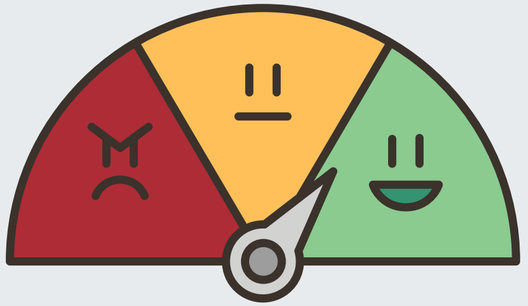
Amazon Reviews Analysis with NLP

Group 2
Wambui Githinji, Lynette Mwiti, Felix Njoroge, Wilfred Lekishorumongi, Monica Mwangi, Joan Maina

CONTENTS

PROBLEM STATEMENT



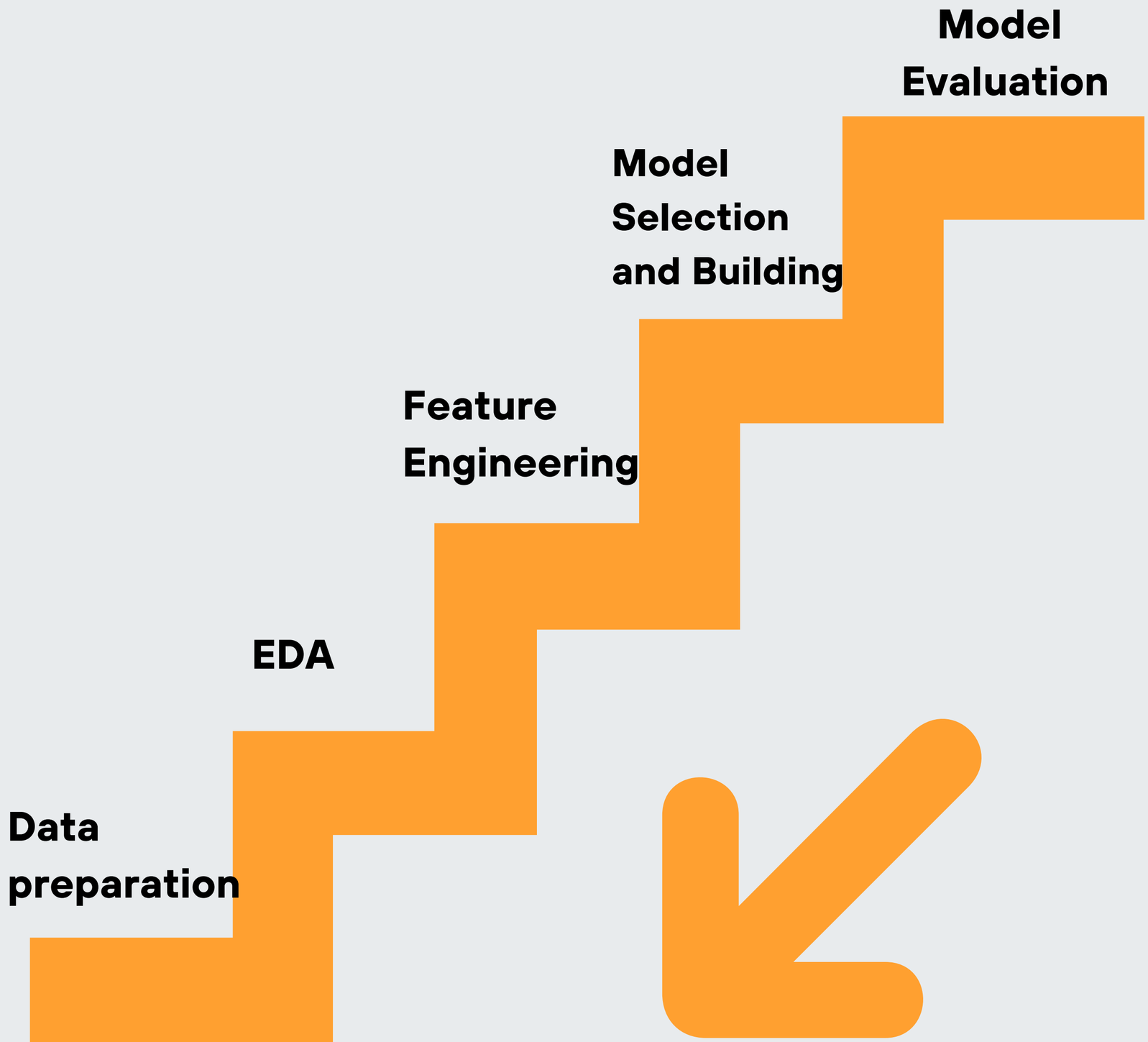
- Reviews are critical to businesses as they offer insights into customer satisfaction, preferences and areas of improvement.

- Businesses need to understand and interpret these reviews in order to cut through the competition. Lots of reviews are generated daily and manually analyzing them is impractical.

