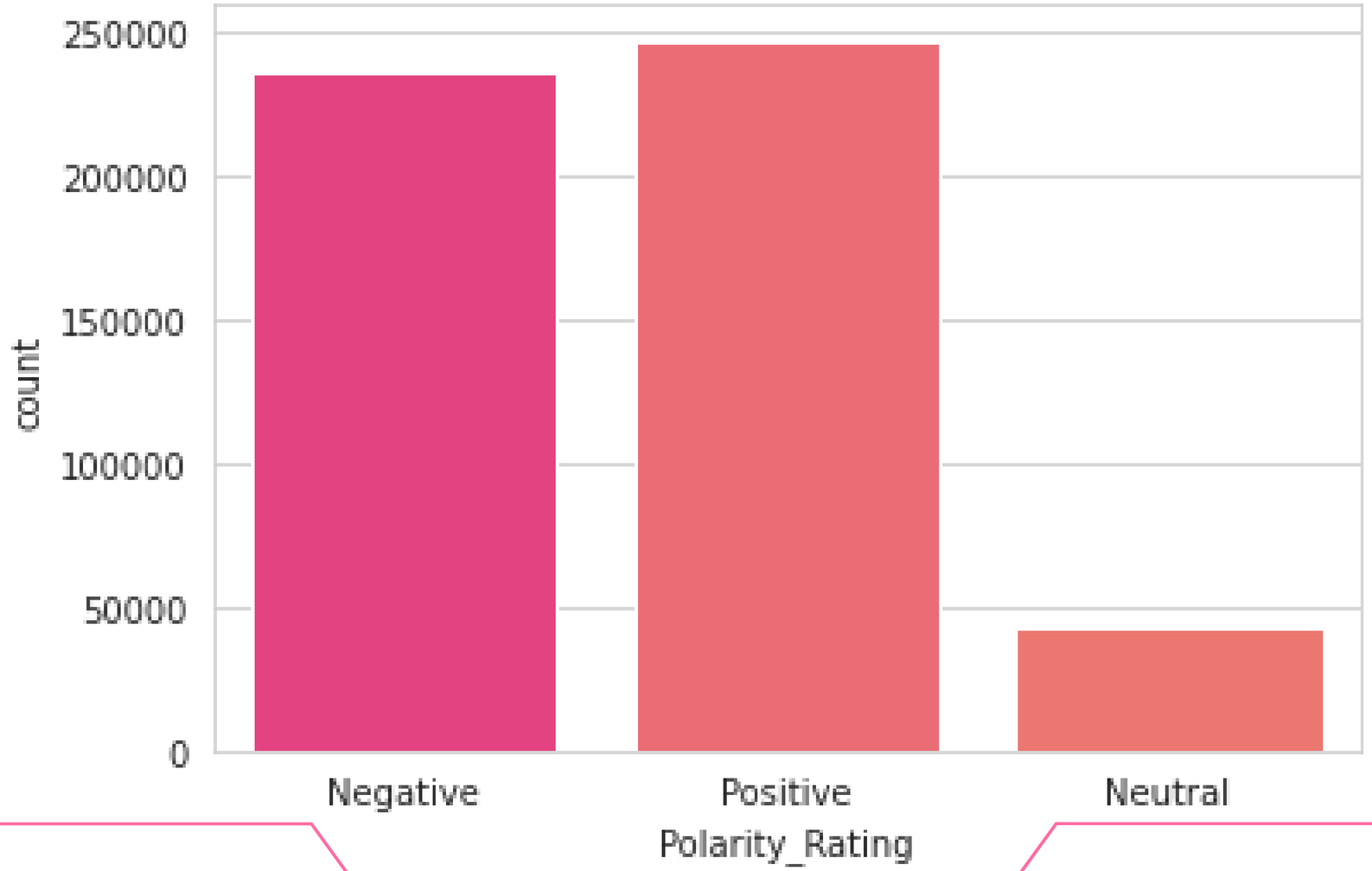
- Most scores were either 1's or 5's (Mostly positive and negative not so many in the middle.)
- A rating of 1 was the most common rating
- Positive words were more common than negative words
- Positive Polarity was the most common as well
- Negative reviews spike in March, and also from 2019-2022


alright'
app'
Good
Eh'
Nice'
Meh'
ok'
Great'
Gud'
Average'
bad'
Cool'
Nothing
okay'
aight'
Fun'
Nice
good'
Awesome'
Perfect'
Great'
app'
Easy
use'
good'
Love
Cool'
Wonderful'
Good
bad'
Excellent'
far
Super
Ok'
Nice'
it'
Fun'
Best'
Great
Nice
Amazing'
Awesome'
Best
Waste
Worst'
time
good'
Cant
Meh'
app
Bad'
Sucks
app'
log
in'
work'
Doesnt
Fake'
ever'
Worst
Horrible'
Terrible'
Useless'

