SUBREDDIT NLP MODELING

Created by Veda Patel

TABLE OF CONTENTS
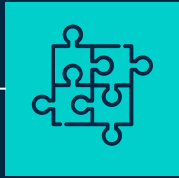
01
PROBLEM STATEMENT
Why do this project?

02
THE PROCESS
Cleaning, exploratory data analysis, modeling

03
CONCLUSIONS & RECOMMENDATIONS
What I learned and what I recommend

Background

What is Reddit?

MY SUBREDDITS

r/AskWomen

- Has 5.3m members
- Rules say a question must be posed

r/AskMen

- Has 5.3m members
- Rules say to have a question in title

AskWomen: Questions about women's thoughts, lives, and experiences
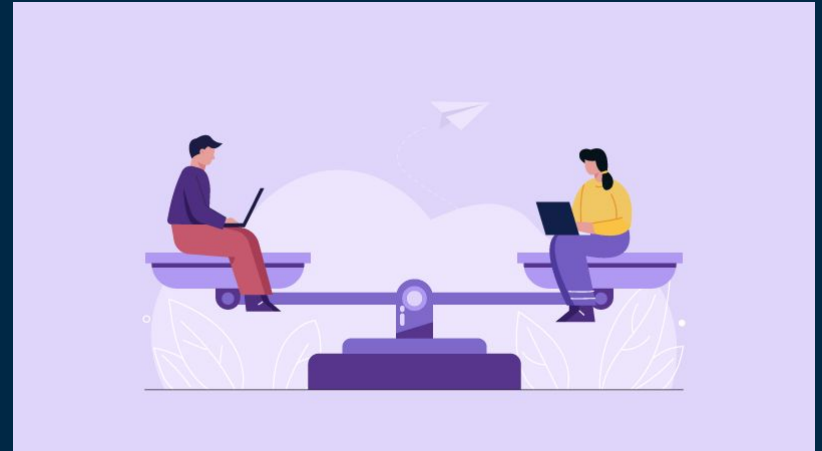r/AskWomen

We don't know, but we'll answer anyway
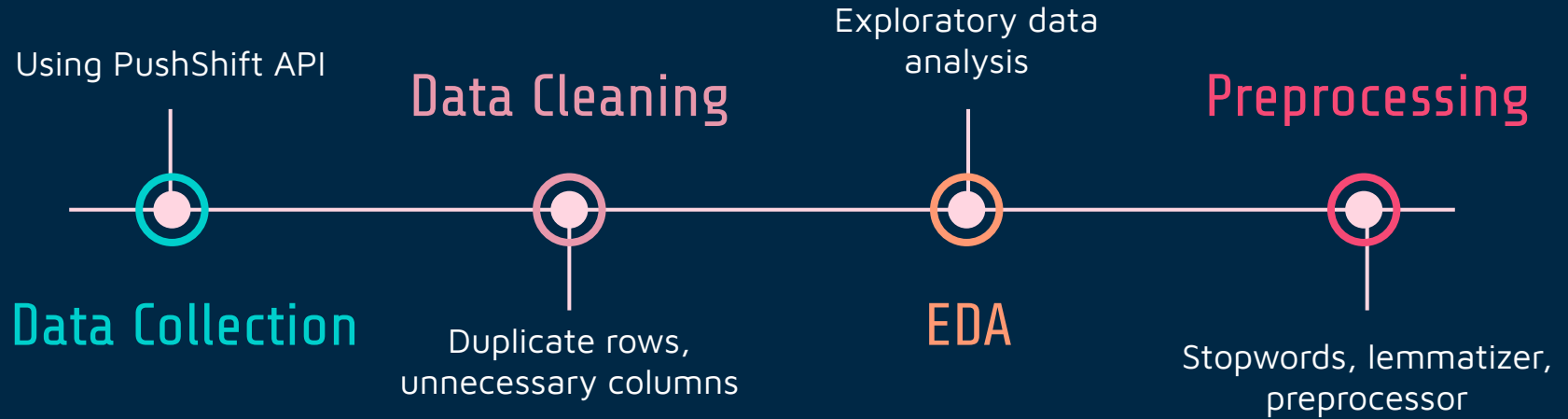r/AskMen

Problem Statement

WHY AM I CONDUCTING THIS PROJECT?