OBJECTIVES

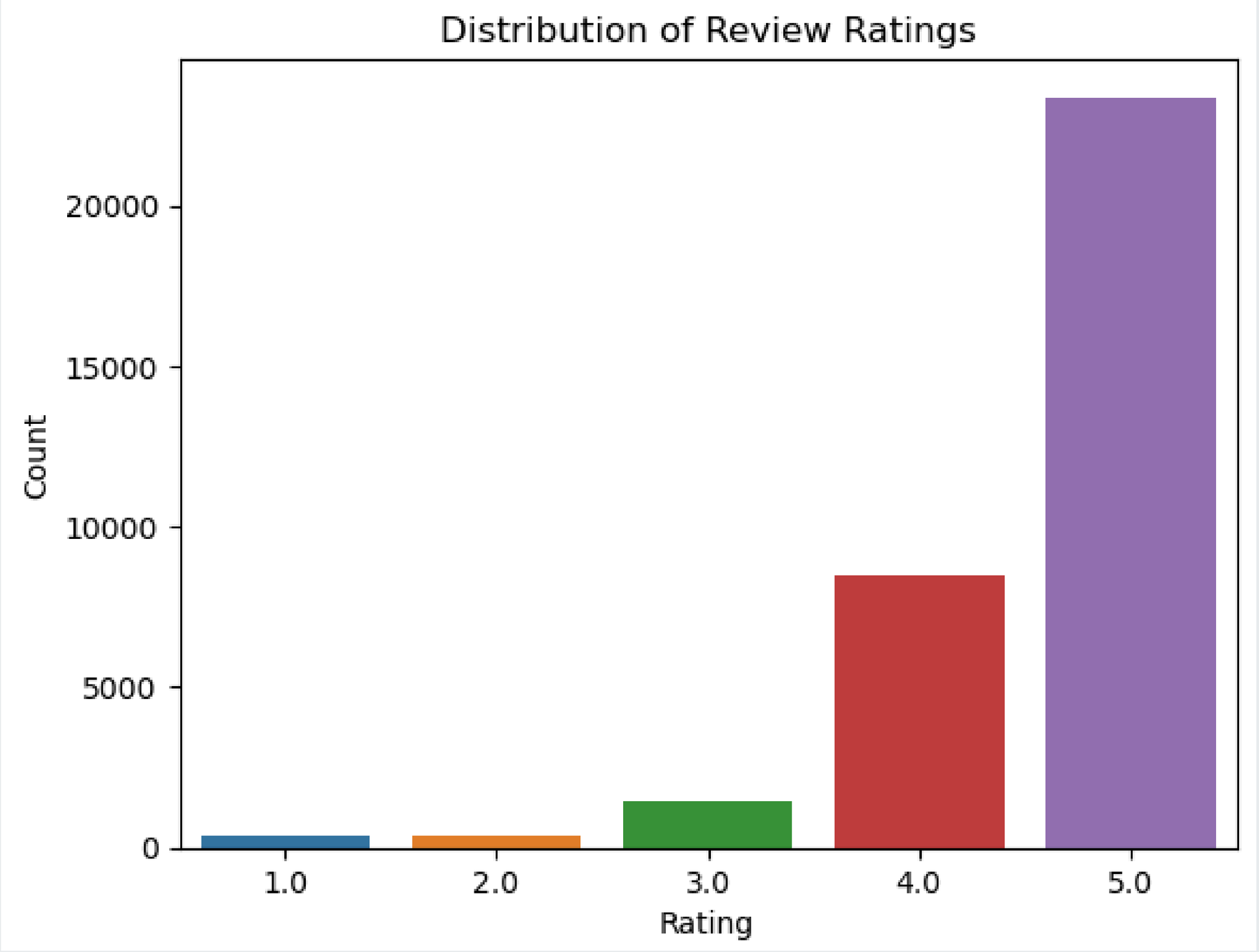
- **Broad Objective**: Use Sentiment analysis to help the businesses get actionable insights from the feedback received from customers.
- **Specific objectives**:
  - Conduct exploratory data analysis to understand the distribution of reviews over time, across brands and products.
  - Determine the sentiment of the reviews (positive or negative) to understand overall customer satisfaction and feedback.
  - Utilize sentiment analysis to help our stakeholders understand customer preferences across various products.
  - Leverage customer reviews to identify areas for improvement in products based on user experience.
  - Build a classifier model to help predict reviews as positive or negative

