

# AUTISM IN WOMEN

---

A look at Reddit forums through  
Natural Language Processing

# INTRODUCTION

Terminology: autism, aspergers, neurodiversity

Autism in women and girls

Goal of this project

# PROJECT WORKFLOW

01



Data Collection  
and Cleaning

02



Text  
Preprocessing

03



Topic  
Modeling

04



Sentiment  
Analysis

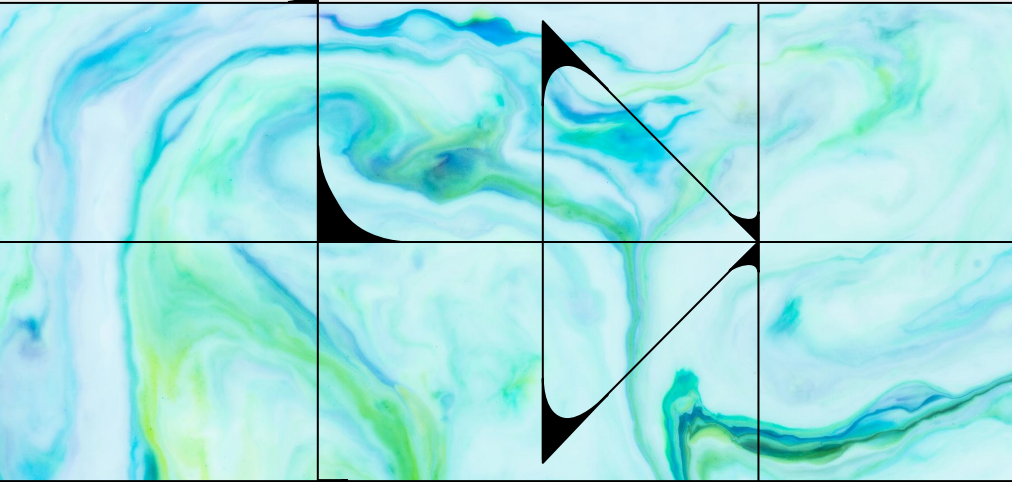
05



Recommender  
System

# REDDIT FORUMS

13k posts scraped using PRAW



**female specific:**

r/AutismInWomen

r/aspergirls

**general:**

r/autism

r/aspergers

# DATA CLEANING AND TEXT PROCESSING

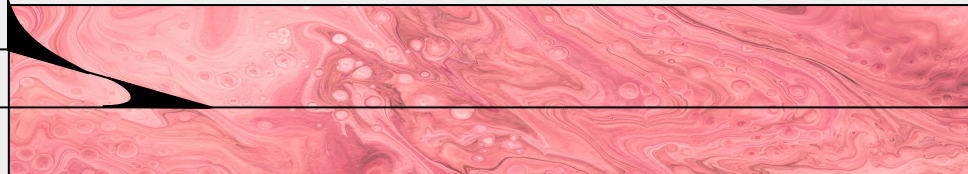
## NLTK and Regex

- Emojis
- Web links
- Punctuation
- Newline and other formatting characters
- Stop words



## SpaCy

- POS tagging
- Lemmatization



# ANALYSIS OF TOP WORDS

Nouns that were in the top 200 terms of the female-specific forums,  
but NOT in the general forums:

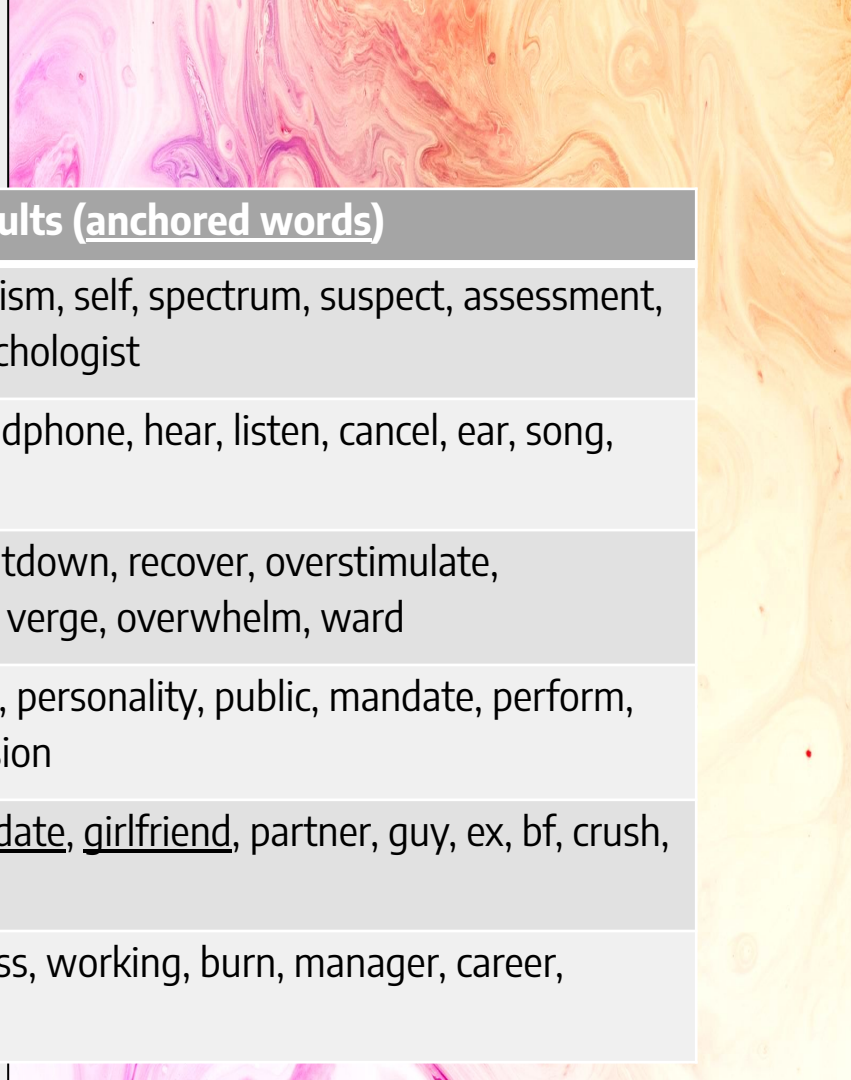
'husband', 'burnout', 'thank', 'mask', 'trauma', 'office', 'masking', 'meeting', 'hell', 'skin', 'overload',  
'adhd', 'cat', 'routine', 'career', 'deal', 'friendship', 'period', 'store', 'daughter', 'tone', 'evaluation', 'boss',  
'comfort', 'weird', 'manager', 'title'

**Subjects: trauma, masking, burnout/overload,  
work/career, relationships**



<h1>TOPIC</h1>	<h1>MODELING</h1>
<p>LSA NMF LDA</p> <p><b>CorEx</b> anchored topic modeling</p>	<p>Restricted POS: nouns verbs adjectives</p> <p>Remove common words</p>

# COREX TOPIC ANCHORS



Topic Interpretation	CorEx Topic Word Results ( <u>anchored words</u> )
Autism Assessment and Diagnosis	<u>diagnose</u> , <u>diagnosis</u> , autism, self, spectrum, suspect, assessment, symptom, research, psychologist
Sound Sensitivity	<u>sound</u> , <u>noise</u> , <u>music</u> , headphone, hear, listen, cancel, ear, song, voice
Meltdowns and Burnout	<u>meltdown</u> , <u>burnout</u> , shutdown, recover, overstimulate, overstimulation, induce, verge, overwhelm, ward
Masking	<u>mask</u> , <u>masking</u> , unmask, personality, public, mandate, perform, persona, pandemic, version
Relationships	<u>relationship</u> , <u>boyfriend</u> , <u>date</u> , <u>girlfriend</u> , partner, guy, ex, bf, crush, sex
Work Life	<u>work</u> , <u>job</u> , <u>interview</u> , boss, working, burn, manager, career, coworker, office