- Self interest

- Social experiment

THE PROCESS

Using PushShift API
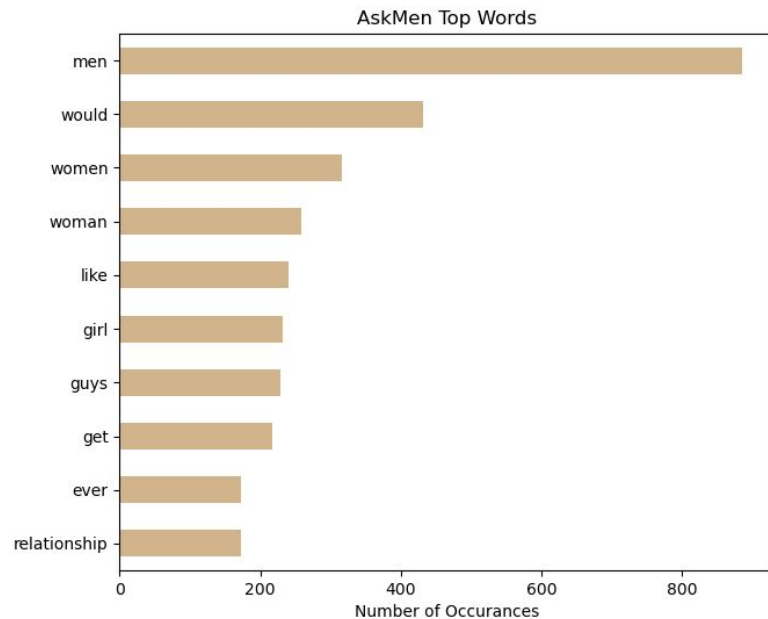
Data Cleaning
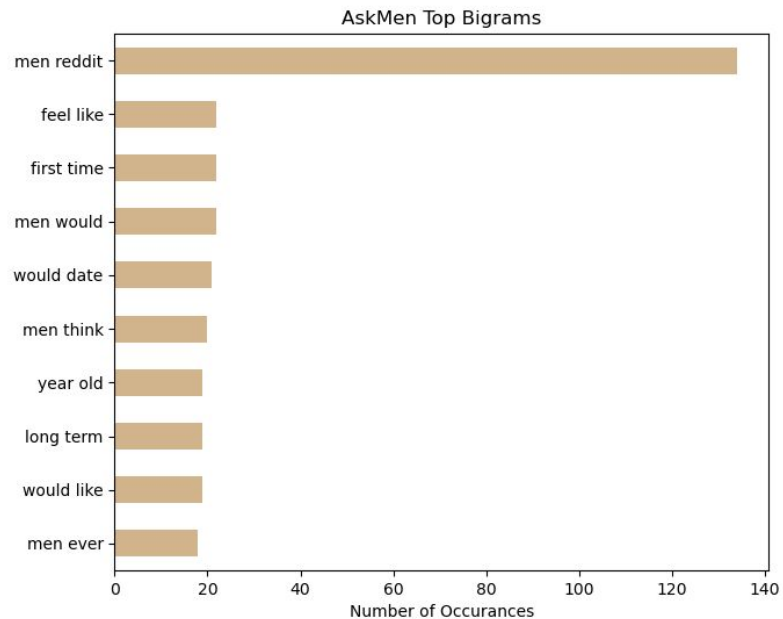
Exploratory data analysis