STEPS

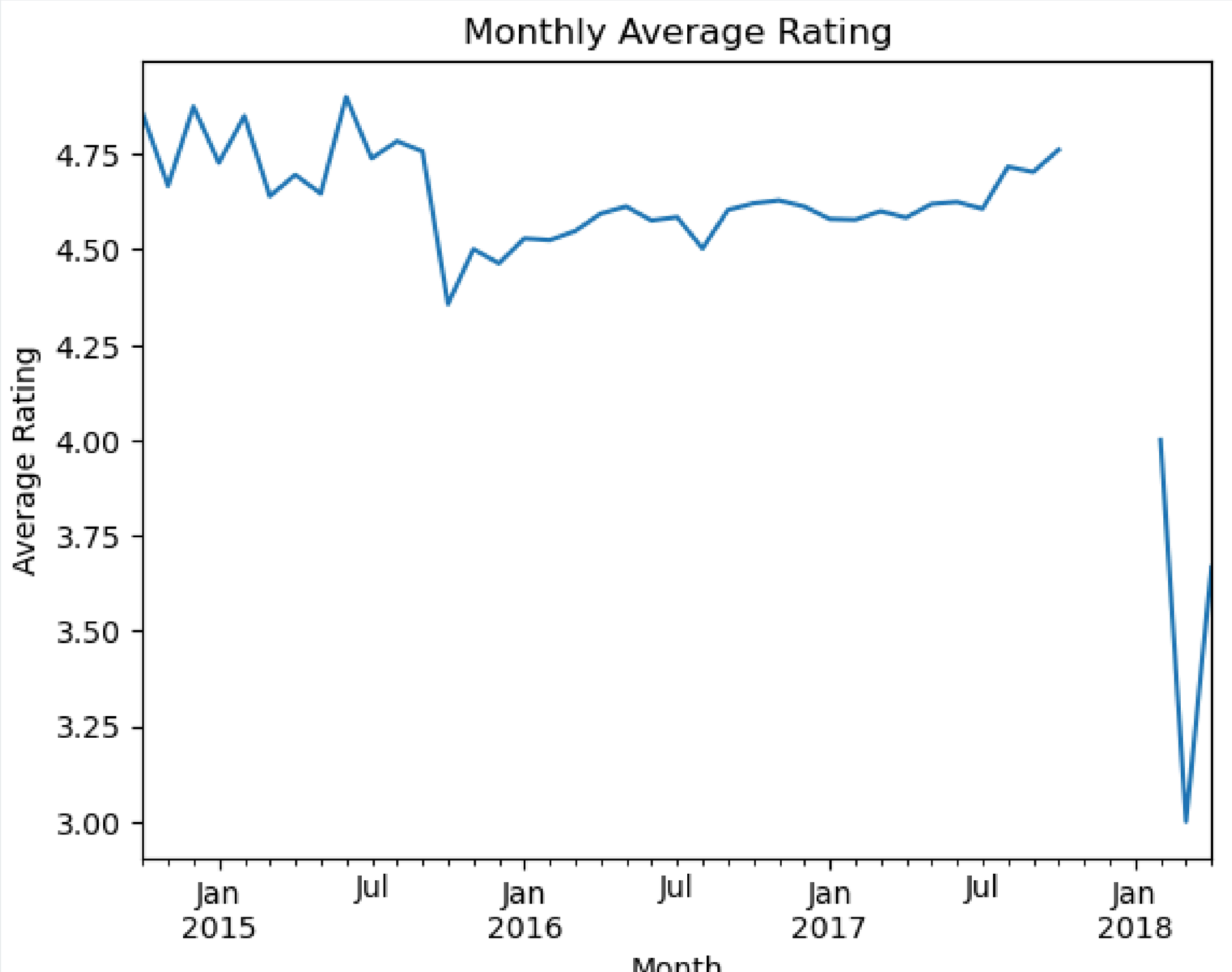

Model
Evaluation
Model
Selection
and Building
Feature
Engineering
EDA
Data
preparation

EDA

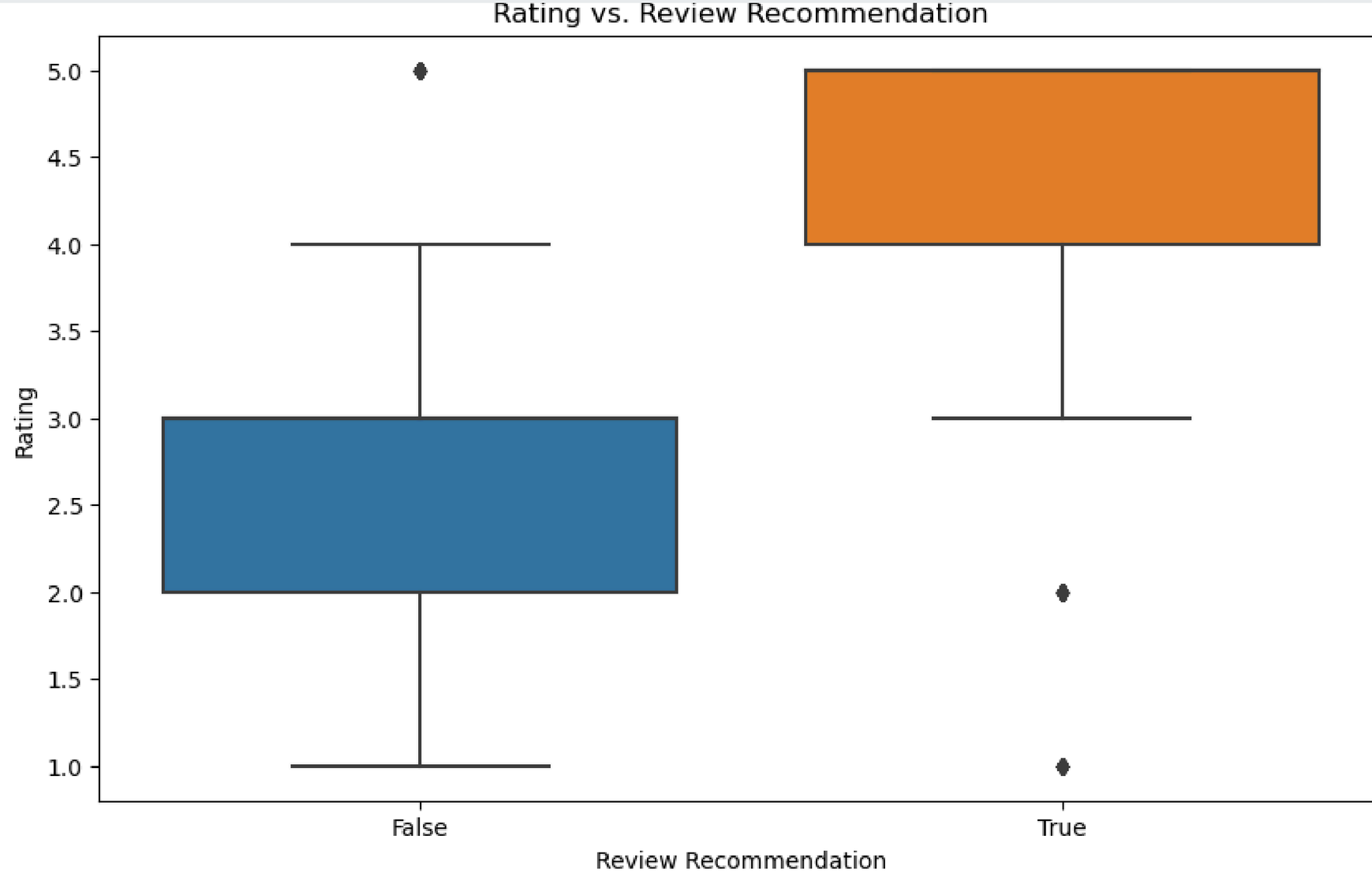
Reviews with higher ratings tend to have a higher count.