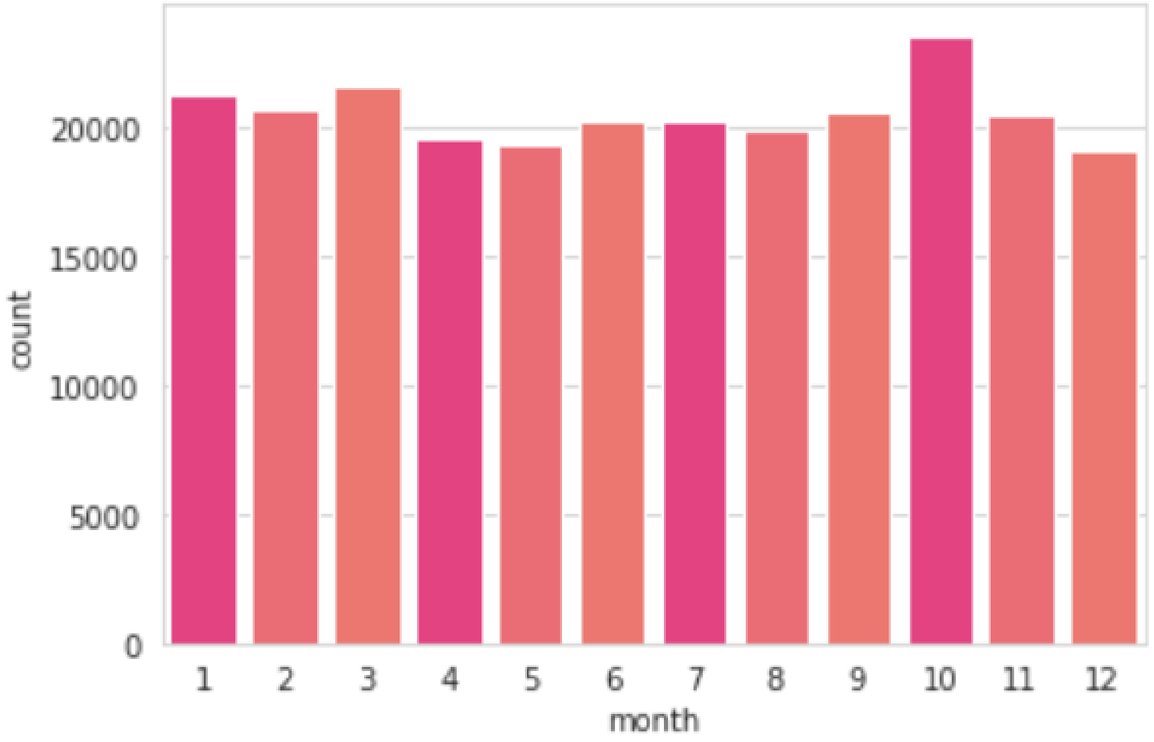
Graph of Score Count

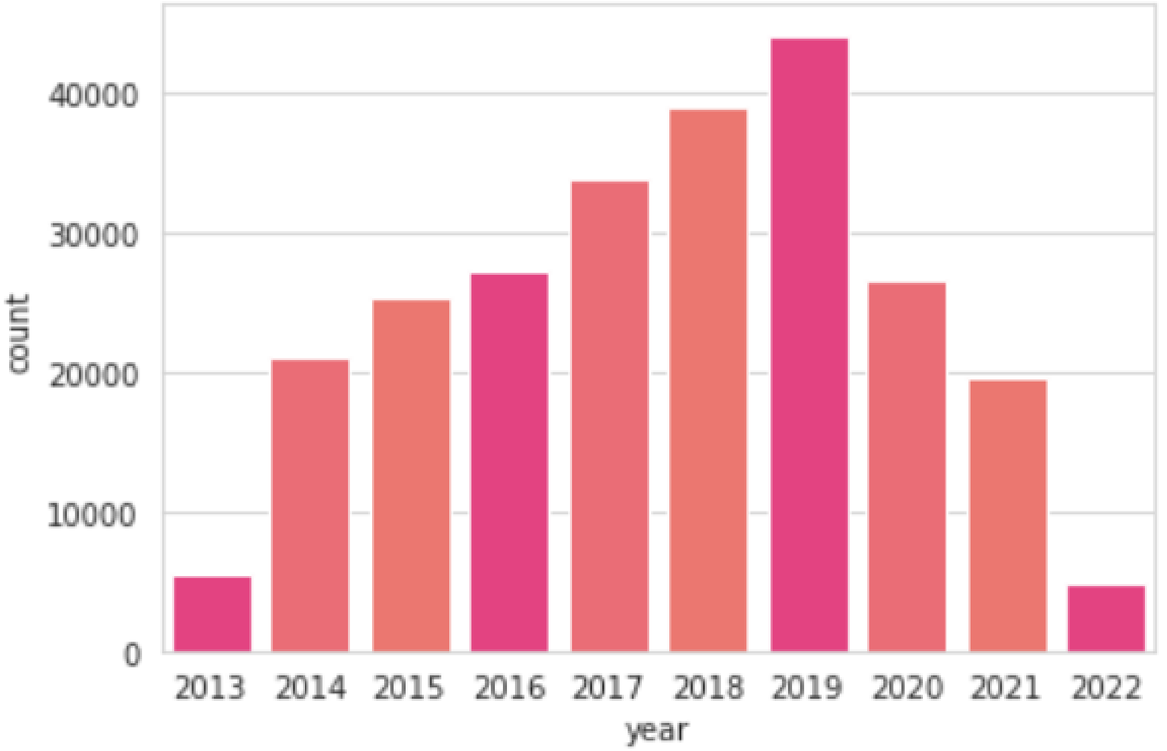


Polarity Rating Count

