

NATURAL LANGUAGE PROCESSING - SUBREDDIT CLASSIFICATION

GANESH MORYE

BUSINESS PROPOSITION

- A baby food start up company has decided to use Reddit for their social media campaign to market their yet to be released baby food product
- Guide the marketing team at your company to design a data-driven marketing campaign

WHICH TRANSLATES TO....



**IDENTIFY THE
CUSTOMER**



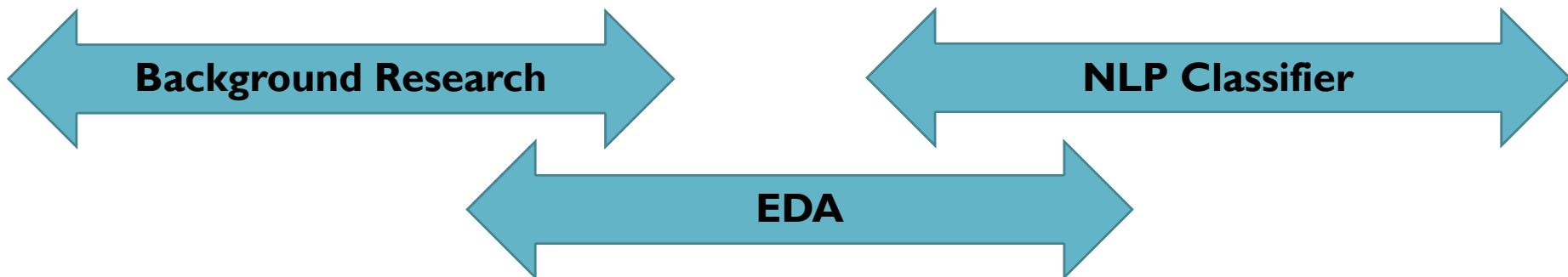
**RESEARCH THE
CUSTOMER**



**GENERATE AD-
RELEVANT DATA
ABOUT YOUR
CUSTOMER**



**EFFECTIVELY
TARGET THEM**



WHY ADVERTISE ON REDDIT?

- #20 in global internet traffic and engagement
- #10 most popular social media sites in the USA
- Almost 40% of the site traffic comes from USA
- 52 million daily active Reddit users worldwide
- 2.8 million subreddits
- Ad competition is low, and inventory is cheap

AD STRATEGIES

- How to design a campaign to target specific subreddits
- How to engage with the communities to gauge the pulse
- Know the Reddit lingo to create ads that will catch user attention
- Messages unique to the subreddit for better response
- Promote your content

All this requires a lot of research and getting to know which subreddits are most relevant to your audience.

IDENTIFY THE CUSTOMER

r/BabyBumps – “*A place for pregnant redditors, those who have been pregnant, those who wish to be in the future, and anyone who supports them.*”

r/beyondthebump (aka Baby Bumps Graduates) – “*A place for new parents, new parents to be, and old parents who want to help out. Posts focusing on the transition into living with your new little one and any issues that may come up. Ranting and gushing is welcome!*”