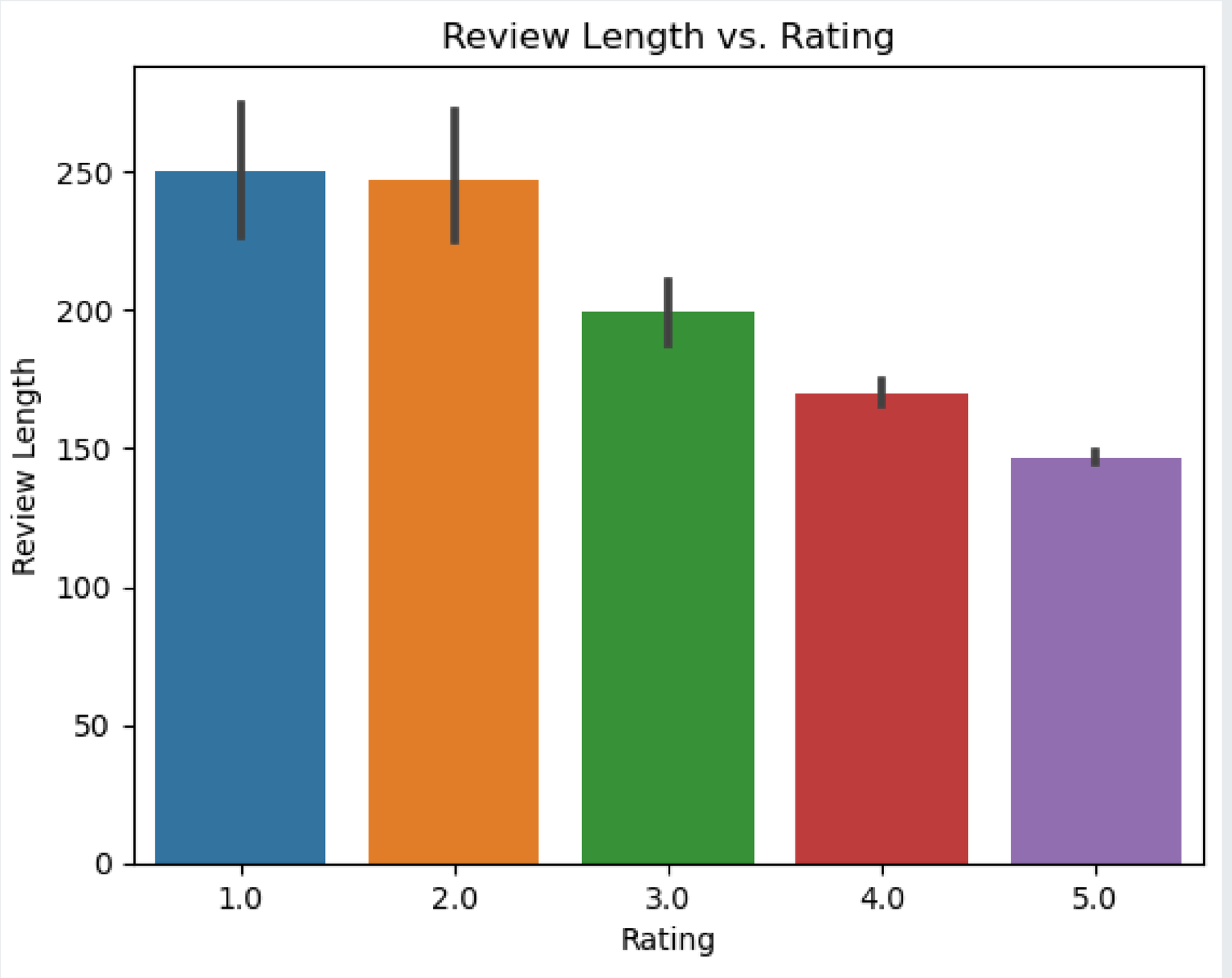
There is a slight fluctuation in average ratings over time, but no clear trend is evident from the monthly average ratings plot.



