

Shop around the Corner

07-22-2024



WISHED CONSULTING

Prepared for: Little Book Shop Holdings

Prepared by: Crystal Grant, Lead Data Scientist

Meet the Client:



LITTLE BOOK SHOP

LBS Holdings:

Owned and operated 22 Small Book Shops in the NE from 1995 - 2008

Barnes & Nobles & Borders were the first big hits to the business - causing LBS to close 13 of their locations between 2002 - 2008. The remaining stores closed by the end of 2013 with the rise of Amazon Books + Kindle.

Kathleen Kelly of LBS is hoping to re-open one location this year and would like Wishes Consulting to assist with all aspects of the reopening to avoid the pit falls of the past.

LETS GET STARTED

UNDERSTANDING THE DATA



AMAZON REVIEWS

43% of Book Buyers rely on Amazon Reviews for guidance.

BOOK DATA

Information regarding 3M Book Titles including publishing data

RATINGS/SCORE

Scaled 1-5 reviewers rate their choices and provide feedback

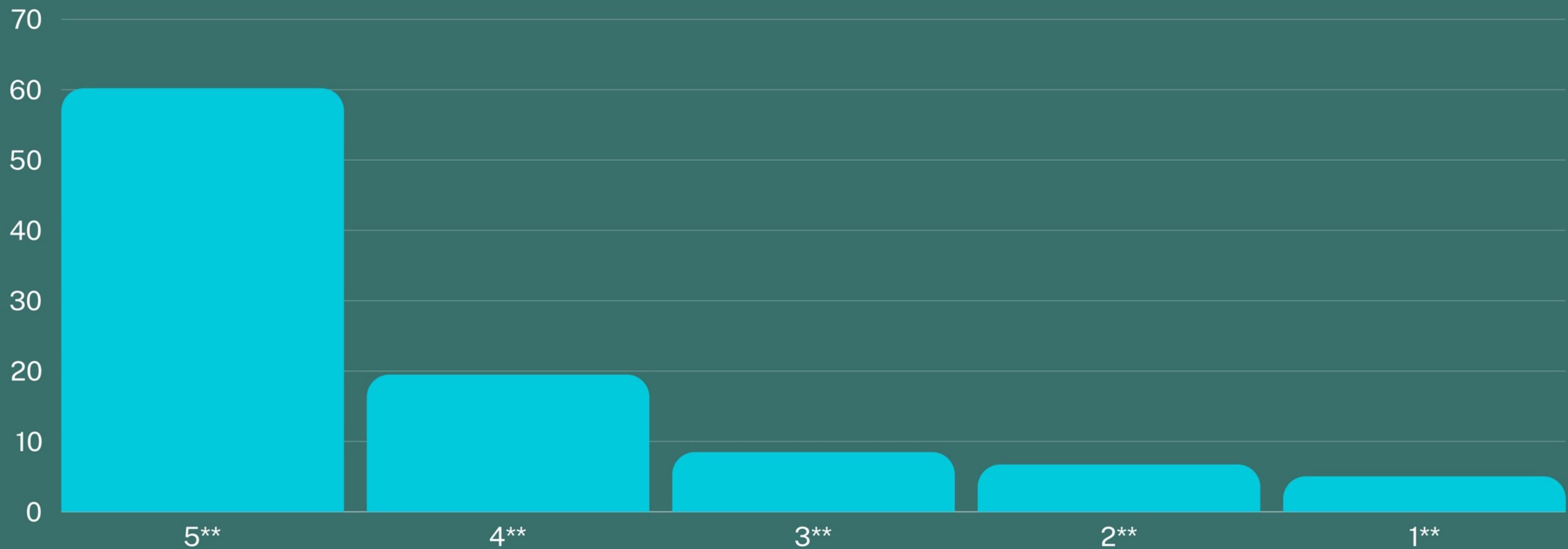
REVIEWS

Information containing over 2M+ Amazon book reviews and scores

REVIEW SUMMARY + TEXT

In addition to a rating score - users can also leave text feedback

PERCENTAGES OF SCORES



60.2%

5 STAR REVIEWS

Most Reviews are positive
- important to decide
which ones hold more
weight

18.7K

TOP REVIEWS

The most Highly rated
book was rated 18k+ times

DATA ANALYSIS APPROACH



SET SENTIMENT

Understanding/ Create our Independent Variable.



LOGISTIC REGRESSION

Classify Sentiment



NAIVE BAYES

Utilize Multinomial, Complement & SVC models to determine the best predictor.