**RESEARCH
THE
CUSTOMER**





**WHAT ARE
THEY TALKING
ABOUT IN
R/BABYBUMPS**

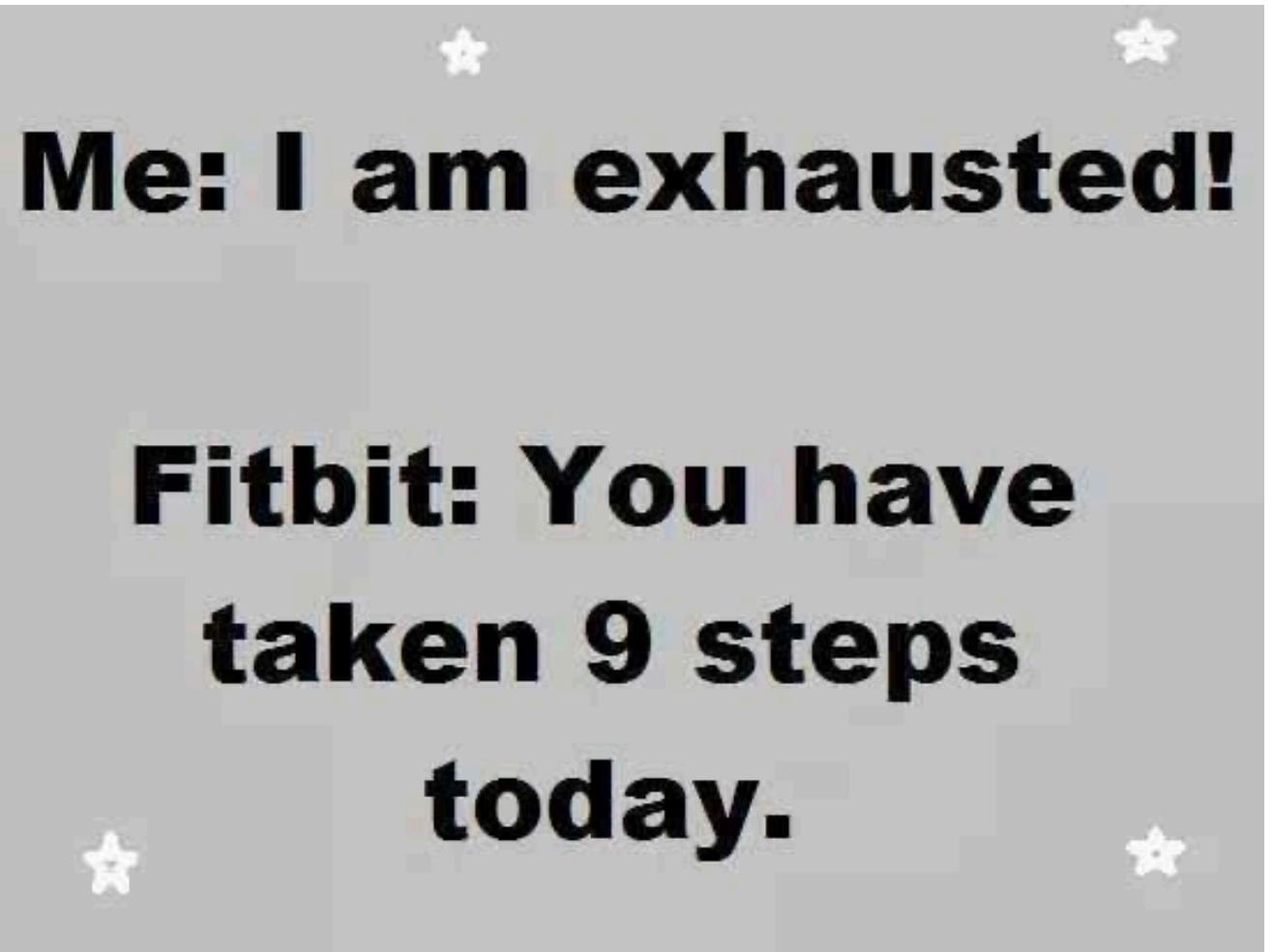
DESIGNING NURSERY





EATING HEALTHY

STAYING ACTIVE



**HOW
EXCITED
THEY ARE...**



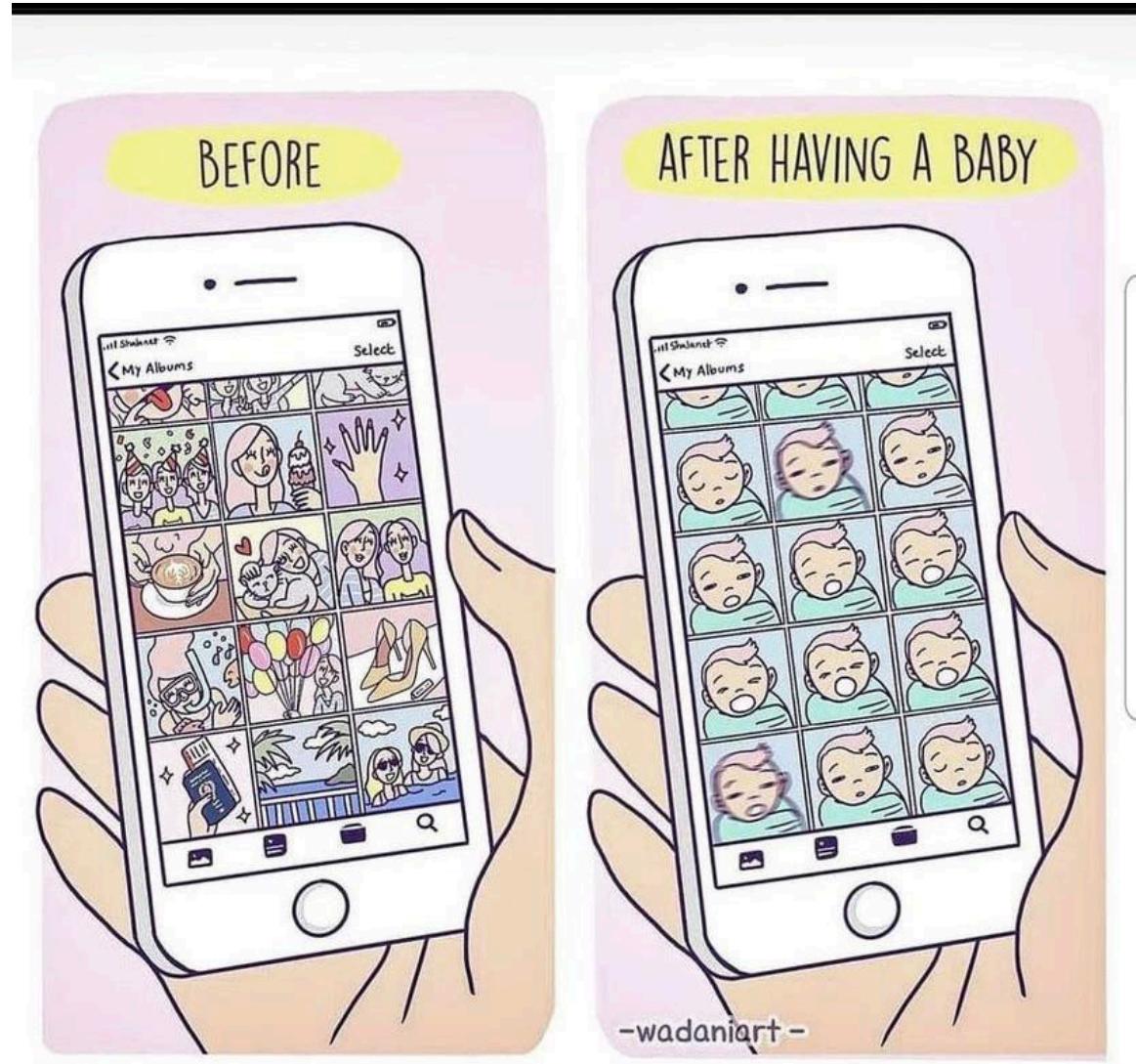
**WHAT ARE THEY
TALKING ABOUT IN
R/BEYONDTHEBUMP**





FEEDING THE BABY...

TRACKING PROGRESS



FUN IN LIFE..

"parenting has to be one
of the easiest jobs
to have."



WORK FROM HOME

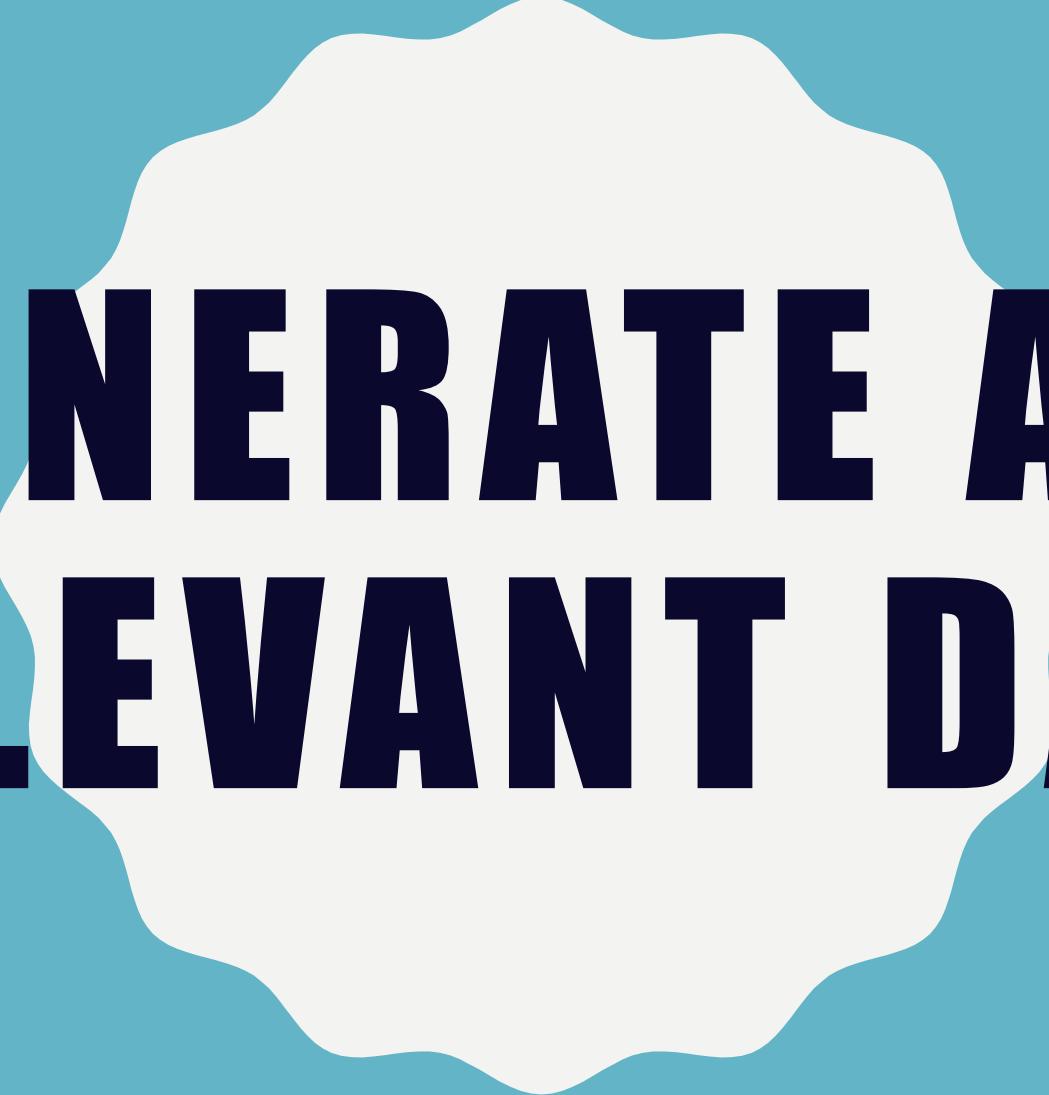
Me: *tries to focus*
My last 3 brain cells:



**IT IS ALL
ABOUT...**



A DAY IN THE LIFE



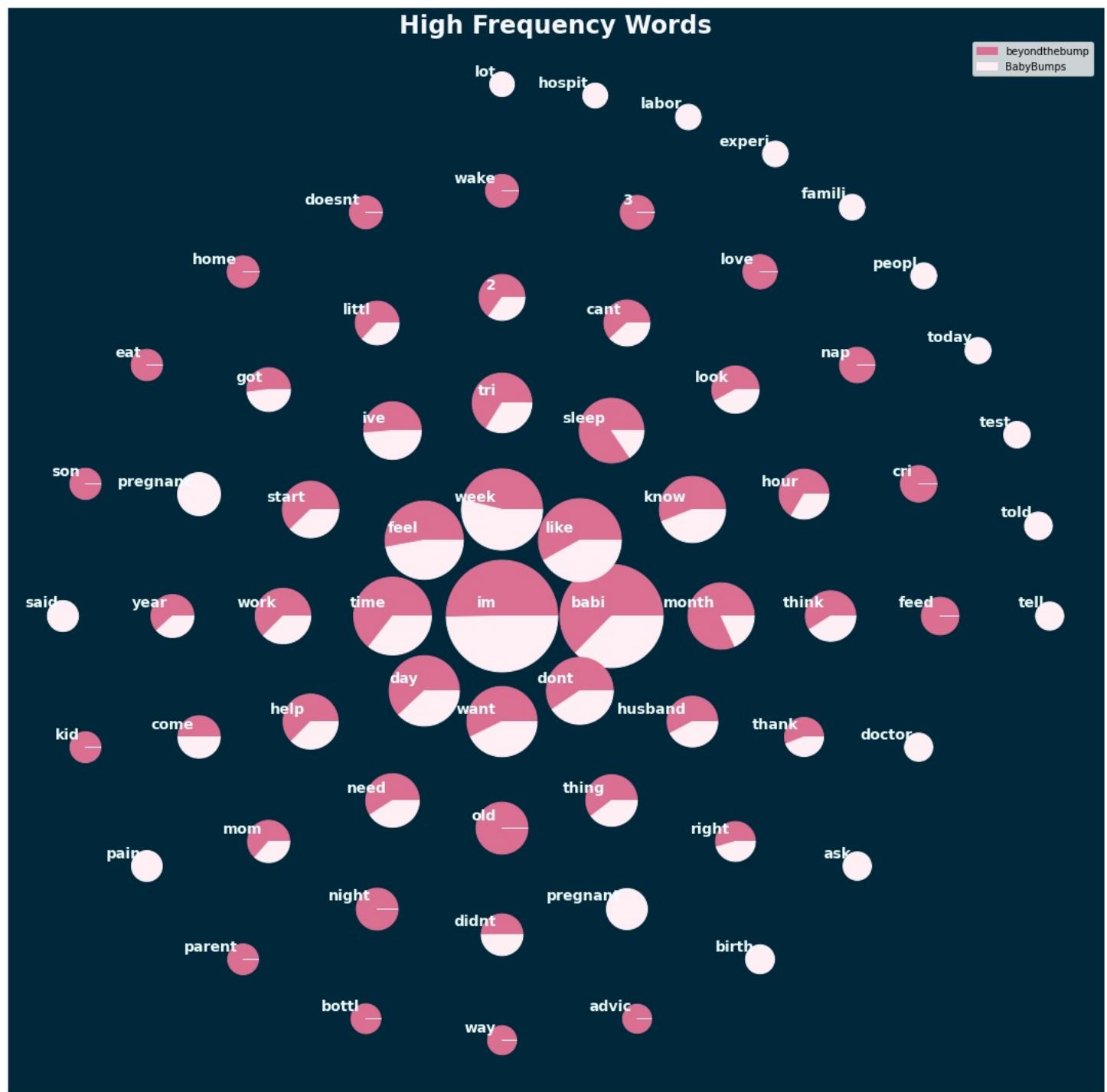
**GENERATE AD-
RELEVANT DATA**

SPACY AND NLTK TO THE EDA

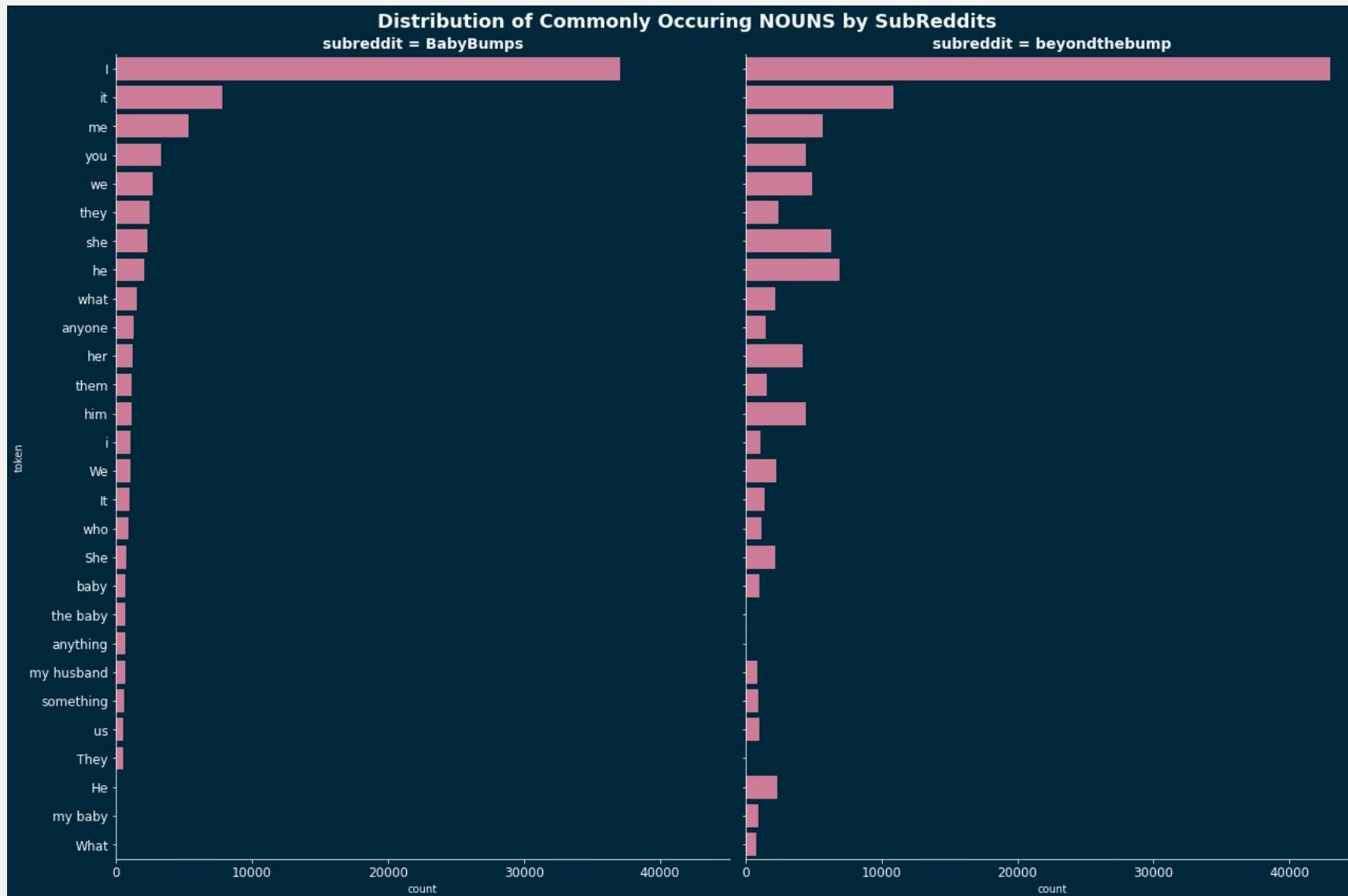
A typical Reddit post

Graduation! FTM, 40+3 DATE . Post almost as long as my labor & org ; unsolicited advice I loved reading these when I was pregnant. Now that my 2-week old DATE (update: revisiting at 1 month DATE LOL) lets me sleep a continuous 3.5 hours TIME , its my turn!Im a first ORDINAL time mom, although in 2020 DATE my first ORDINAL pregnancy ended in a miscarriage around 9 weeks DATE . That experience absolutely shaped how I experienced at least the first half of this pregnancywe didnt announce to close family until ~18 weeks DATE , after our NST ORG showed a boy with no chromosomal abnormalities and we felt it was 1 CARDINAL) real and 2 CARDINAL) safe to announce. At the 20 week DATE scan we found out our lil guy had a 2 CARDINAL -vessel umbilical cord, which gave us more monitoring in the form of ultrasounds at 28 and 36 weeks DATE , and weekly DATE non-stress tests starting at 36w PRODUCT . These were to check for growth restriction and poor blood flow. Thankfully lil bug was doing great. I tried to stay active and relatively stress-free throughout pregnancy in part to try to keep my blood pressure low so the single artery didnt have to work against high blood pressure. So I was running until 35 weeks DATE , biking until 39 DATE ,