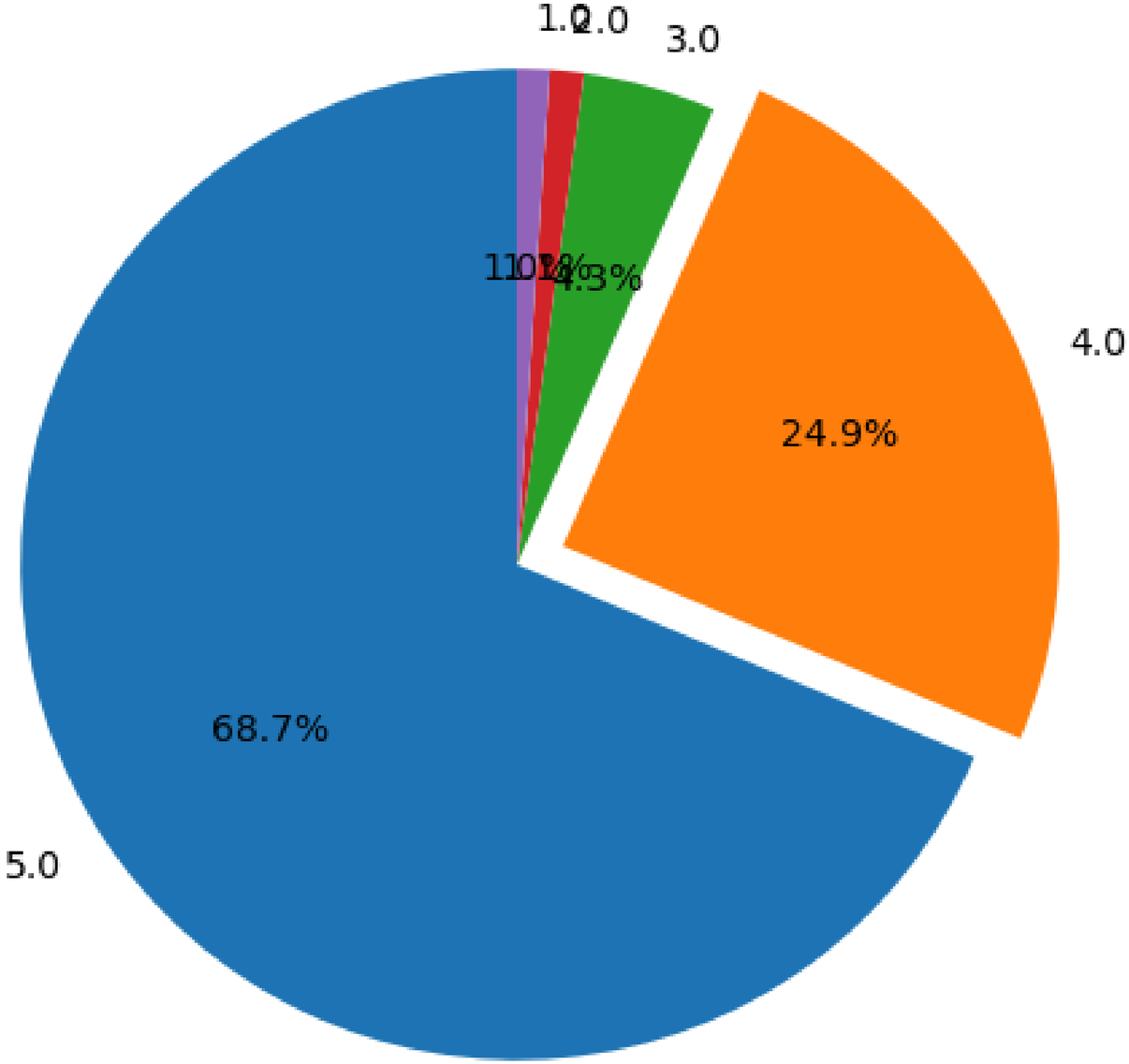
The analysis shows that reviews with a positive recommendation (review_do_recommend = True) generally have higher ratings compared to those without a recommendation.



It is evident that shorter reviews tend to receive higher ratings.