EXPLORATORY DATA ANALYSIS

1

COMBINE DATA

Utilize Title Column to Inner Join
two data sets

3

REDUCE DATA SIZE

Started with 3M Rows - Final
Analysis on 2k**

PRE- PROCESSING

Stemming & Lemmitization

5



2

OPTIMIZE DATA

Removed unnecessary features &
prioritize “helpfulness”

4

PRE- PROCESSING

Eliminate contractions, Uppercase,
& punctuation + Stop Words

6

CREATE SENTIMENT

Use cleaned data to understand
Sentiment

** More rows would have been ideal with more computing power.

MODELS

TEXT CLASSIFICATION

BAG OF WORDS

Bag of words - 90% Accuracy



TF - IDF

Improves model - but by a fraction -- 89.5 --> 89.75



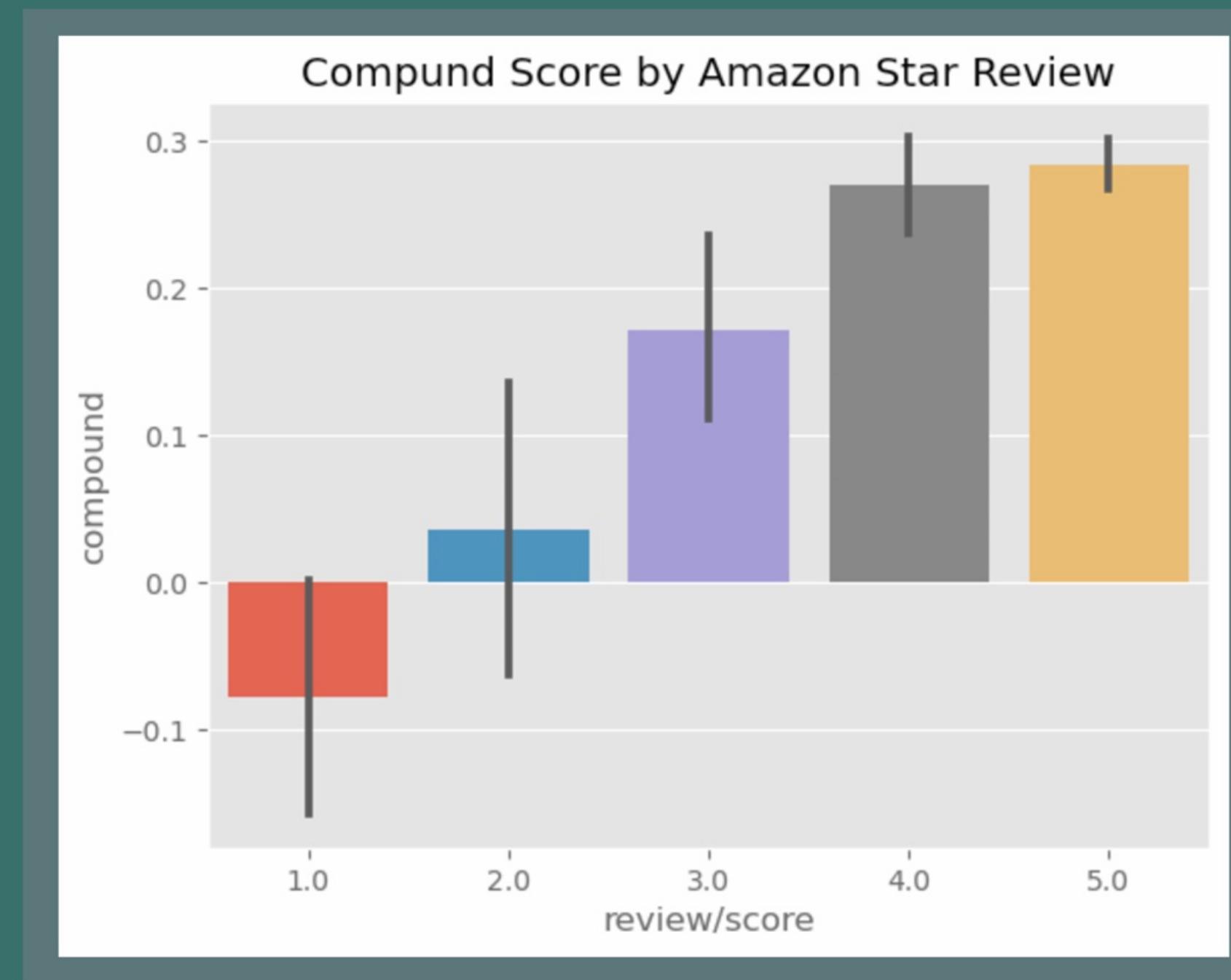
VADERS

Predicting Negative, Neutral , Positive and Combined Sentiment - Predicted as expected based on Scores