Frequency



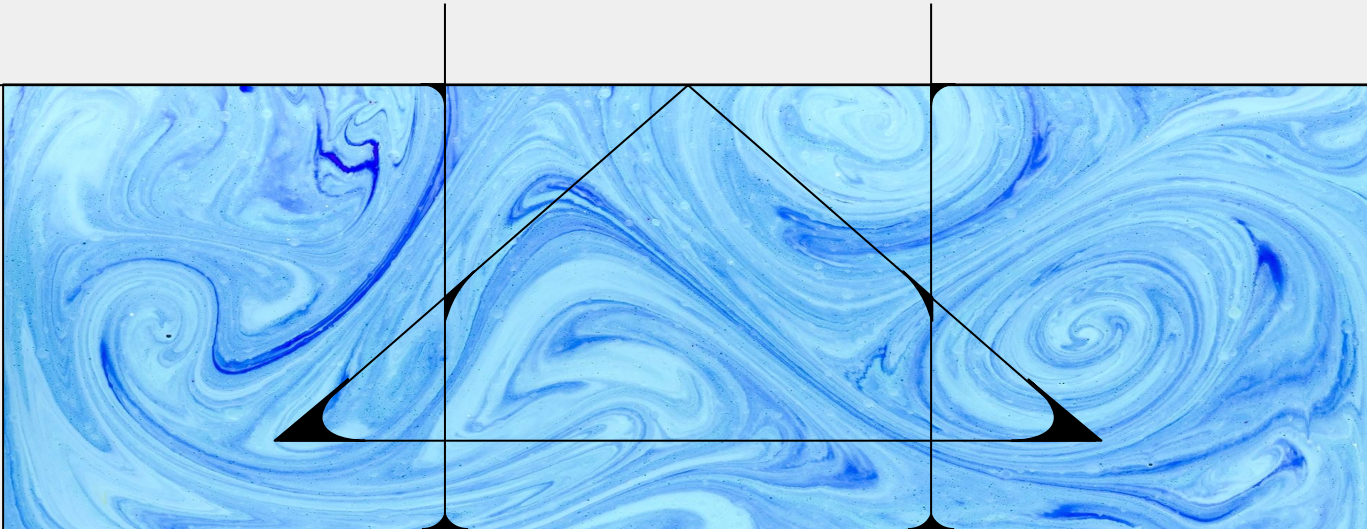
general Frequency  
|  
Frequent

Characteristic  
meltdown  
neurotypical  
clothe  
coworker  
trait  
stim  
stimme  
hyperfixation  
overstimulation  
hobbie  
adhd  
burnout  
idk  
empathy  
psychiatrist  
asd  
stims  
aspergers  
covid  
symptom  
ableism  
earplug  
therapist  
neurodiversity  
anxiety  
nts  
raads  
diagnosis  
emotion  
youtube

# SENTIMENT ANALYSIS

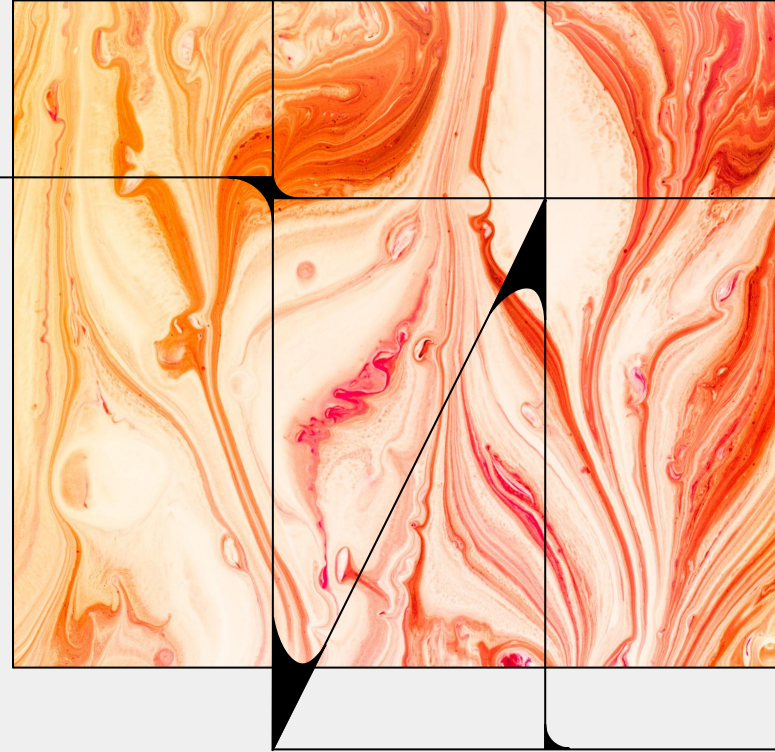
VADER

Posts had high neutrality score, resulting in extremes for the compound score



# RECOMMENDER SYSTEM

- Search by keywords or post title
- Return most relevant Reddit post based on Cosine Similarity



# FUTURE WORK

- Sentiment Analysis using TextBlob, DeepMoji, or something not trained on review data
- Determine the topic distribution for each post using CorEx



# APPENDIX

# COREX CORRELATION DISTRIBUTION