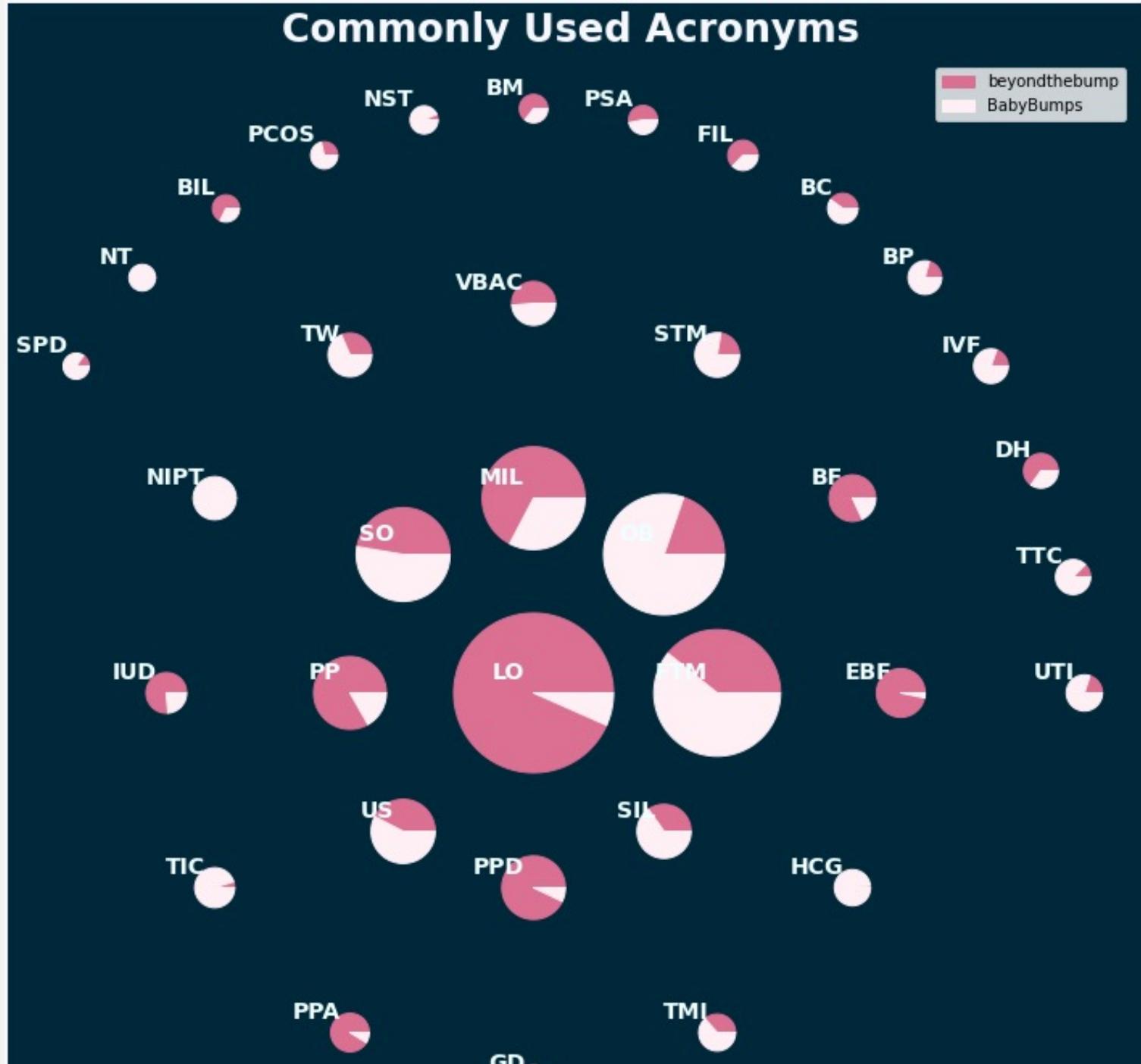
WHAT ARE THE DIFFERENT TOPICS OF CONSERVATION ??



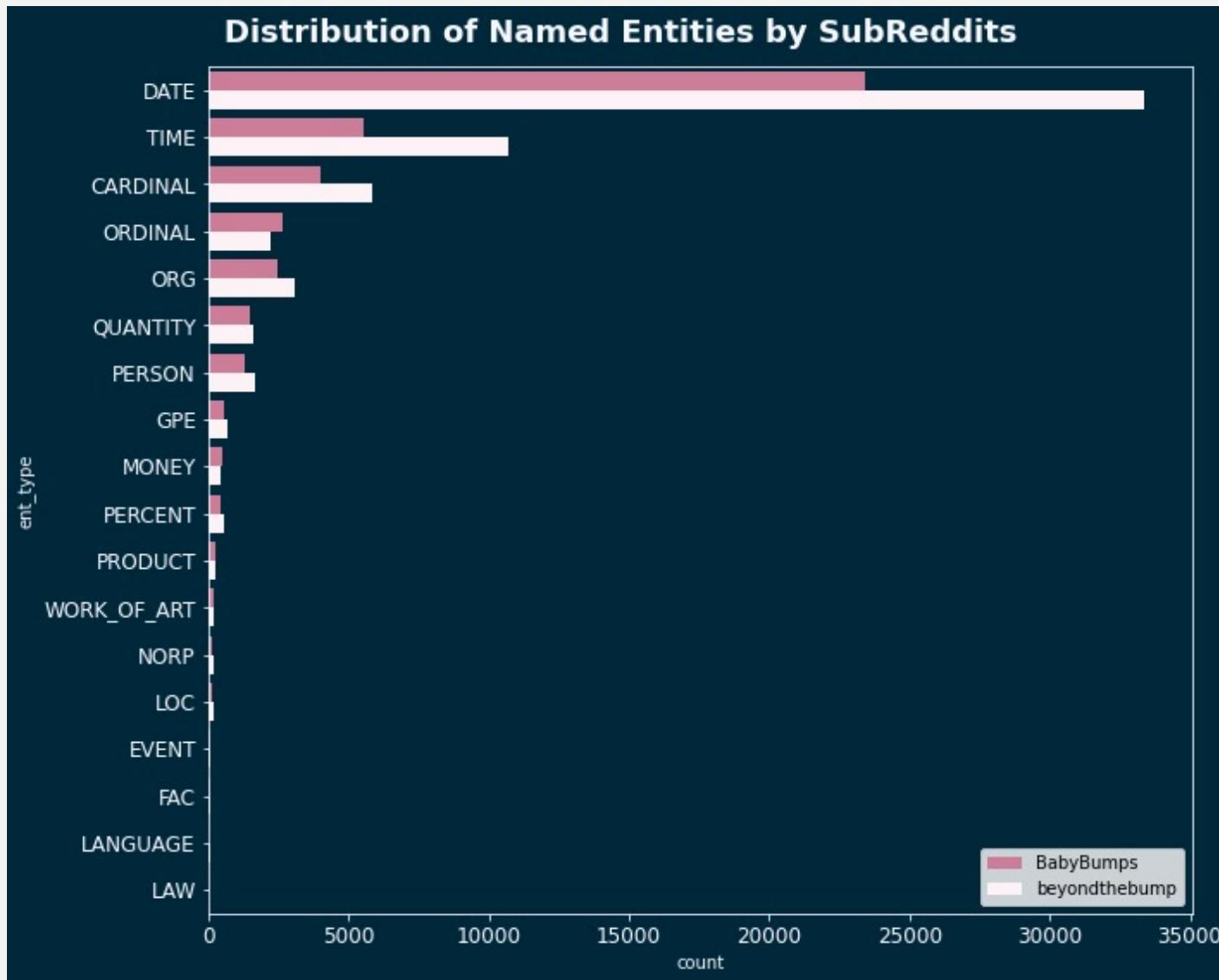
WHO ARE THEY TALKING ABOUT??