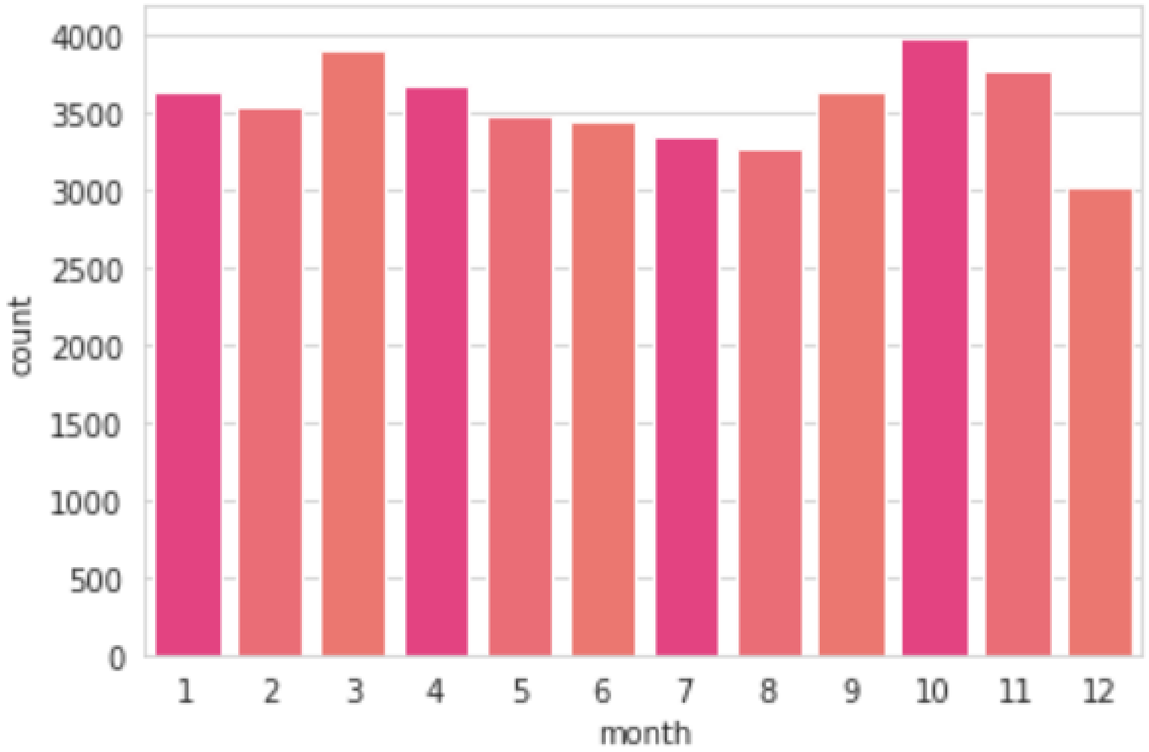


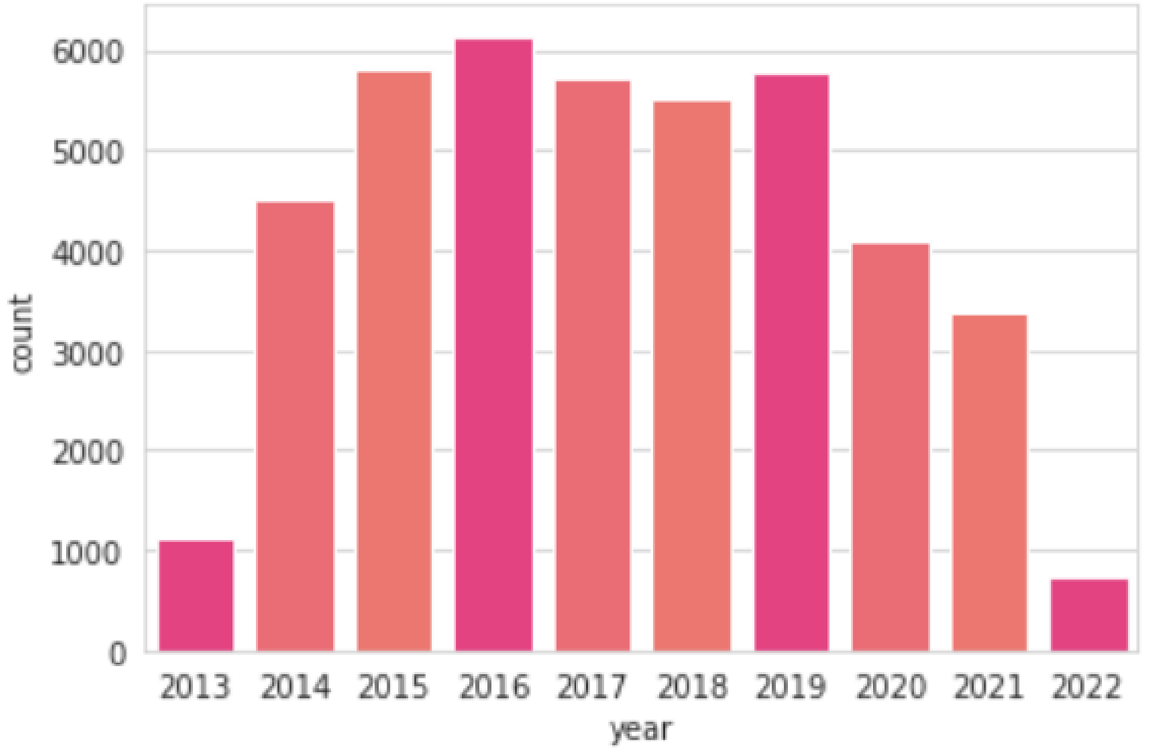
Positive


count
20000
15000
10000
5000
0
1
2
3
4
5
6
7
8
9
10
11
12
