



WEB APIs & NLP

PROJECT 3: GENERAL ASSEMBLY DSIR 0124



COLE POPPSINGER:

passionate
about beverages
for over
40 years





TABLE OF CONTENTS

01

DATA SCIENCE PROBLEM

What do we want to solve?

02

MODELING + ANALYSIS

How will we try?

03

RECOMMENDATIONS

What's next?



01

DATA SCIENCE PROBLEM:

Using NLP, can customer insight be gleaned from user activity on reddit?

Will the insight be useful in targeted marketing?



DEVISE A STRATEGY

Scrape user posts from r/espresso and r/CraftBeer subreddits

Train a model to identify similarities and differences

Apply findings to recommendations and next steps

A close-up photograph of a hand holding a black smartphone, positioned next to a chrome faucet. The background is blurred, showing what appears to be a kitchen or bathroom setting. The text is overlaid on the right side of the image.

reddit users:
430,000,000 monthly
52,000,000 daily

The potential impact of
insight and advertising
is significant



02

MODELING + ANALYSIS

A background image showing four men from the chest up, holding beer glasses. They are smiling and looking towards the camera. The man on the far left is wearing a dark shirt, the second is in a red shirt, the third is in a green shirt, and the fourth is in a grey shirt. They are holding four different styles of beer in tulip-shaped glasses. The background is a light-colored wooden wall.

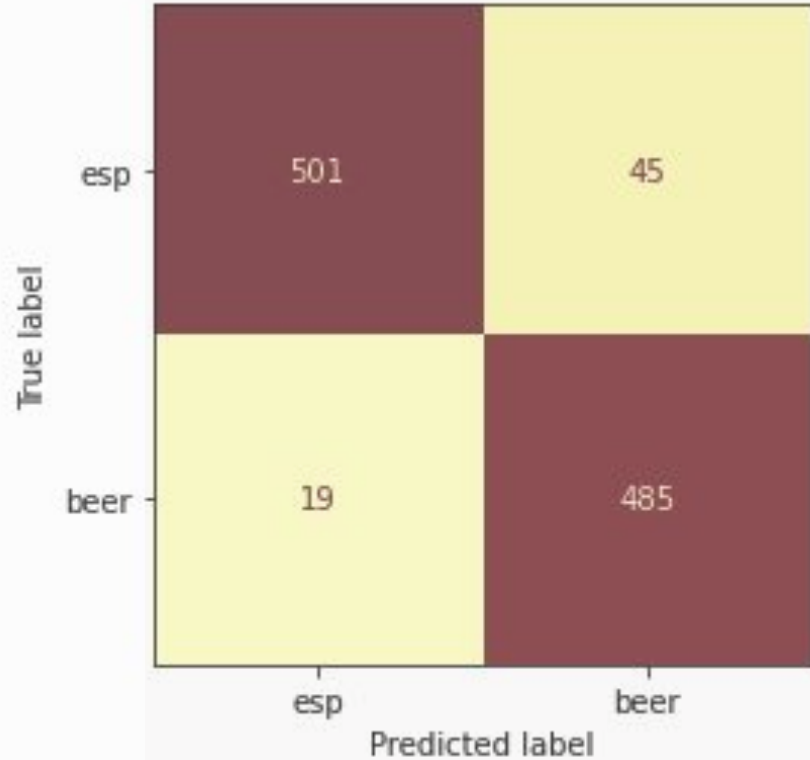
THE MODELS+NLP

- CountVectorizer + Logistic Regression Pipeline
- TfidfVectorizer + DecisionTreeClassifier Pipeline
- RandomForest
- FeatureUnion
- Lemm + Stem
- CountVectorized
- custom stop words

BEST:

- TfidfVectorizer + Logistic Regression Pipeline

CONFUSION MATRIX:

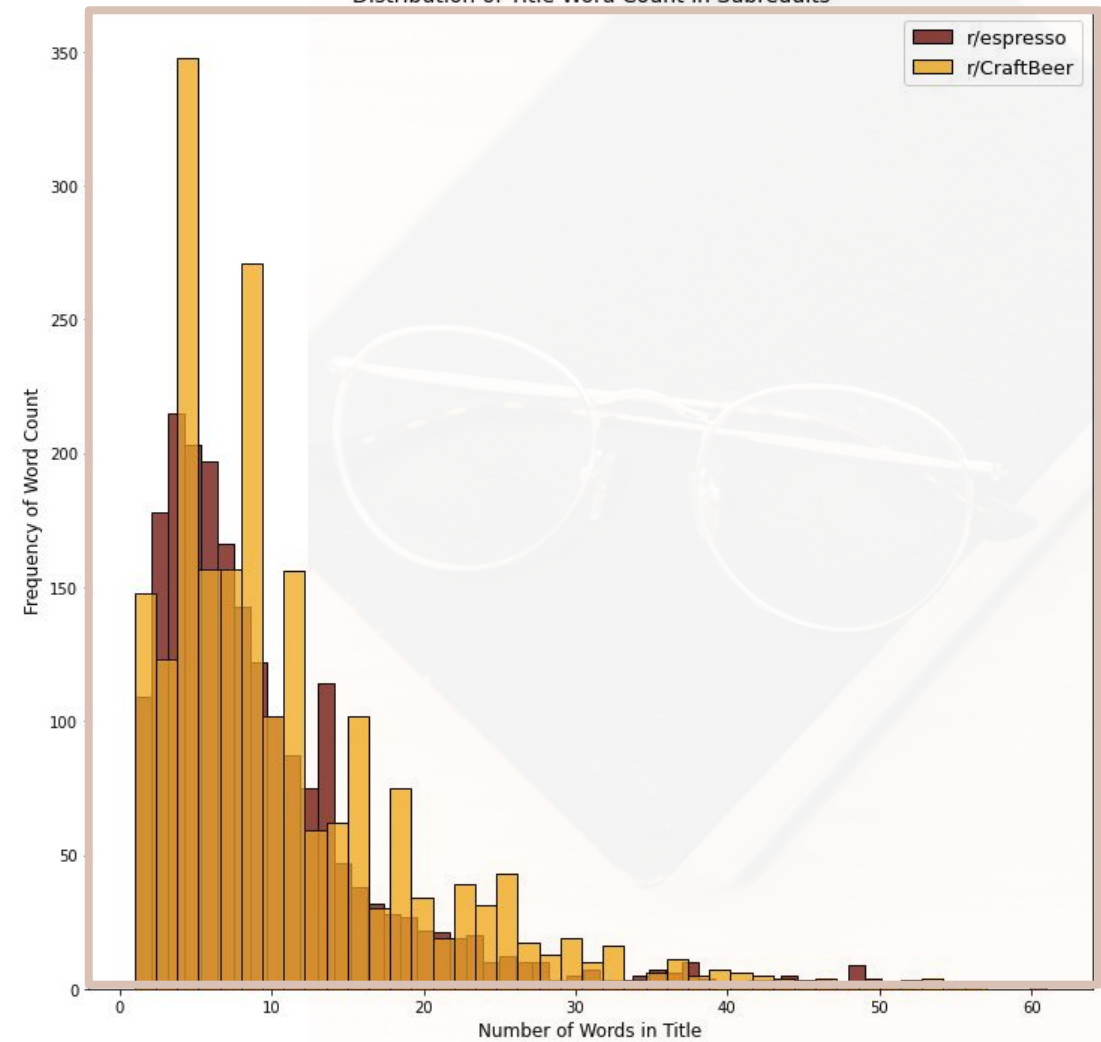


BEST MODEL

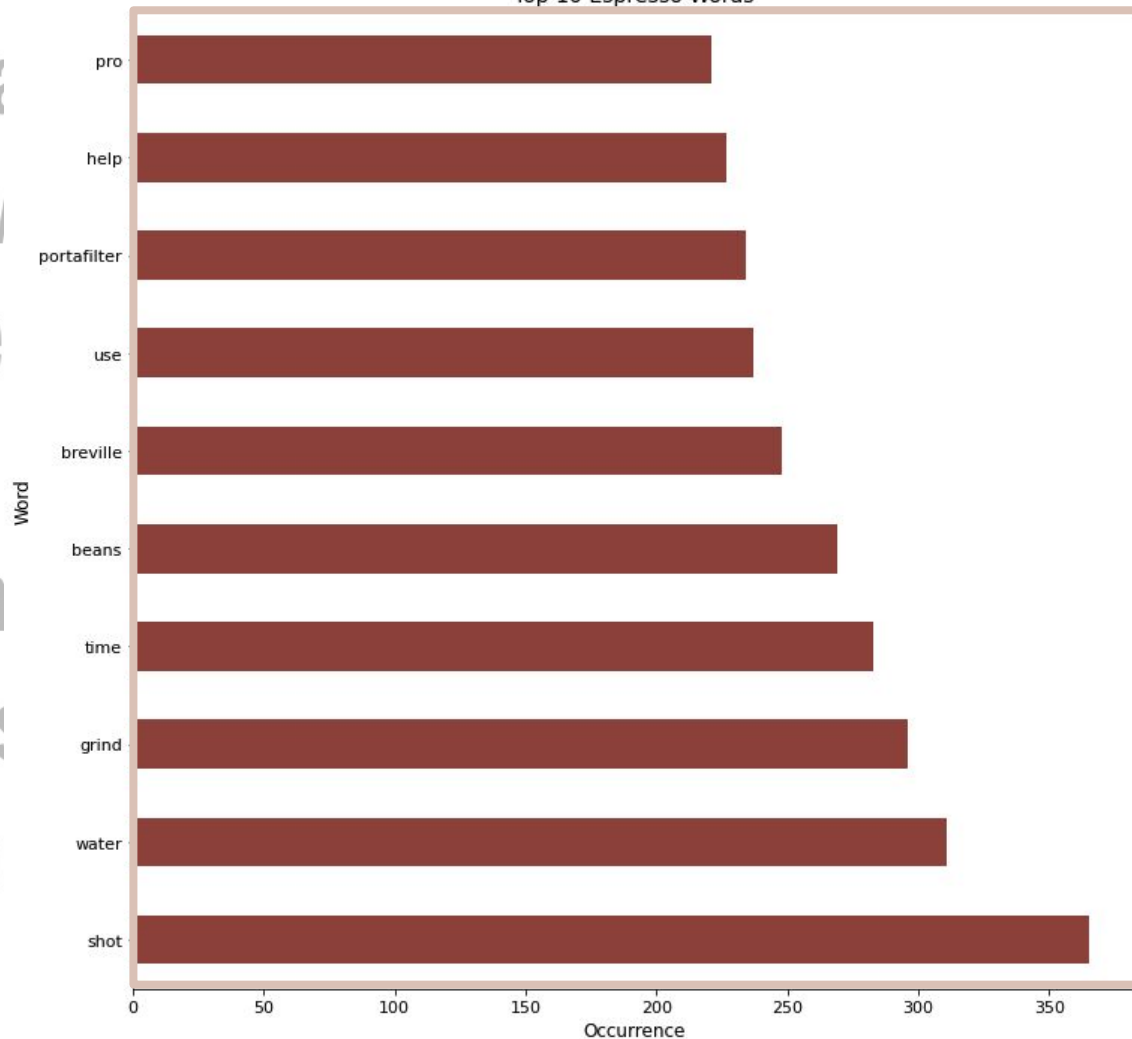