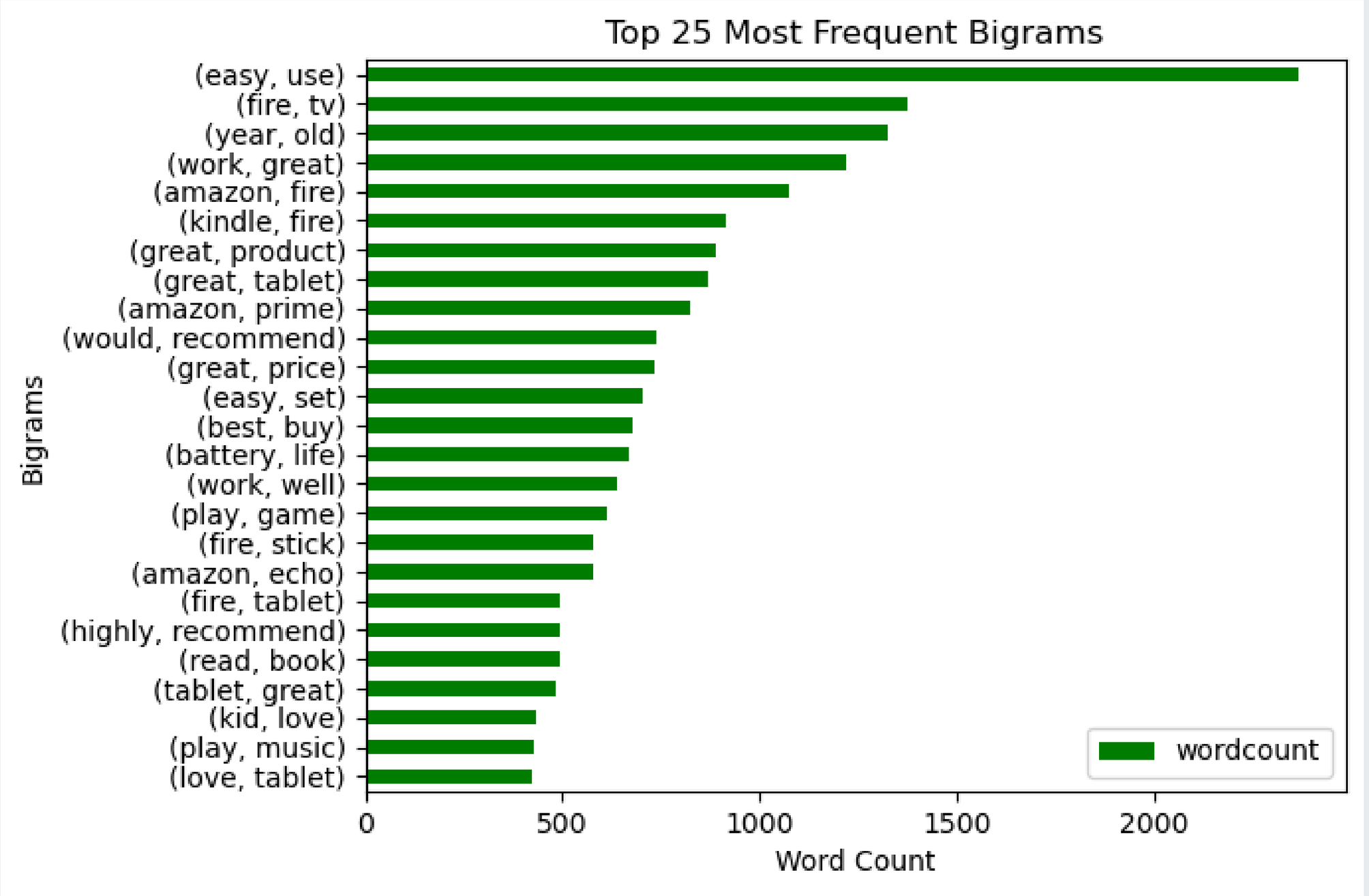


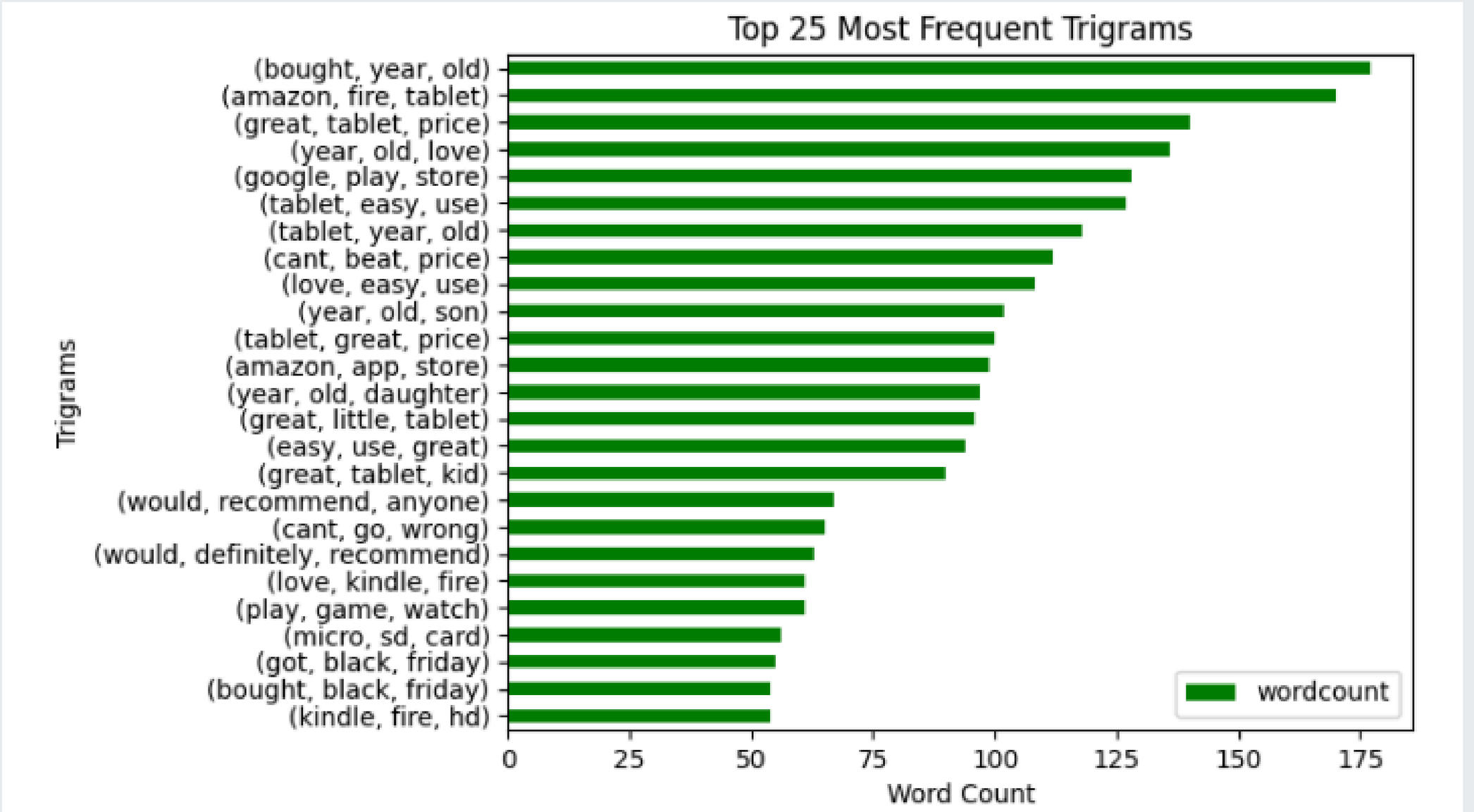
The analysis shows that reviews with a positive recommendation (review_do_recommend = True) generally have higher ratings compared to those without a recommendation.



Sentiment Analysis

- Sentiment analysis was conducted to classify reviews as positive, negative, or neutral using NLP techniques


Top 25 Most Frequent Bigrams
(easy, use)
(fire, tv)
(year, old)
(work, great)
(amazon, fire)
(kindle, fire)
(great, product)
(great, tablet)
(amazon, prime)
(would, recommend)
(great, price)
(easy, set)
(best, buy)
(battery, life)
(work, well)
(play, game)
(fire, stick)
(amazon, echo)
(fire, tablet)
(highly, recommend)
(read, book)
(tablet, great)
(kid, love)
(play, music)
(love, tablet)
Bigrams
wordcount
0
500
1000
1500
2000
Word Count


Top 25 Most Frequent Trigrams
(bought, year, old)
(amazon, fire, tablet)
(great, tablet, price)
(year, old, love)
(google, play, store)
(tablet, easy, use)
(tablet, year, old)
(cant, beat, price)
(love, easy, use)
(year, old, son)
(tablet, great, price)
(amazon, app, store)
(year, old, daughter)
(great, little, tablet)
(easy, use, great)
(great, tablet, kid)
(would, recommend, anyone)
(cant, go, wrong)
(would, definitely, recommend)
(love, kindle, fire)
(play, game, watch)
(micro, sd, card)
(got, black, friday)
(bought, black, friday)
(kindle, fire, hd)
Trigrams
wordcount
0
