Topic Modeling
w/ Male Mental Health Issues

By: Edward Mendoza

Background

- Male Mental Health has been a recent mainstream buzzword in Psychology Today

- Some even cite male mental health being a "silent epidemic"

- The dire consequences of male mental health may lead to depression, weight gain, and thoughts of suicide

"Since that day [I opened up about my emotions], it's just been so much easier to live and so much easier to enjoy life." – Michael Phelps

Problem Statement

❖ Explore what the main topics of discussion in regards of Male Mental Health on social media– more specifically, within the certain Reddit communities involving male mental health.

The Topic Modeling Process

1

The Topic Modeling Process

Data Gathering and Cleaning — Source of Data - Reddit

EDA — Exploratory Data Analysis including:
Sentimental Analysis, Bi-Gram\Tri-Gram Analysis

Modeling — Latent Dirichlet Allocation

Interpretation — Top 5 Main Topics Discovered

Data Gathering and Cleaning

- **Subreddits:**

- r/askmen

- r/malementalhealth

- r/bropill

- r/mensLib

- How data was cleaned:

  - Lowercase

  - Removed words longer than 25 characters and shorter than 3 characters long

  - Removed Autobots

  - Hyperlinks, punctuation marks and urls were removed

  - All words were lemmatized and stemmatized

Exploratory Data Analysis – 10 most common words

Ten Most Common Words in the Dataset
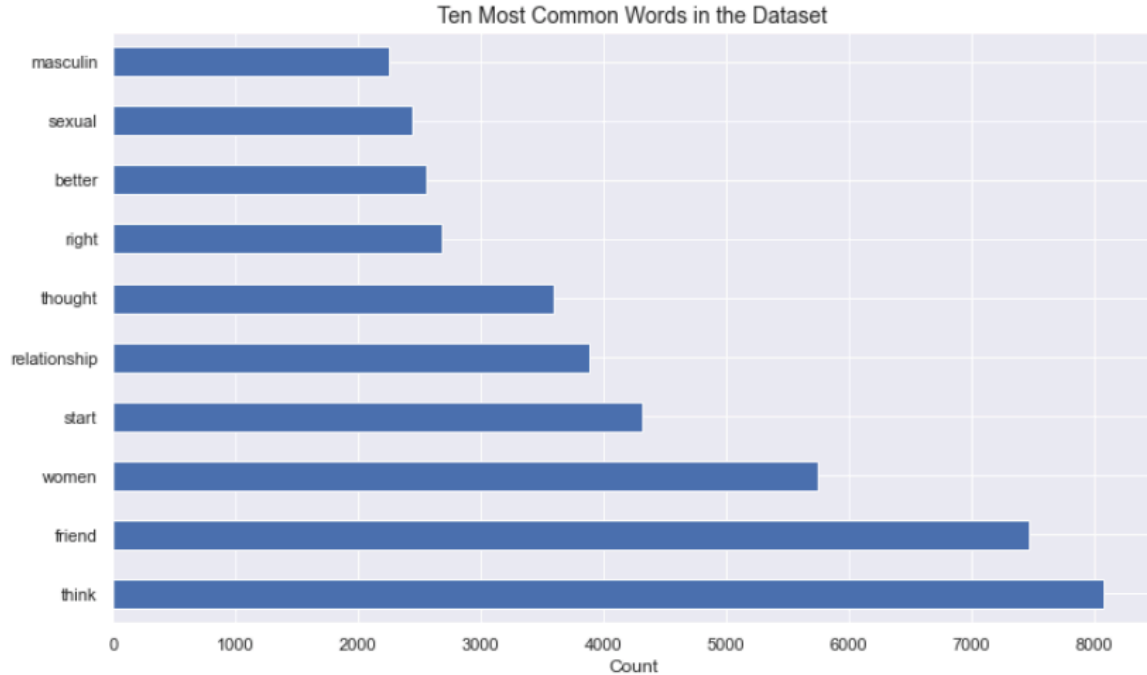
Exploratory Data Analysis – Sentiment Analysis

Post Count Categorized as Positive or Negative
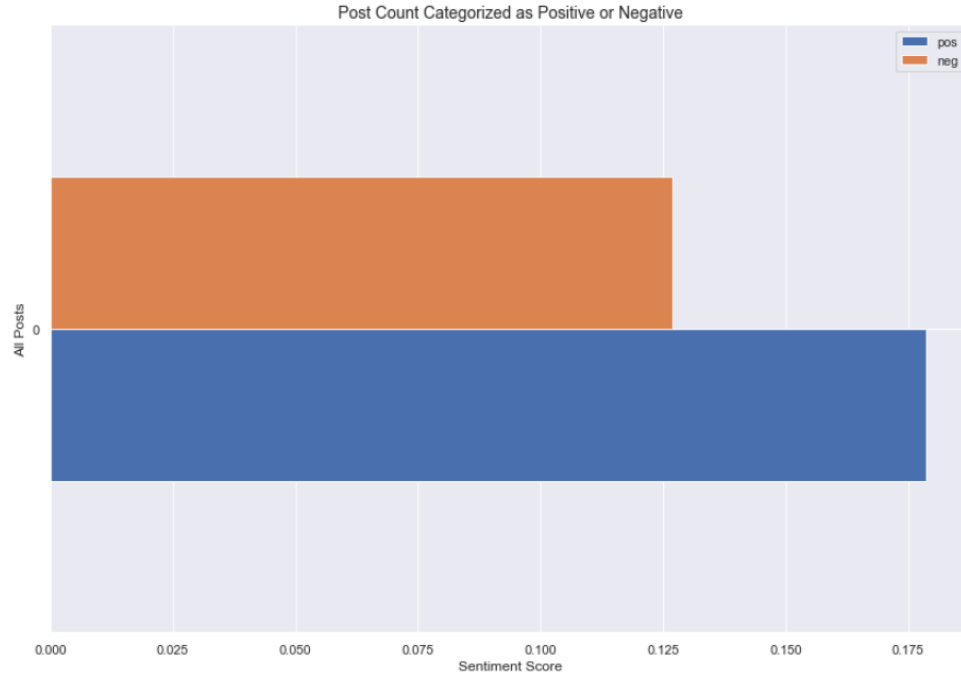
Most posts were categorized as positive rather than negative