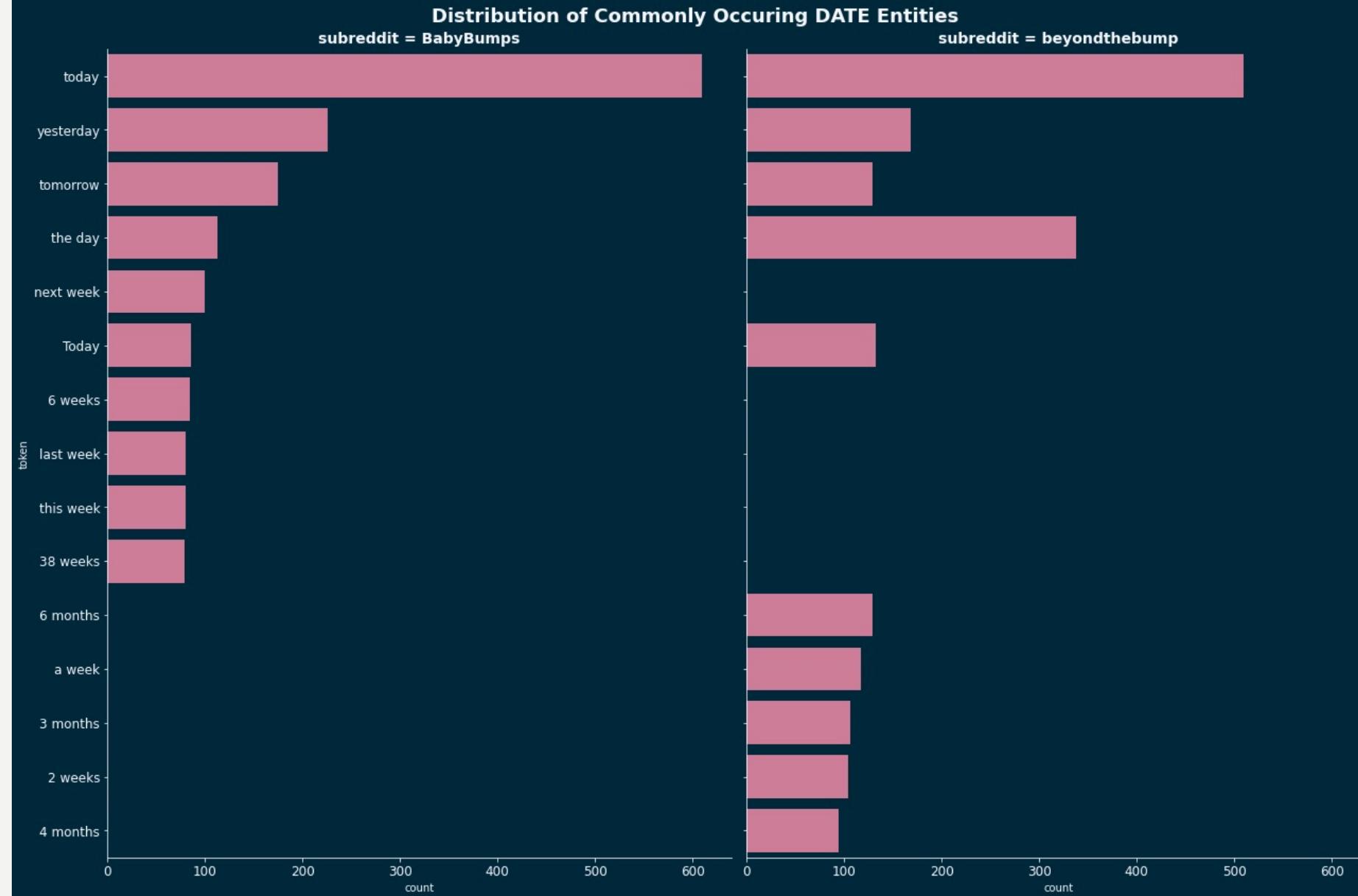
**REDDITORS
LOVE THEIR
LINGO...**

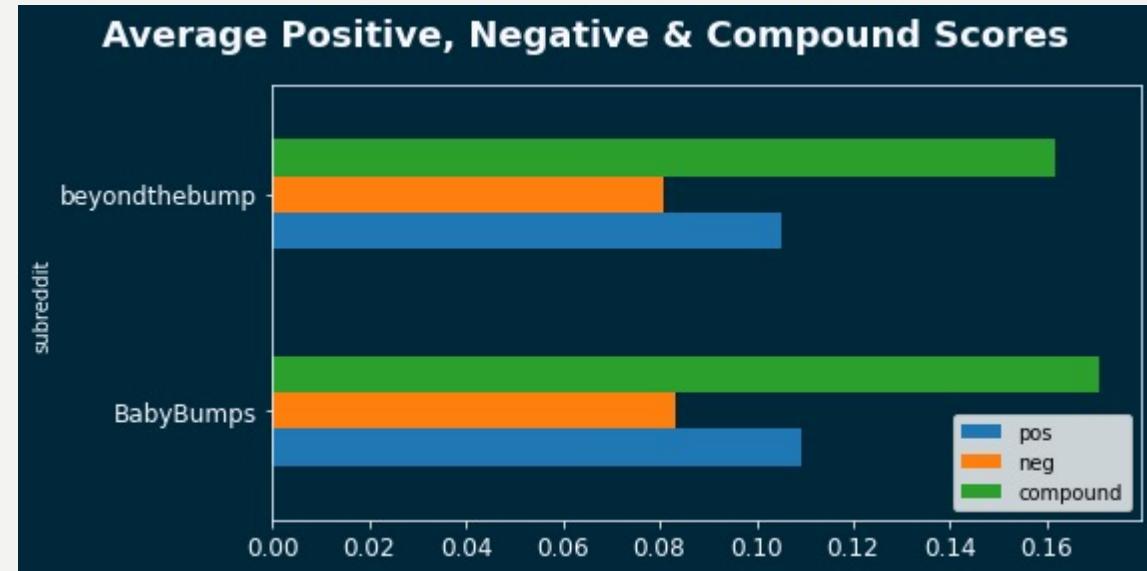
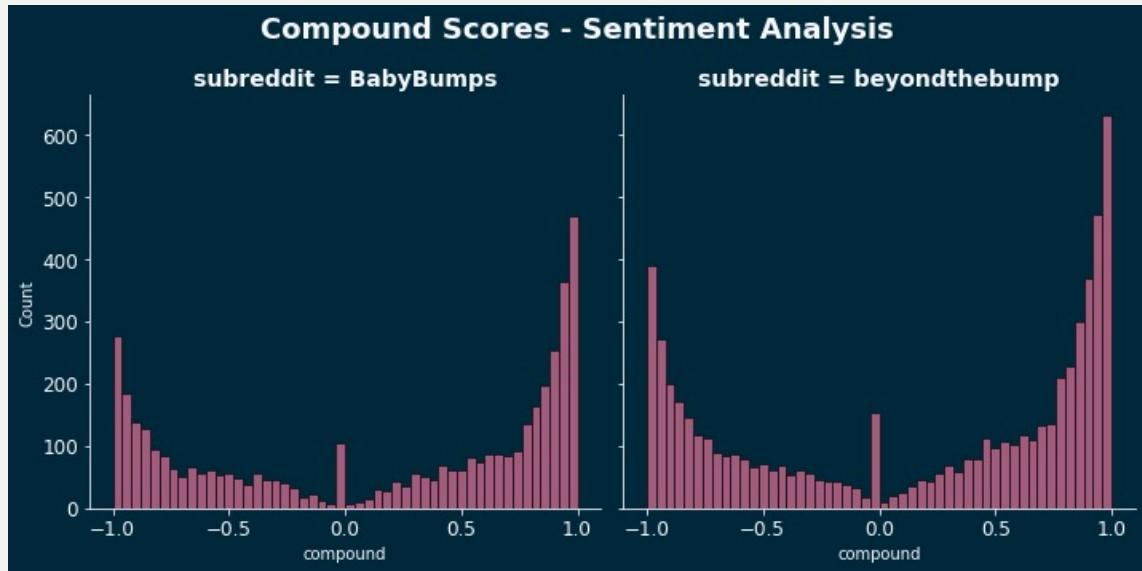