month

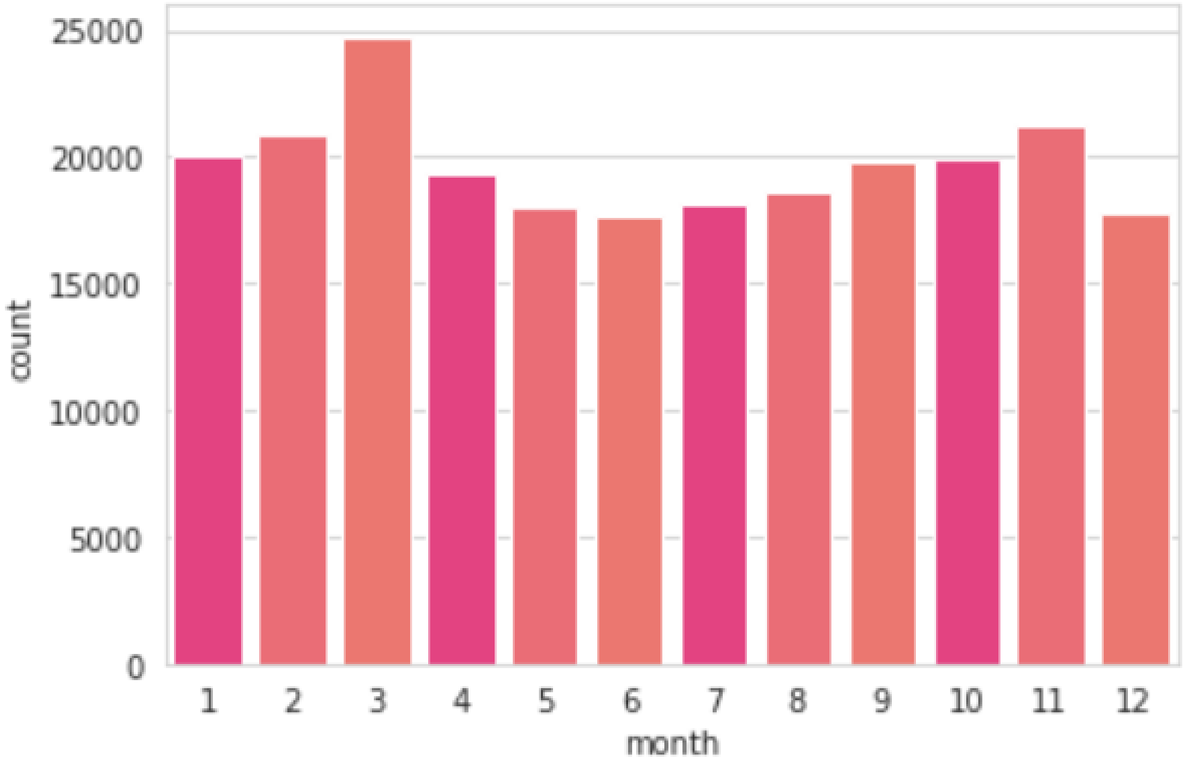

count
40000
30000
20000
10000
0
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
year