TfidfVectorizer +
LogisticRegressor

94% accuracy score

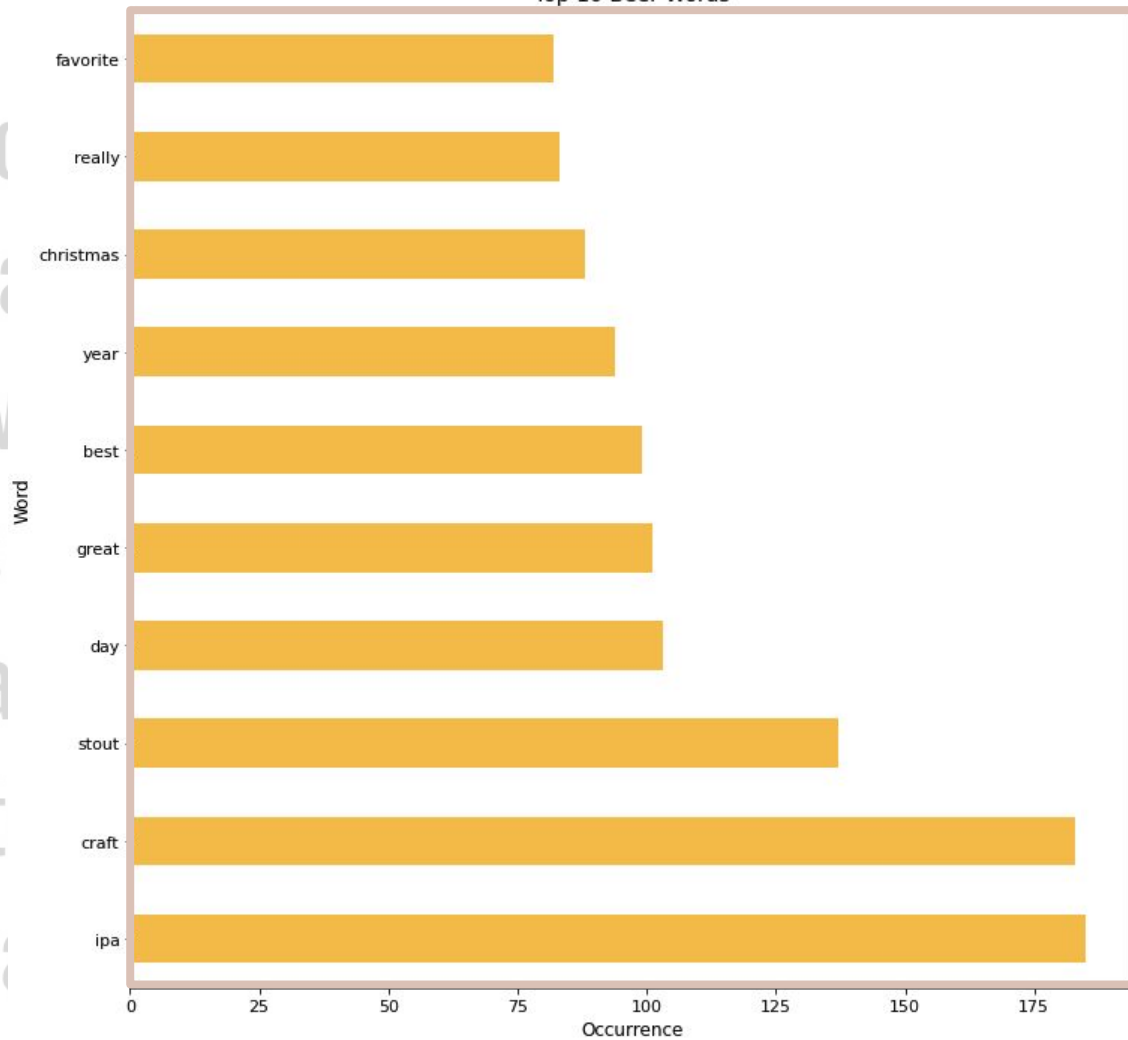
Distribution of Title Word Count in Subreddits