25
50
75
100
125
150
175
Word Count

Negative Reviews


Positive Reviews
great
easy
use
tablet
love
kindle
bought
good
price
one
amazon
kid
recommend
christmas
thing
excellent
store
paperwhite
purchased
battery
awesome
alexa
loved
internet
im
long
need
much
reading
set
device
little
read
book
better
best
perfect
feature
well
music
year
happy
new
make
got
play
screen
really
time
son
get
light
fire
old
purchase
daughter
buy
small
first
child
google
family
money
reader
phone
memory
video
ipad
still
wife
echo
wanted
everything
highly
day
quality
would
like
product
game
lot
gift
movie
access
free
looking
around
work
apps
also
could
email
right
find
way
two
friendly
web
enough
using
size
nice
easier
issue
worth
doesnt
deal
back
absolutely
problem
weight
wish
etc
easily
turn
friday
carry
charge
eye
parental
pretty
fast
option
didnt
night
storage
fit
download
android
user
resolution
amazing
getting
us
ive
grandson
definitely
available
wifi
control
item
without
since
sound
color
playing
value
older
camera
enjoy
cant
version
black
every
watching
case
many
prime
hold
able
keep
overall
page
picture
watch
fun
hd
ereader
anyone
sale
look
think
dont
go
app
say
last
far
grandaughter
eyes
simple
hand
mother
always
another
put
especially
home
life
used
card
friend
even
learning

MODELING

Two deep learning models are used;

- a Simple RNN
- an LSTM,

They are built to predict the sentiment of the reviews.