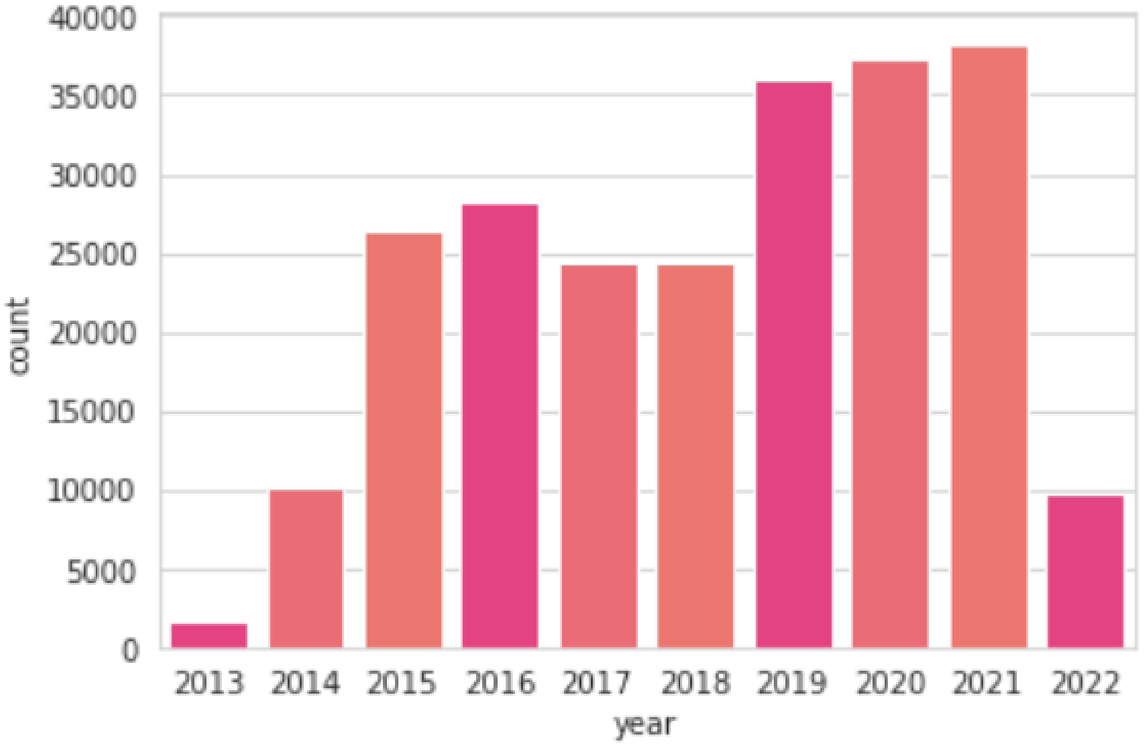
Neutral


4000
3500
3000
2500
2000
1500
1000
500
0
count
1
2
3
4
5
6
7
8
9
10
11
12
month


6000
5000
4000
3000
2000
1000
0
count
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
year

Negative


25000
20000
15000
10000
5000
0