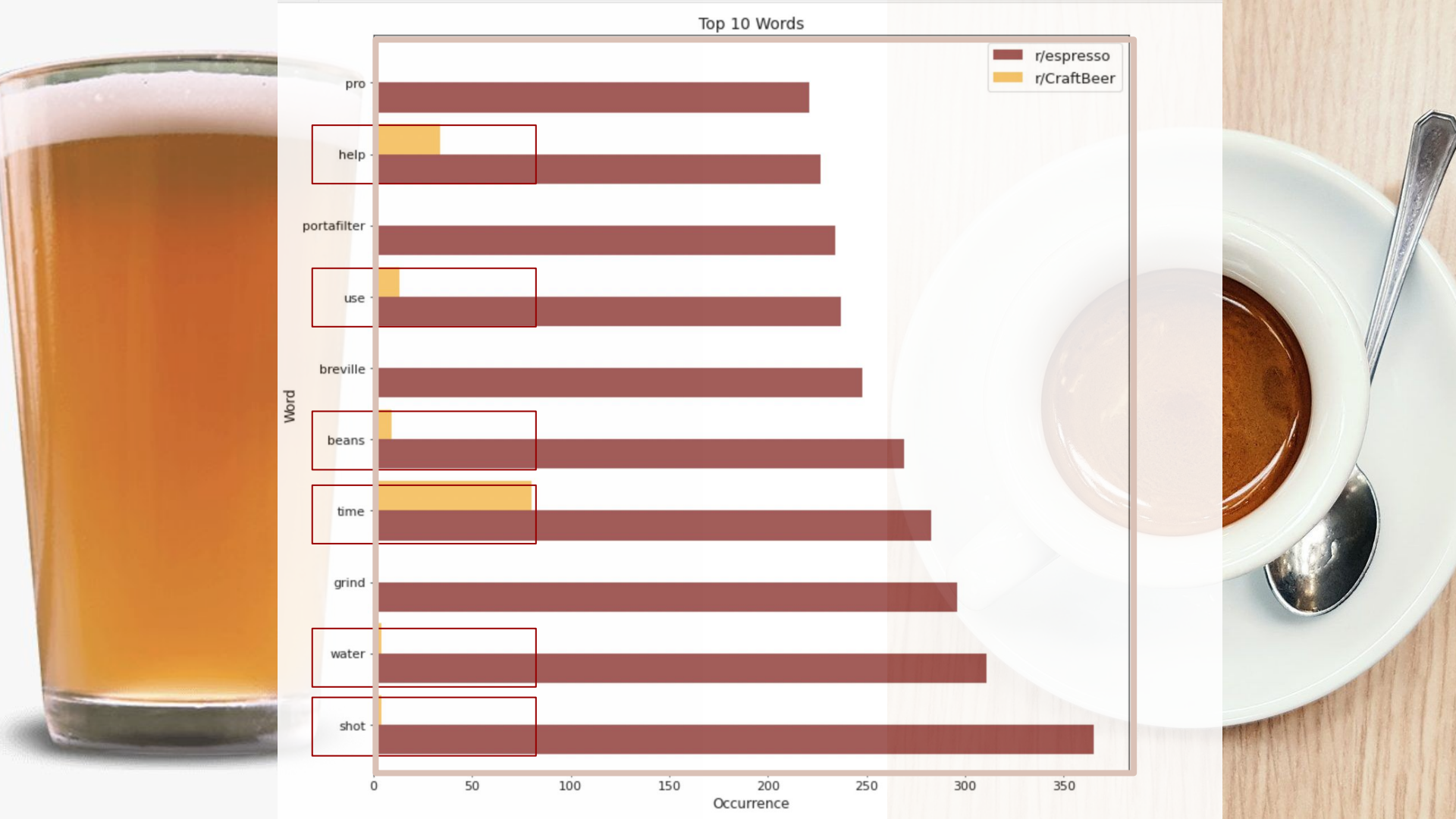


Top 10 Espresso Words

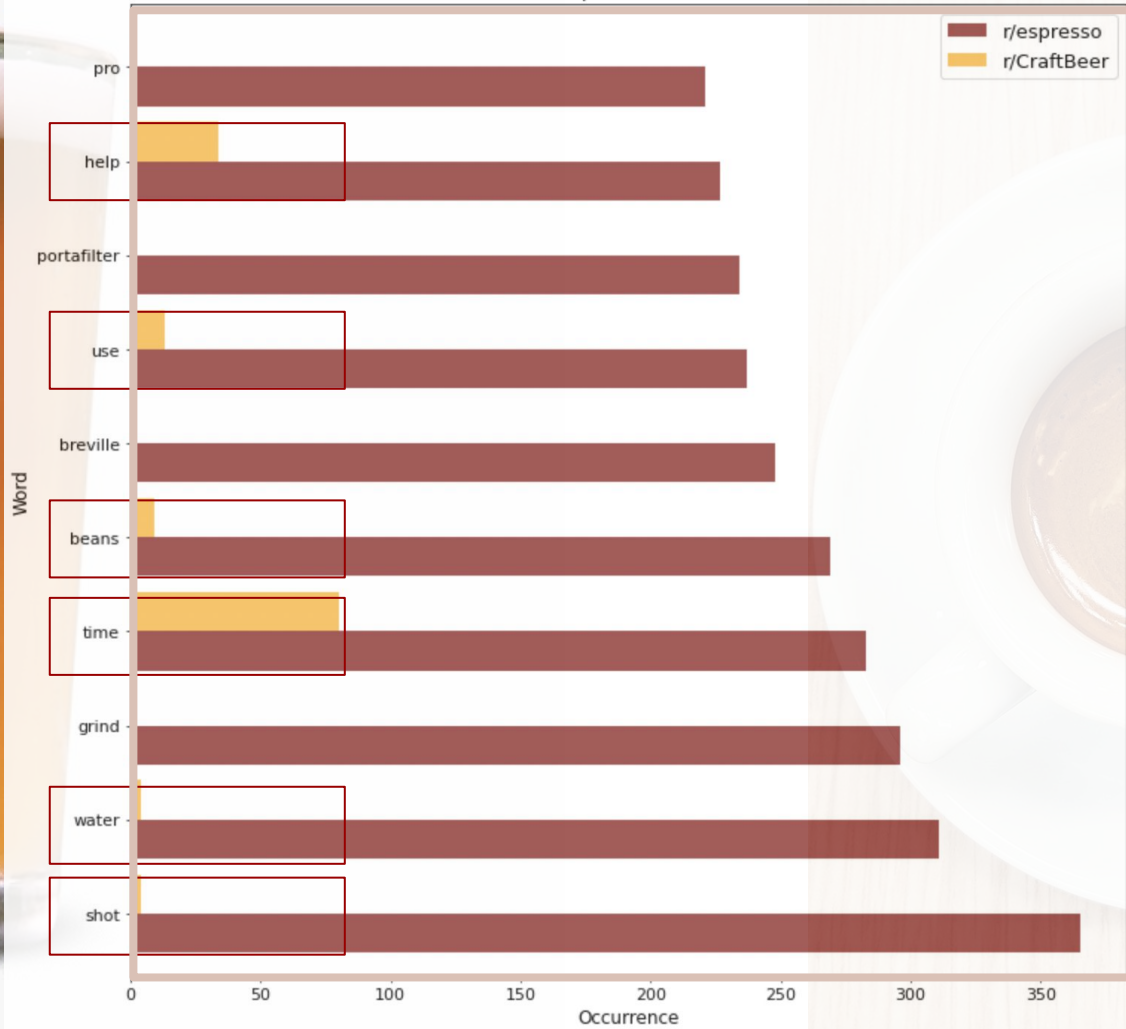


Top 10 Beer Words

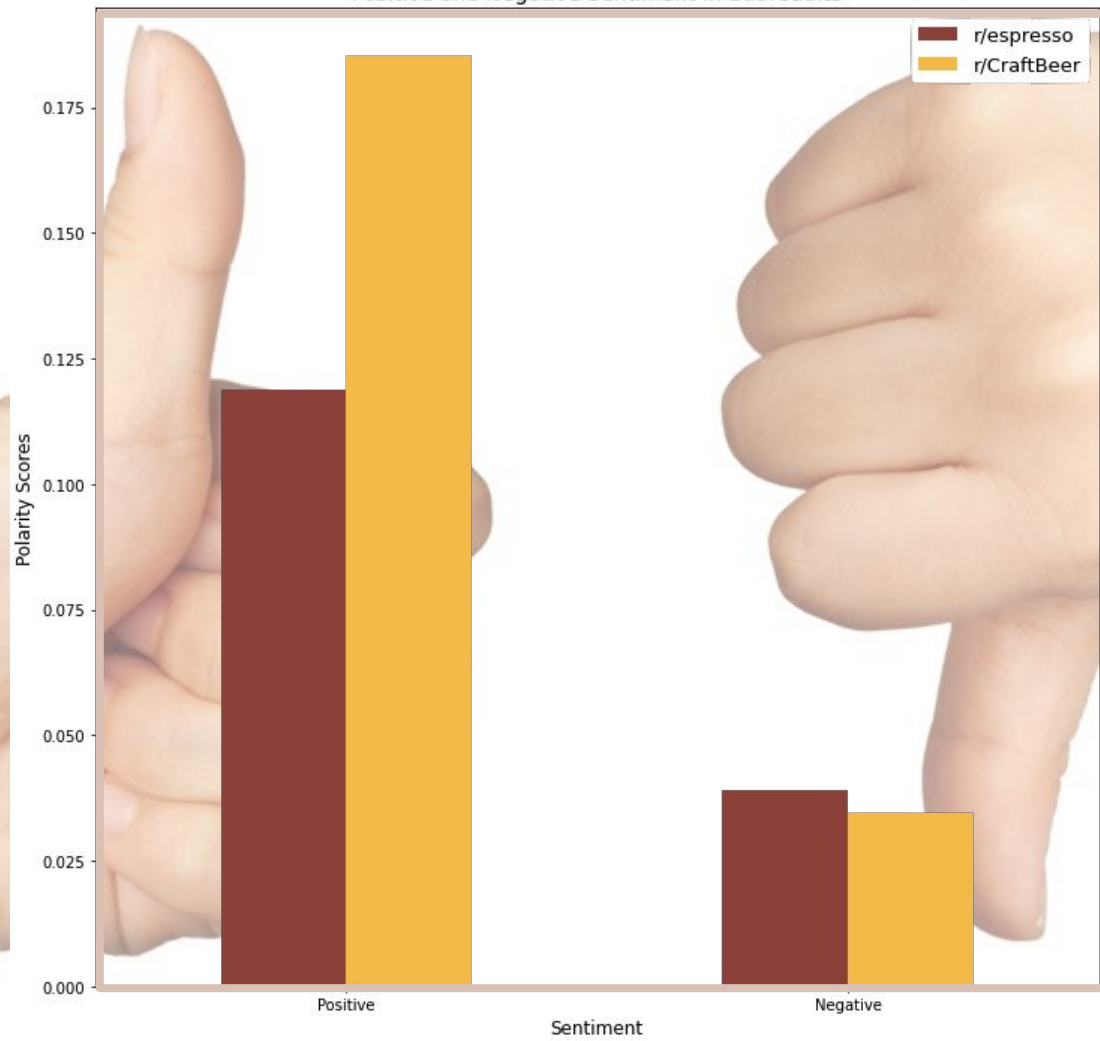




Top 10 Words



Positive and Negative Sentiment in Subreddits



03

RECOMMENDATIONS



Customer insight can be gleaned from applying NLP to user activity on reddit and the insight could be useful in targeted marketing.

Perhaps we need more coffee-flavored beer, or beer-flavored coffee.

Perhaps also this model can be generalized to other data to determine interest across other beverages or food.

Recommendations include but are not limited to:

- applying this model to text in other social media to look for new markets for existing products
- investigating the mis-categorizations within the model for insight into interest in new cross-over products.

“Half of the day I wonder if
it’s too late for coffee, the
other half I wonder if it’s
too early for alcohol.”

—RONALD ULYSSES SWANSON

A close-up photograph of two hands holding beer glasses, clinking them together in a toast. The glasses are filled with a golden beer topped with a thick white head of foam. The background is blurred, showing what appears to be an indoor setting with warm lighting. A semi-transparent white rectangular box is centered over the image, containing the text 'CHEERS!' and 'Questions?'.

CHEERS!

Questions?

Credits:

PHOTOS:

<https://www.sacurrent.com/sanantonio/texas-craft-brewers-react-to-emerging-beer-trends-with-both-new-and-backward-looking-styles/Content?oid=27319151>

<https://www.travelandleisure.com/food-drink/beer/best-breweries-in-seattle>

<https://www.educatall.com/page/1157/Thumbs-up-or-down.html>

<https://www.timeout.com/seattle/bars/best-breweries-in-seattle>

<https://twigscafe.com/what-is-espresso-shot-all-you-need-to-know-about-espresso/>

Beer Day Campaign by Slidesgo
Popic Café Company Profile by Slidesgo