Preprocessing

Data Collection

Duplicate rows, unnecessary columns

EDA

Stopwords, lemmatizer, preprocessor

From EDA

From EDA

AskWomen Top Words / AskMen Top Words
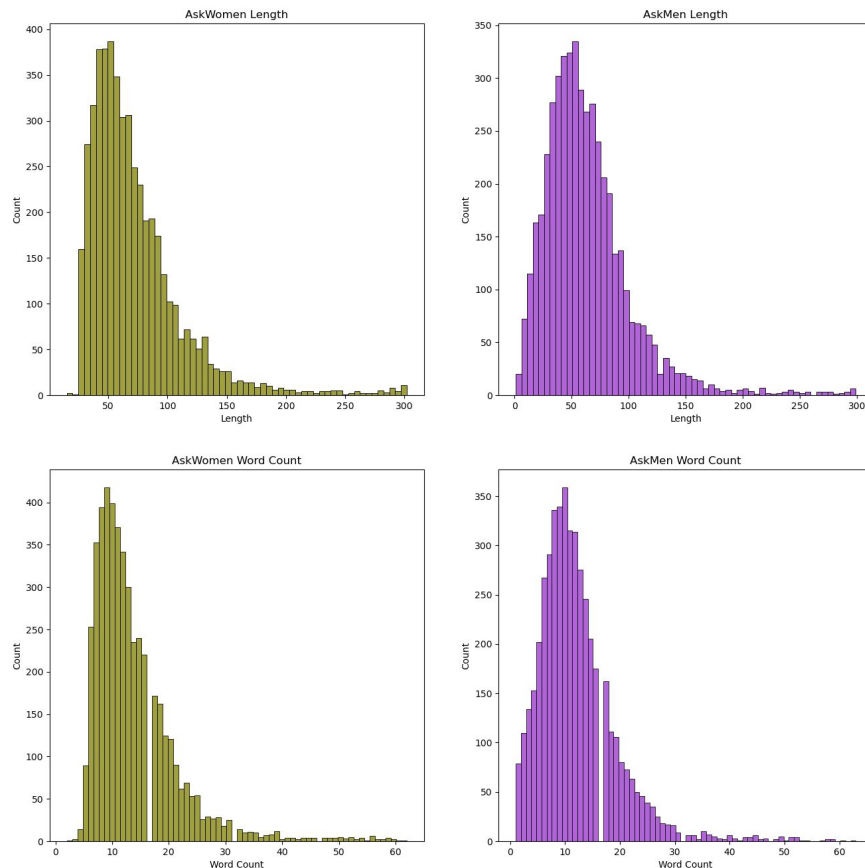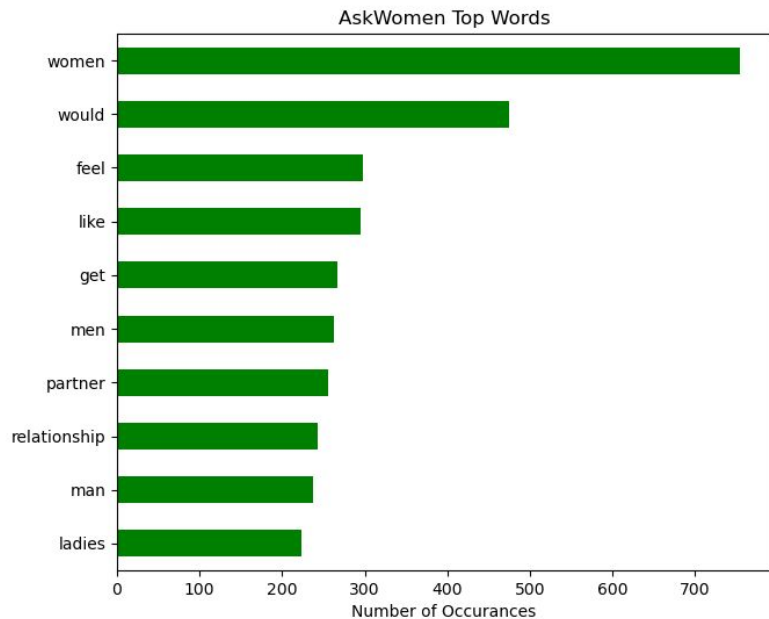
From EDA
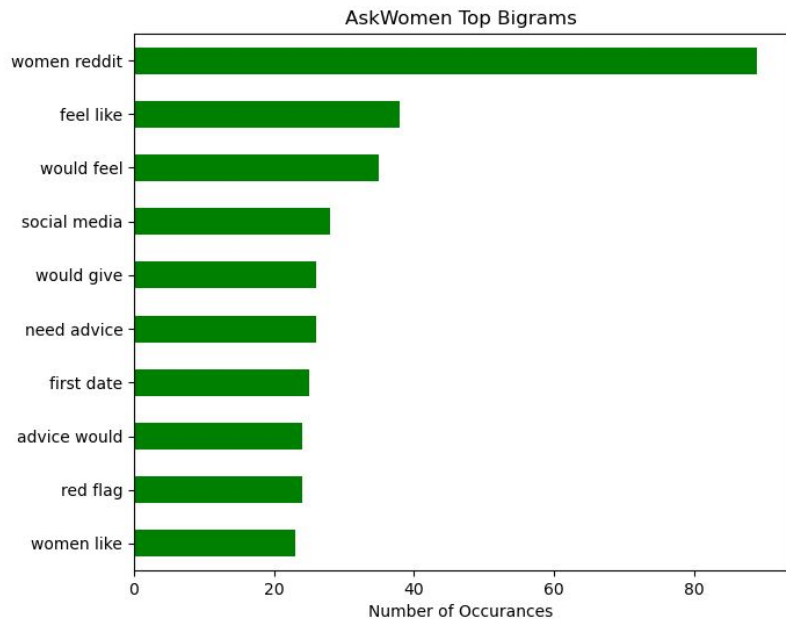
Bigrams after adding to stopwords