count
1
2
3
4
5
6
7
8
9
10
11
12
month


40000
35000
30000
25000
20000
15000
10000
5000
0
count
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
year

Model Types

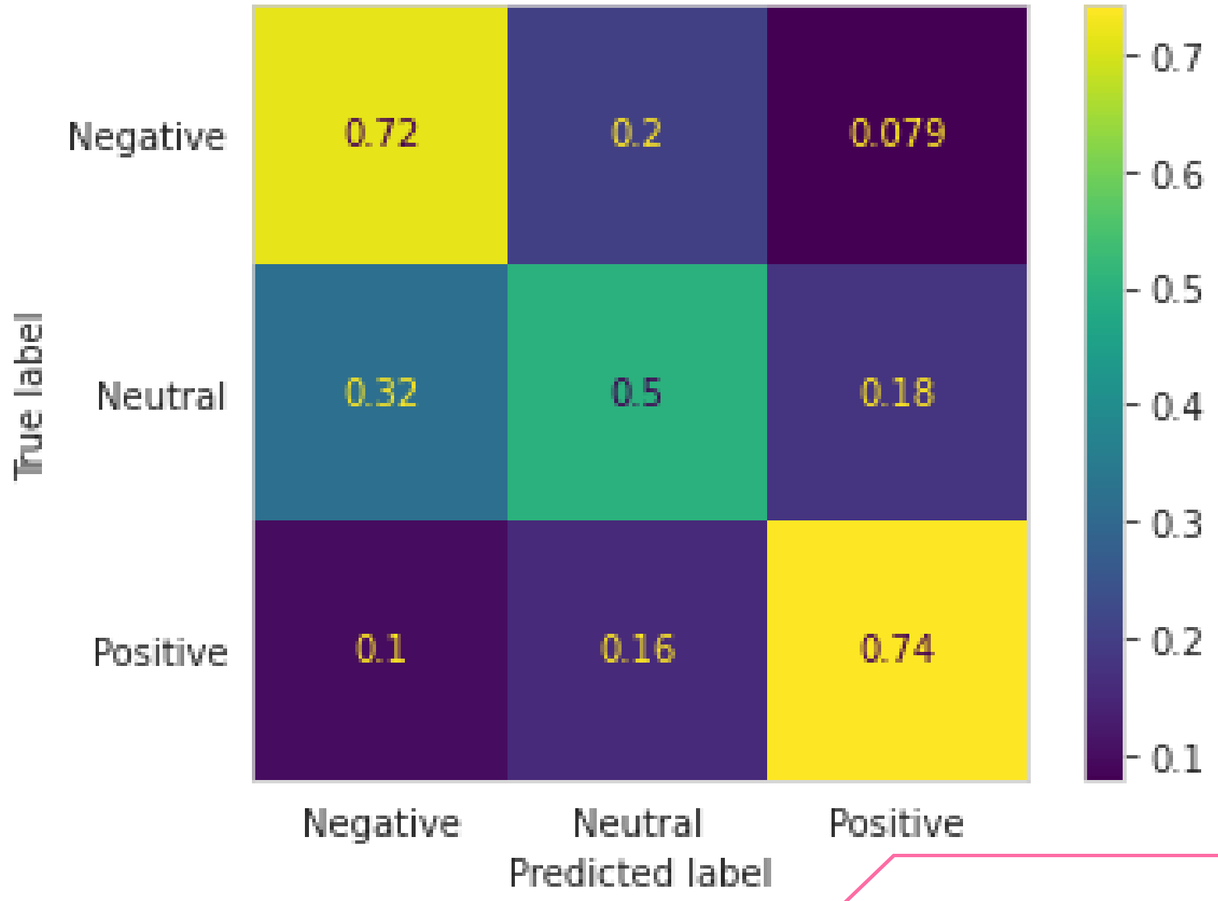
- Bag of Words
- Random Forest
- TFIDF
- Logistic Regression