WHAT IS MOST TALKED ABOUT??



WHY ARE THEY SO INTERESTED IN DATES??





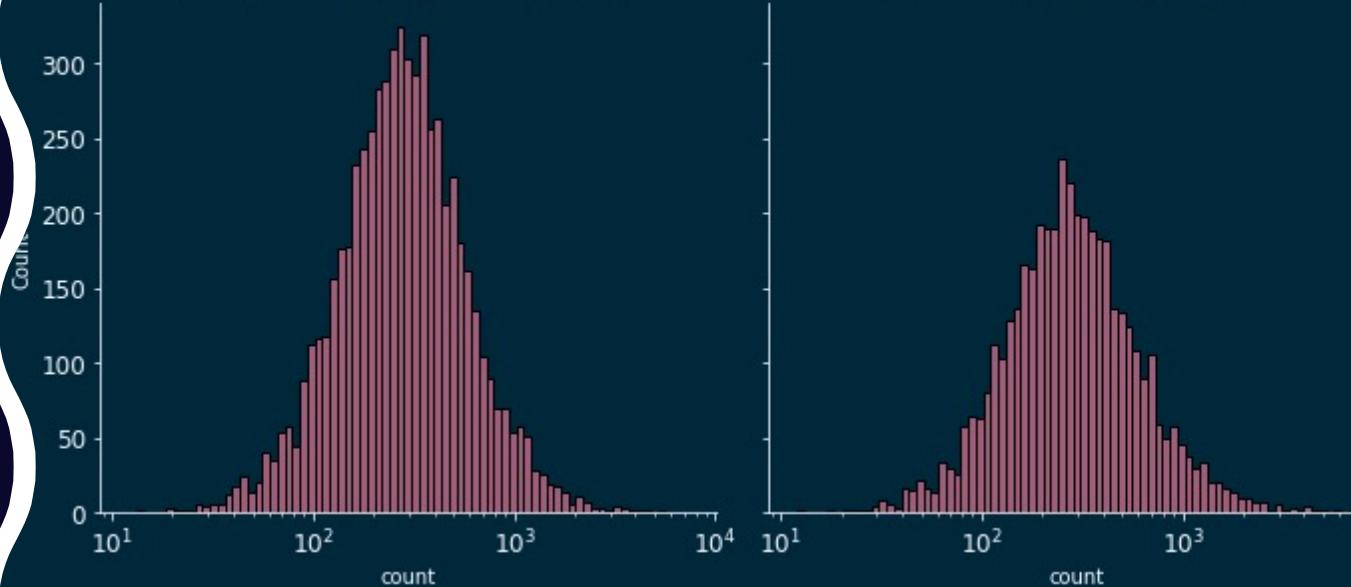
HOW'S THE MOOD

WHO IS THE CHATTY ONE?