Model Performance

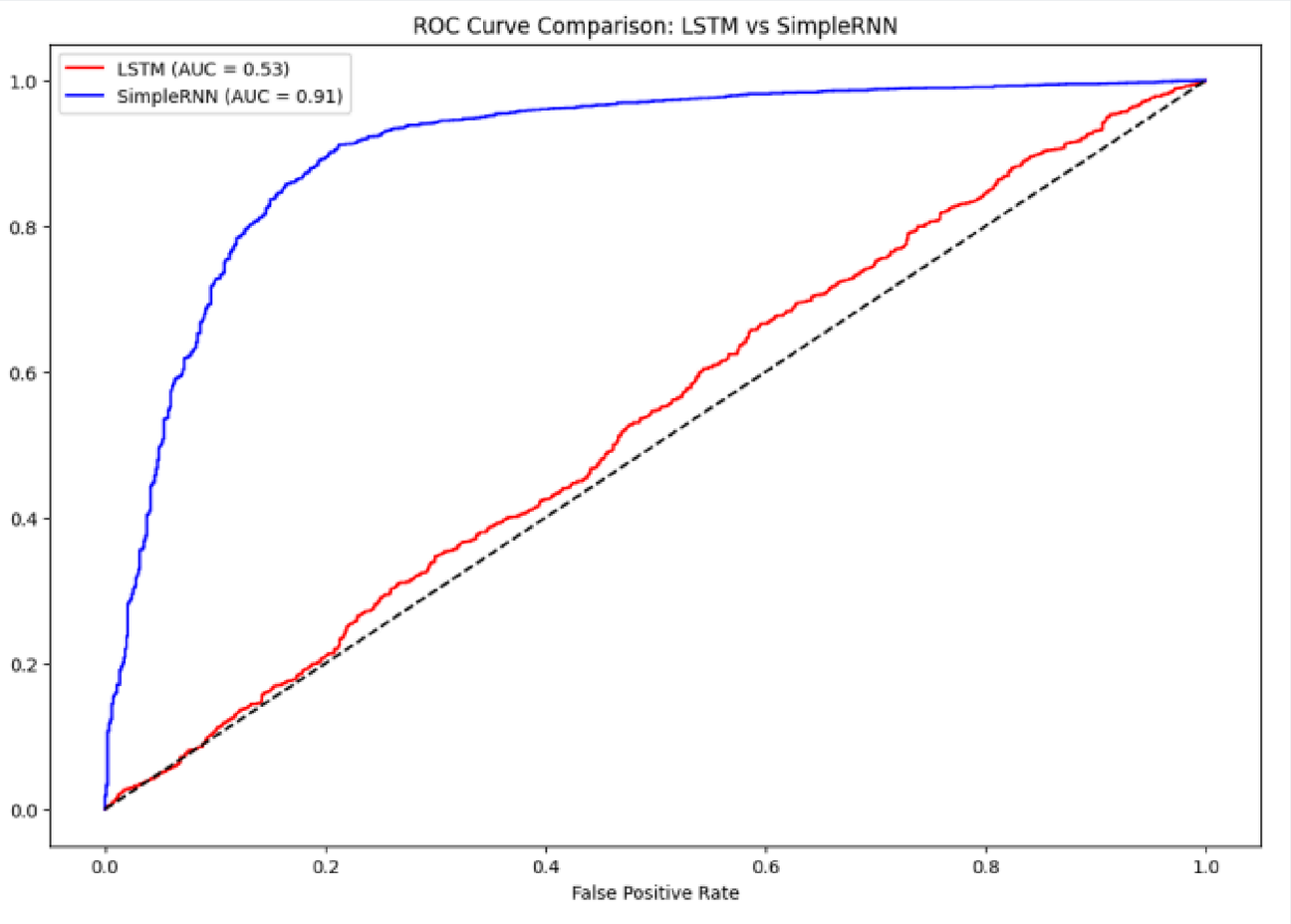
These are the model performance results;

- SimpleRNN Model: With an AUC of 0.91, the SimpleRNN model performs much better than the LSTM model in distinguishing between the classes. It has a high true positive rate and a low false positive rate across various thresholds.

- LSTM Model: With an AUC of 0.53, the LSTM model is only marginally better than random guessing. This indicates that the LSTM model is ineffective for this particular task or Amazon reviews dataset.

Model Evaluation

- Hyperparameter tuning is performed to optimize the models for better performance.
-
- The Model tuning for simple RNN Accuracy score is 0.8694 which has no huge difference from the model before tuning which was 0.8658.
-
- The score is slightly higher than our objective of achieving 0.85 accurracy score. The model is therefore satisfactory.



RECOMMENDATIONS

- **Customer Feedback:**

Maintain product quality, enhance customer service, and solicit feedback to bolster positive reviews.

- **Product Preference:**

Expand preferred product categories, improve features based on feedback, and maintain competitive pricing.

- **Trend Analysis:**

Correlate sales spikes with launches, updates, or campaigns to adapt strategies.

- **User Experience**:

Incentivize detailed reviews to enhance credibility and trust.


thank you