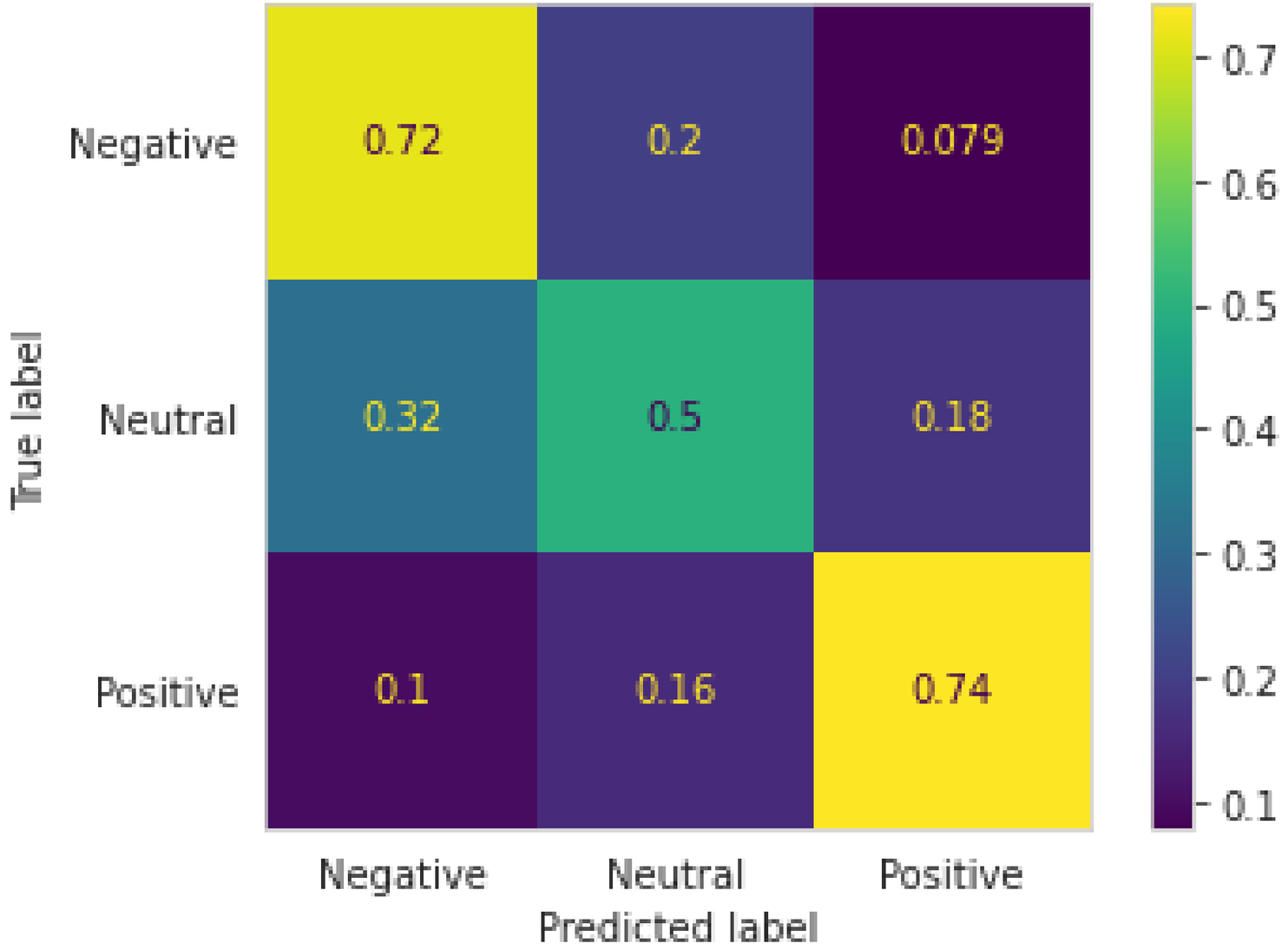
Bag of Words

- Remove stop words
- Remove emojis
- Make punctuation consistent
- Count vectorizer
- PCA