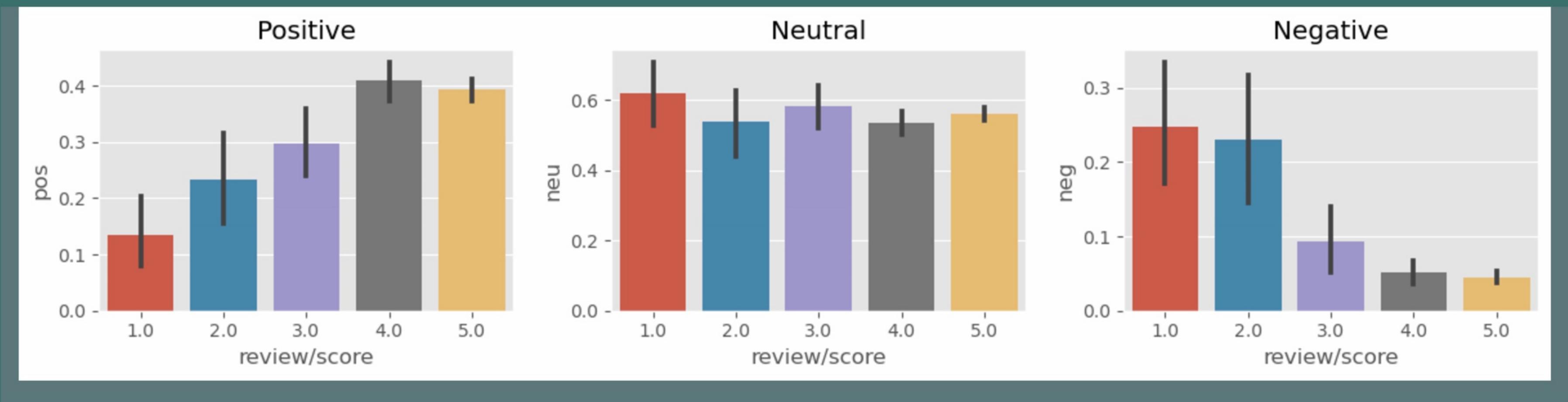
ROBERTA

Predicting values similar to Vafders - however these are more accurate

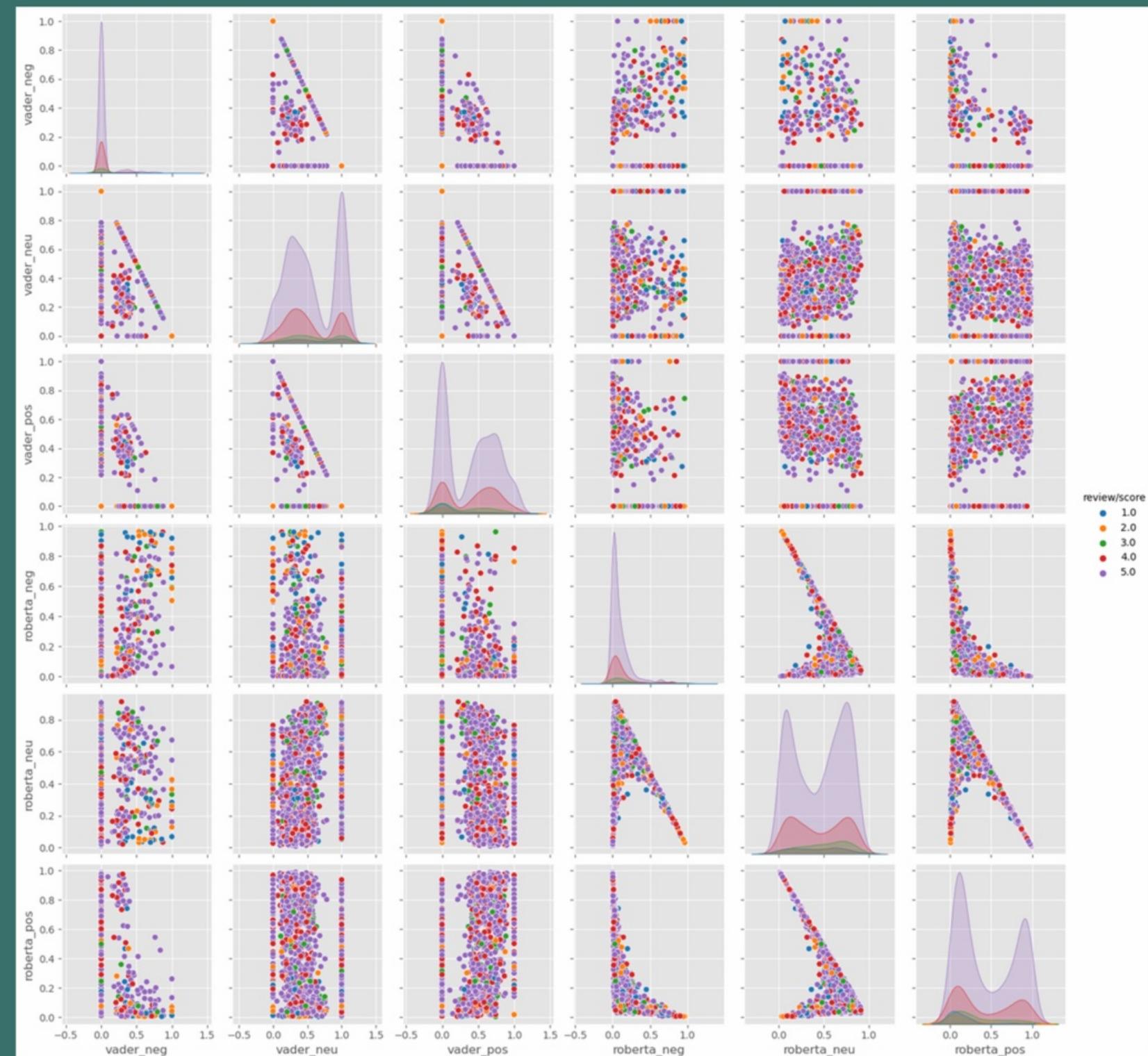
VADERS



VADERS



COMPARING VADERS VS. ROBERTA



INTERESTING FINDINGS

REVIEW SCORE AS PREDICTOR

With Each review - There is an out of 5 Star review -- We would expect to find High Positive scores from the model along with High Ratings



REALITY

Nuance is difficult to balance when it comes to textual data - While Roberta Model does better - its is still not perfect.



RATING



NEGATIVE

'heartbreaking'