social media	28
advice give	28
need advice	26
first date	25
red flag	24
birth control	23
year old	22
best friend	21
first time	21
anyone else	20
best way	20
long take	18
one thing	15
long term	14
date guy	13
please help	12
find attractive	11
relationship someone	11
dating apps	10
body hair	10
dtype: int64	

first time	22
year old	19
long term	19
best way	18
anyone else	18
first date	17
one thing	17
best friend	17
red flag	17
need help	17
need advice	14
guy friend	12
social media	12
reddit ever	11
dating apps	11
dating app	10
advice give	10
even though	10
wife girlfriend	10
someone else	10
dtype: int64	

MODELING

Logistic Regression

Naive Bayes

Random Forest
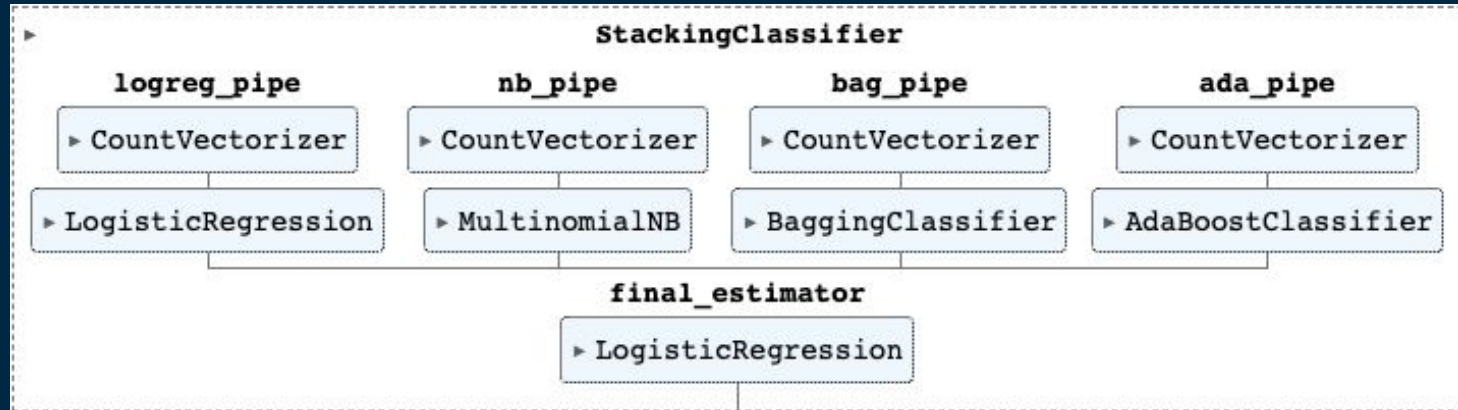
Confusion Matrix and Roc Curve