Post Lengths Distribution by SubReddits

subreddit = beyondthebump

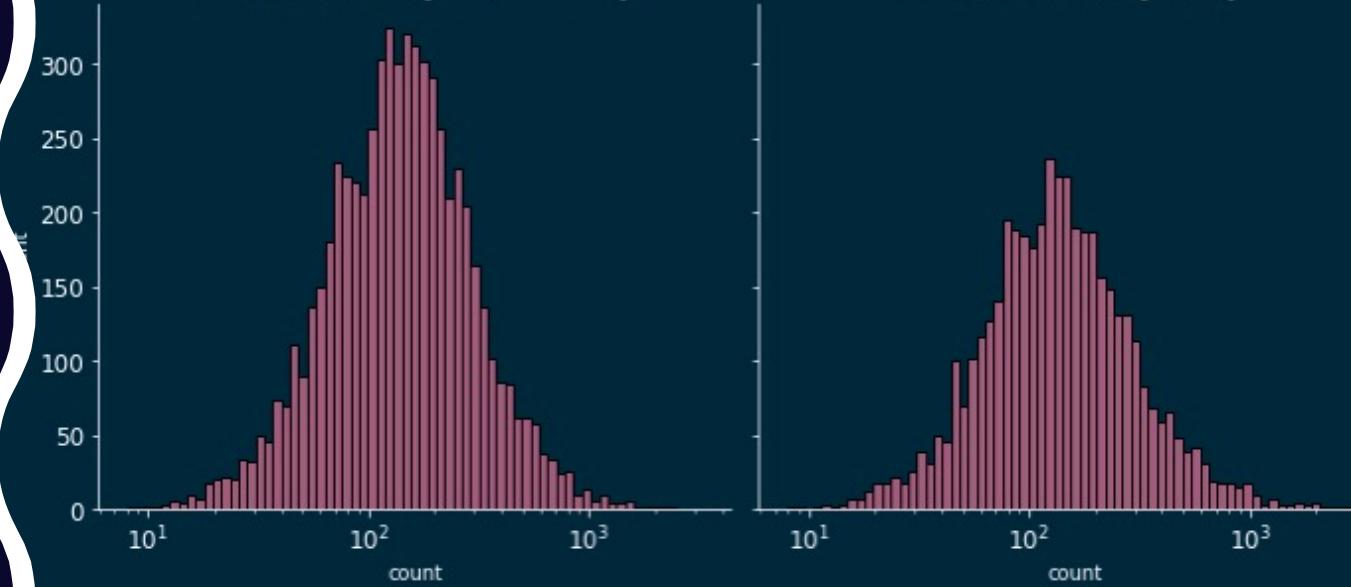
subreddit = BabyBumps



Token Counts Distribution by SubReddits

subreddit = beyondthebump

subreddit = BabyBumps





TARGETING YOUR CUSTOMER

THE MODELING BIT OF IT

4 different classification models

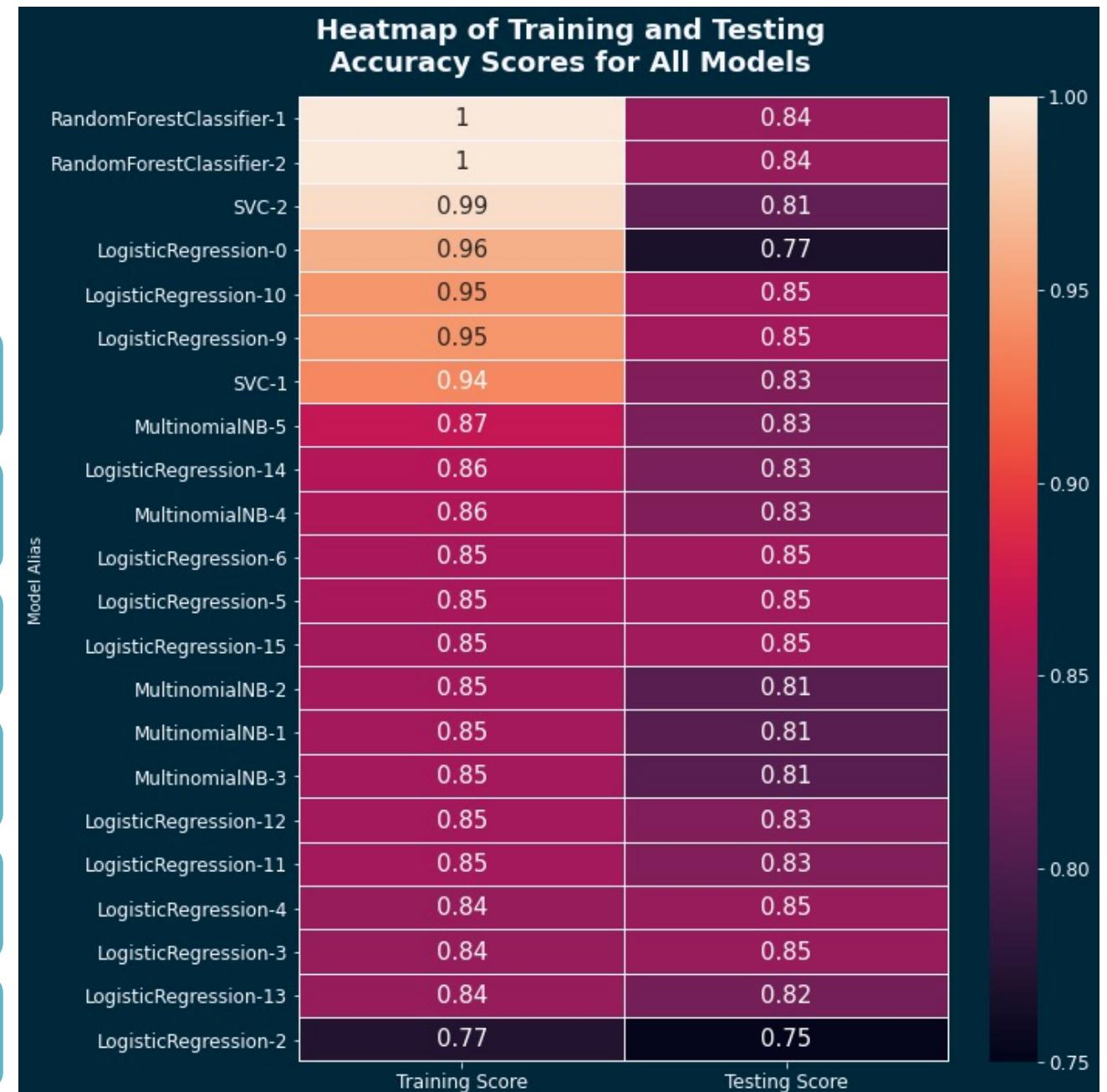
Tuning the hyperparameters to reduce overfitting

RandomForest Classifier is the most overfit

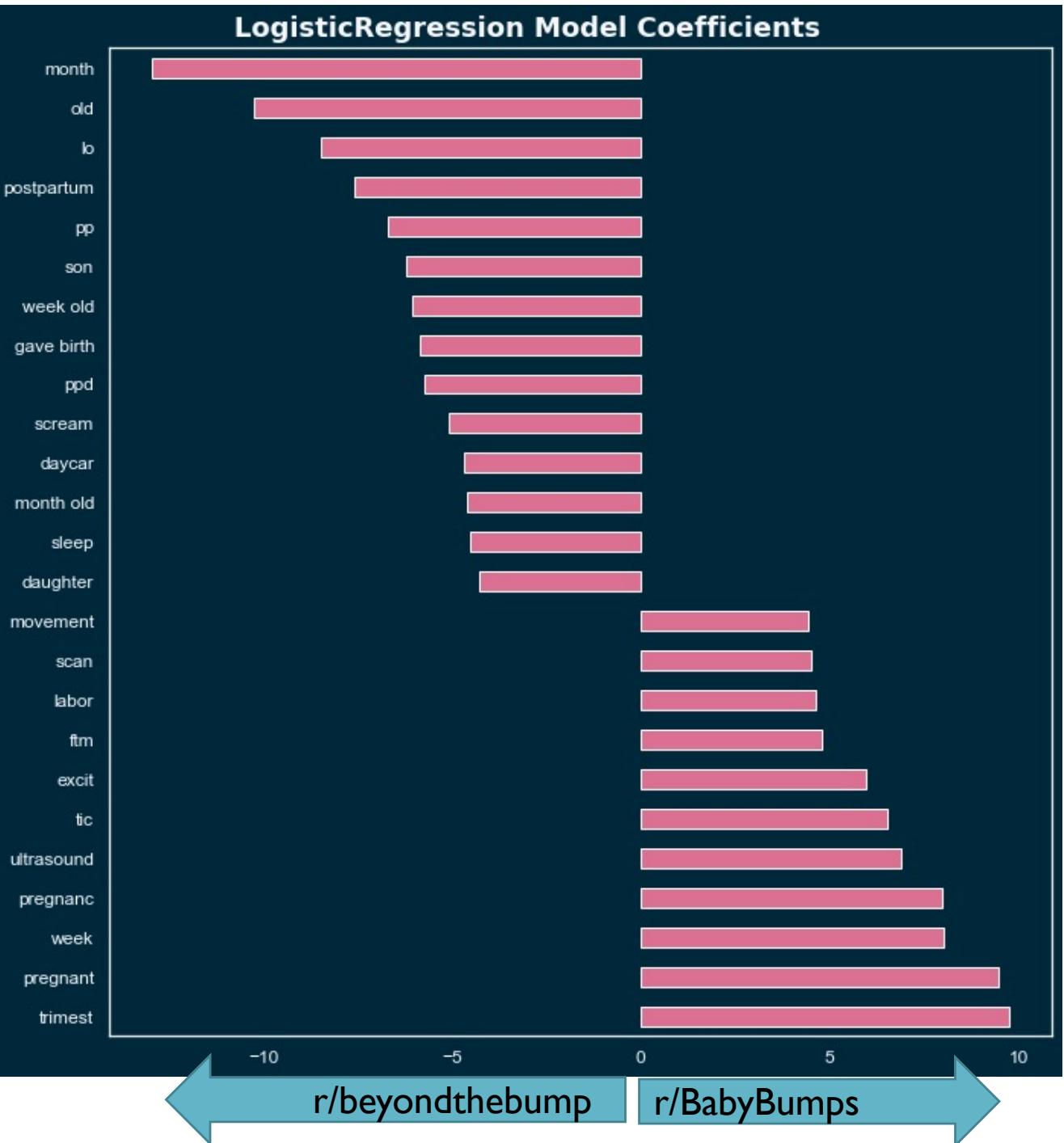
SVC model is overfit even after tuning the hyperparameters