Bag of Words

- **Model predicts polarity**
- **Accuracy of 62%**



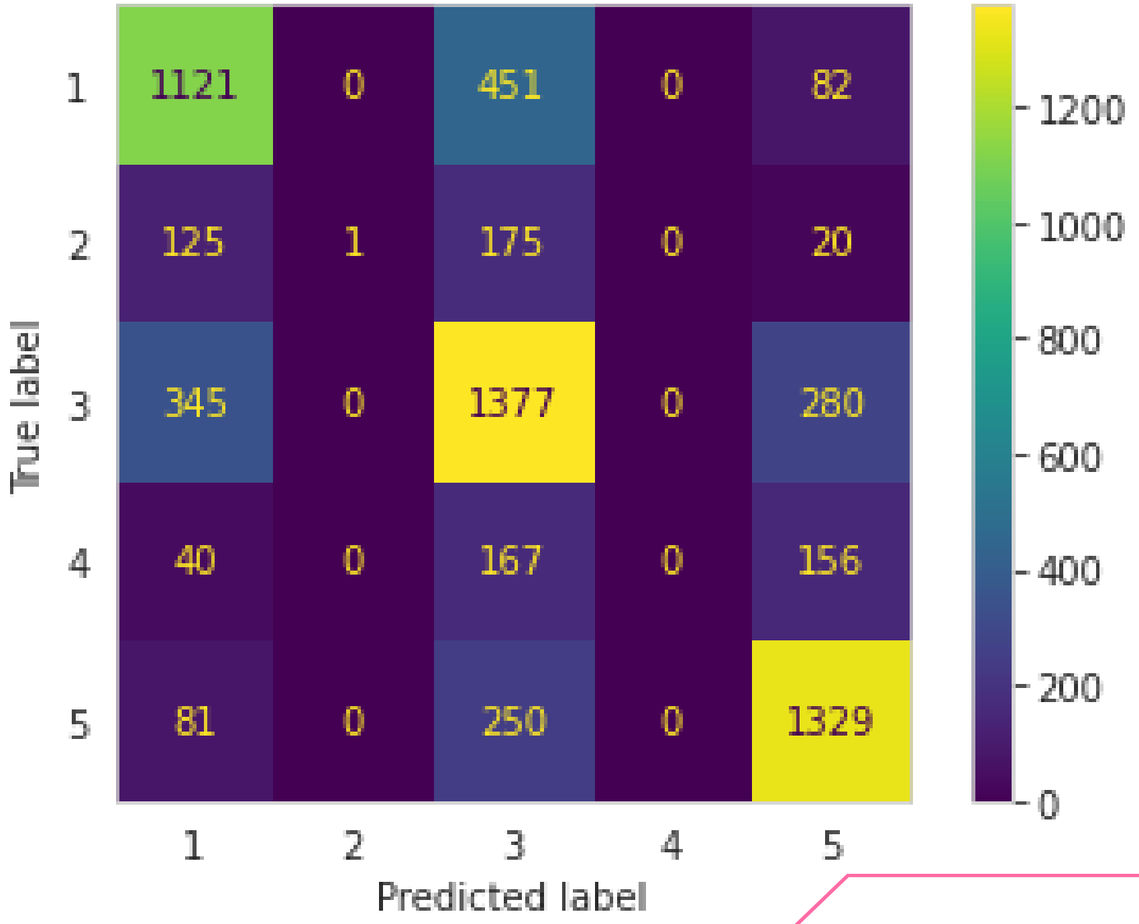

Negative
Neutral
Positive
True label
0.72
0.2
0.079
0.32
0.5
0.18
0.1
0.16
0.74
Negative
Neutral
Positive
Predicted label
0.7
0.6
0.5
0.4
0.3
0.2
0.1

TF-IDF

- Remove stop words
- Remove emojis
- Make punctuation consistent
- TFIDF vectorizer

Logistic Regression

- **Model predicts Score**
- **Accuracy of about 62%**
- **Does not predict 2 or 4**



Conclusions

- Tinder had the most user reviews between 2019-2022
- Overall, their reviews were more positive than negative so they're on the right track.
- When creating a model to predict user satisfaction, it's much easier to predict the extremes than it is to predict the "in-between" reviews

Thank you

We hope you enjoyed our presentation.

Questions?