Misclassified AskWomen Top Words

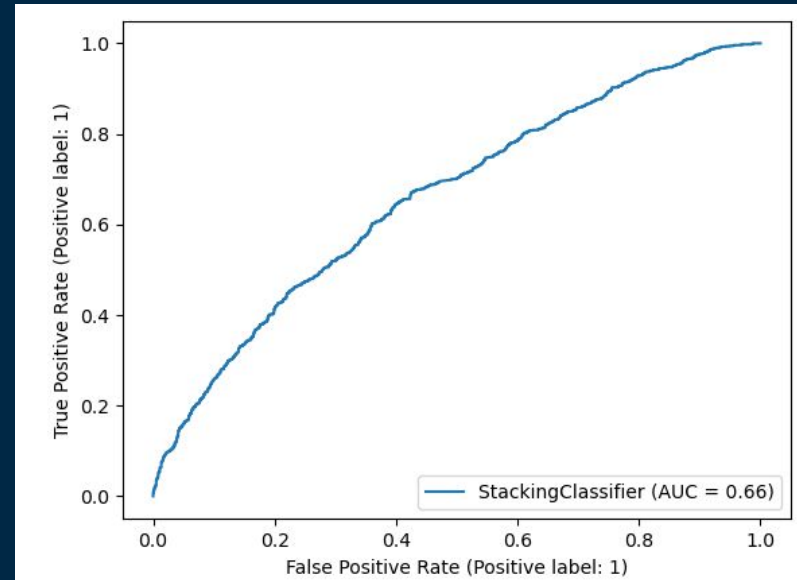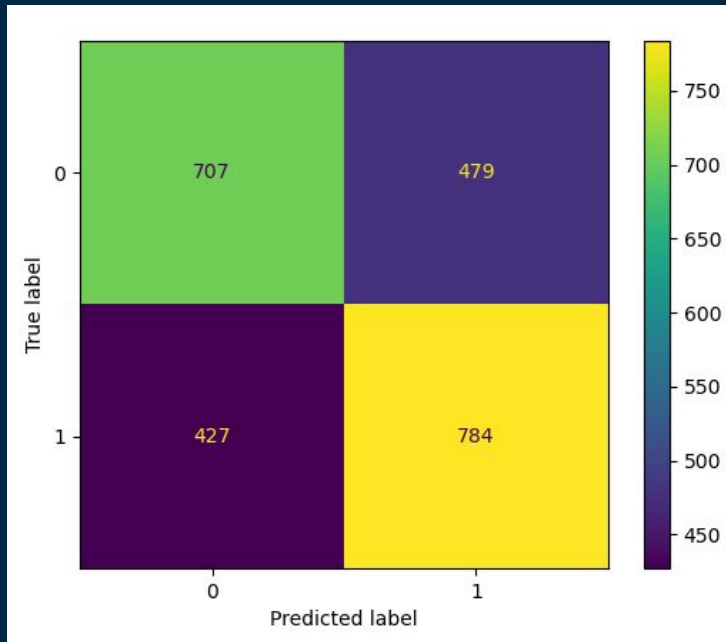| Word	Number of Occurances
relationship	31
think	27
friend	25
time	22
one	19
ever	18
advice	17
female	16
year	16
what's	16
dating	13
way	13
best	12
first	12
thing	12
love	12
old	11
need	11
sex	11
back	11

Misclassified AskMen Top Words

Word	Number of Occurances
partner	29
reddit	25
life	24
friend	24
someone	23
relationship	22
advice	18
say	17
think	17
best	17
dating	15
ex	15
sex	13
actually	12
deal	12
move	12
really	11
way	11
give	11
find	11

Misclassified AskWomen Top Bigrams

Bigram	Number of Occurances
need advice	5
best friend	5
first date	4
year old	4
social medium	3
first time	3
back time	3
year ago	3
time life	2
female bos	2
way approached	2
helping son	2
advice help	2
way helping	2
keep putting	2

MisclassifiedAskMen Top Bigrams

Bigram	Number of Occurances
year old	5
wi fi	3
shave body	3
best friend	3
advice give	3
anyone else	3
long term	3
male friend	3
side relationship	2
react partner	2
watch view	2
fictional character	2
dating apps	2
best advice	2
new gf	2

CONCLUSIONS & RECOMMENDATIONS

My Model

My Problem Statement

Future Research

THANKS

Do you have any questions?