MNB is relatively less overfit but its testing scores are lower

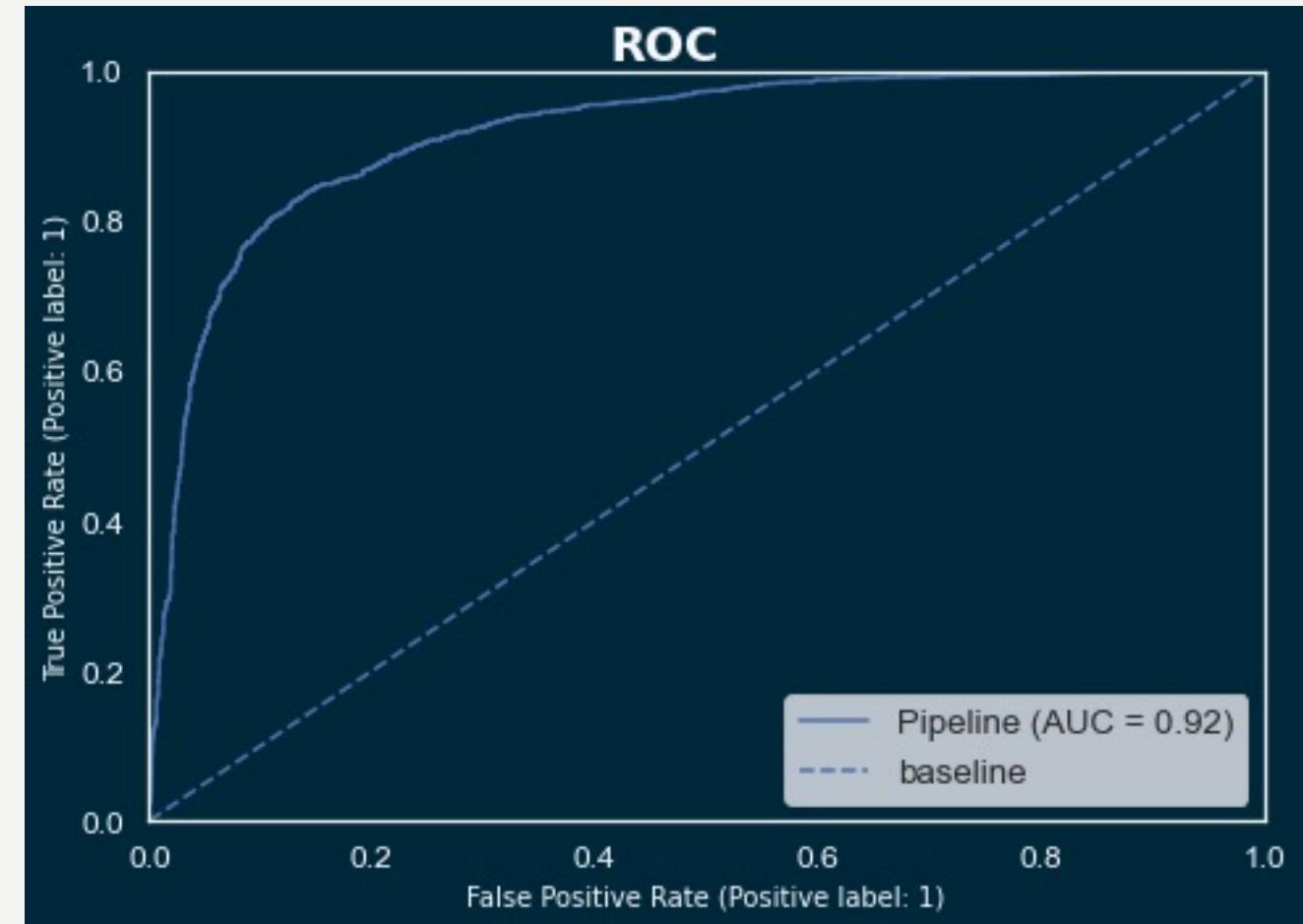
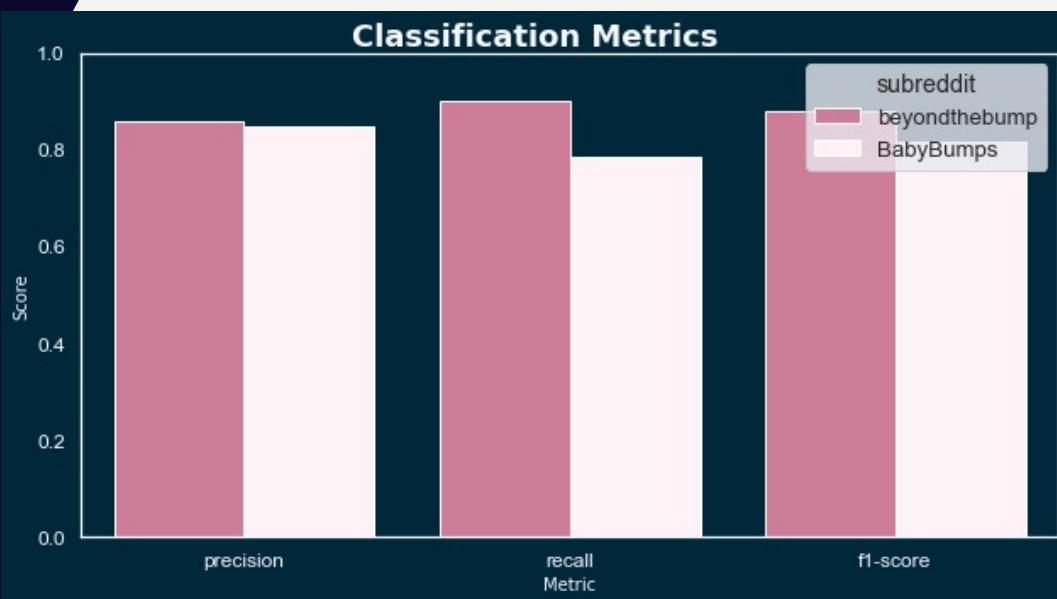
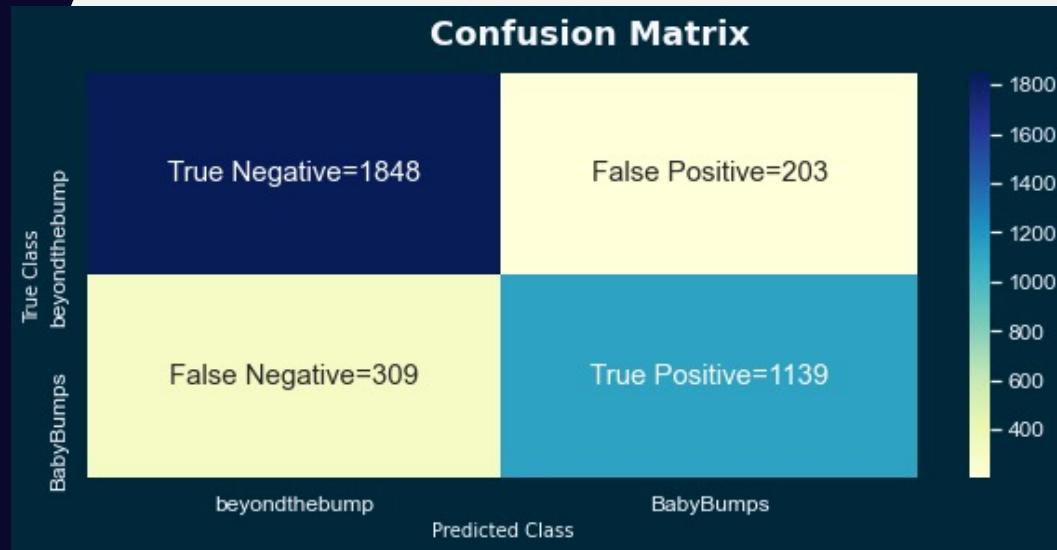
LogisticRegression has the best bias-variance tradeoff and is chosen as the desired model



INFLUENTIAL SEARCH TERMS



METRICS



	precision	recall	f1-score
beyondthebump	0.856745	0.901024	0.878327
BabyBumps	0.848733	0.786602	0.816487

TO CONCLUDE



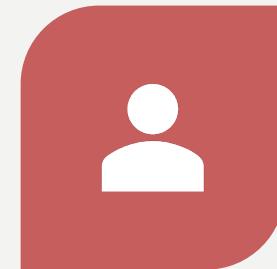
IDENTIFY THE CUSTOMER

r/beyondthebump
r/BabyBumps



RESEARCH THE
CUSTOMER

Pregnancy and their
little one's related
posts



GENERATE AD-RELEVANT
DATA ABOUT YOUR
CUSTOMER

I, DATES, LINGOS,
Positive, trimester,
month, pregnant...



EFFECTIVELY TARGET
THEM

NLP classifier to
help generate
targeted promotion
content