RATING



POSITIVE

'raising teenager 52 brilliant idea'

TRADITIONAL MODELS

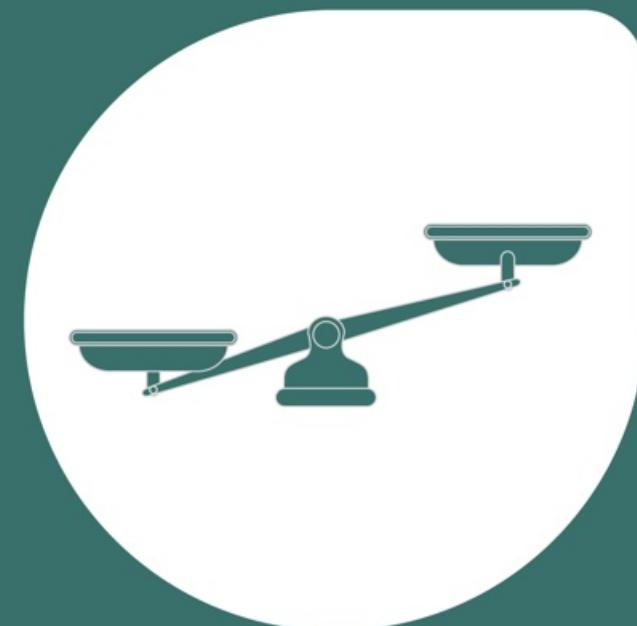
LOGISTIC REGRESSION

88.25% Accurate



NAIVE BAYES

50% Accuracy --- Something Must Be Wrong
Attempted to Predict Categories -- That also did not work - 39% Accuracy



NAIVE BAYES- COMPLEMENT

Utilized Pipeline for all 3 Naive Bayes



NAIVE BAYES- SVC

MNB: 88%
CNB: 79%
SVC: 90%

DELIVERABLES

**CATALOG OF EASILY
ACCESIBLE BOOK & REVIEW -
TO SHARE WITH CUSTOMERS**

**HIGHLY RATED BOOK
SELECTION**

**NEW BOOK SELECTION
MODEL BASED ON
GENERE**





THANK YOU