This is best explained by the subreddit community advocating a supportive environment

EDA – Comparing Most Frequent Words Between Positive and Negative Scores

Words Most Frequent with Positive Score

Words Most Frequent with Negative Score
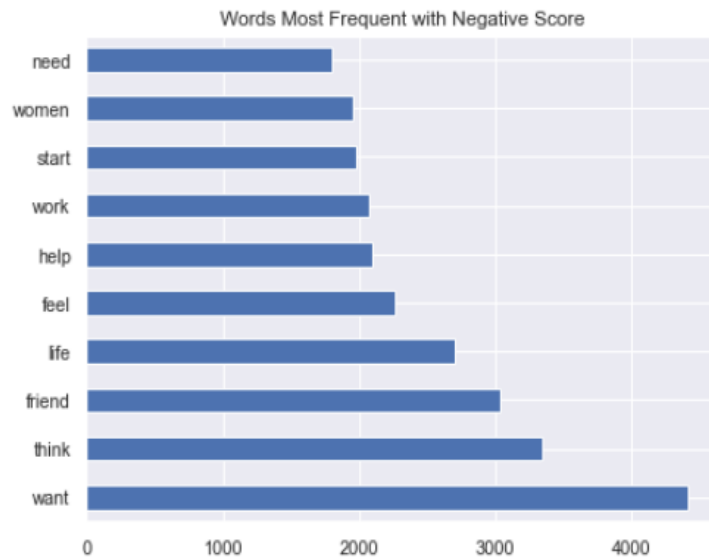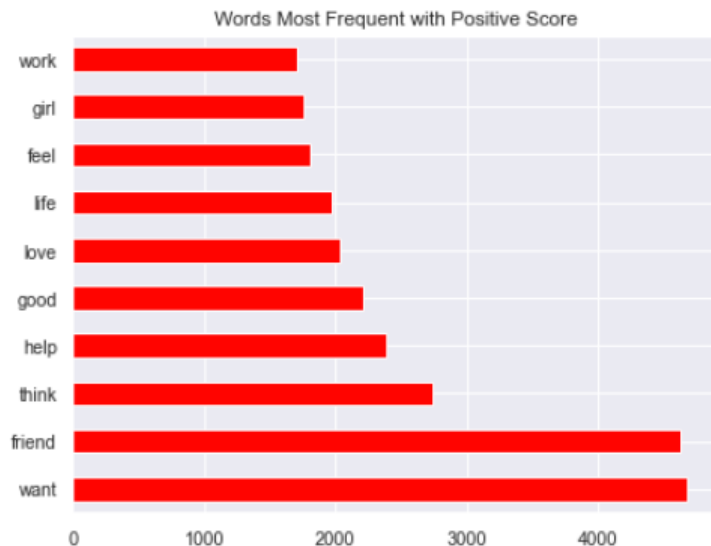
EDA – Bi-Gram Chart of Most Frequent Words
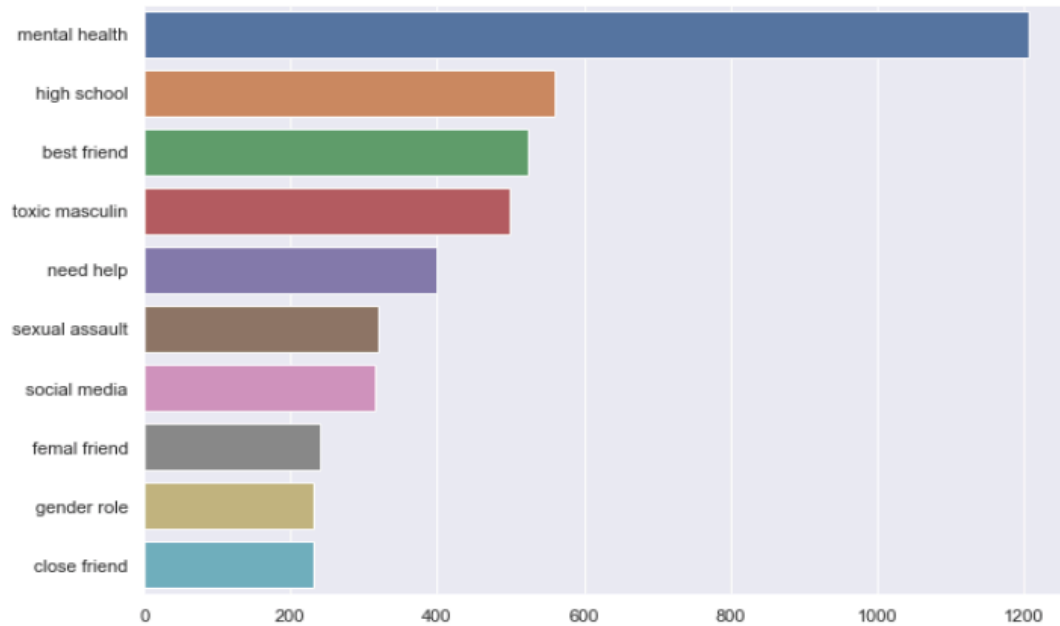
EDA – Tri-Gram Chart of Most Frequent Words
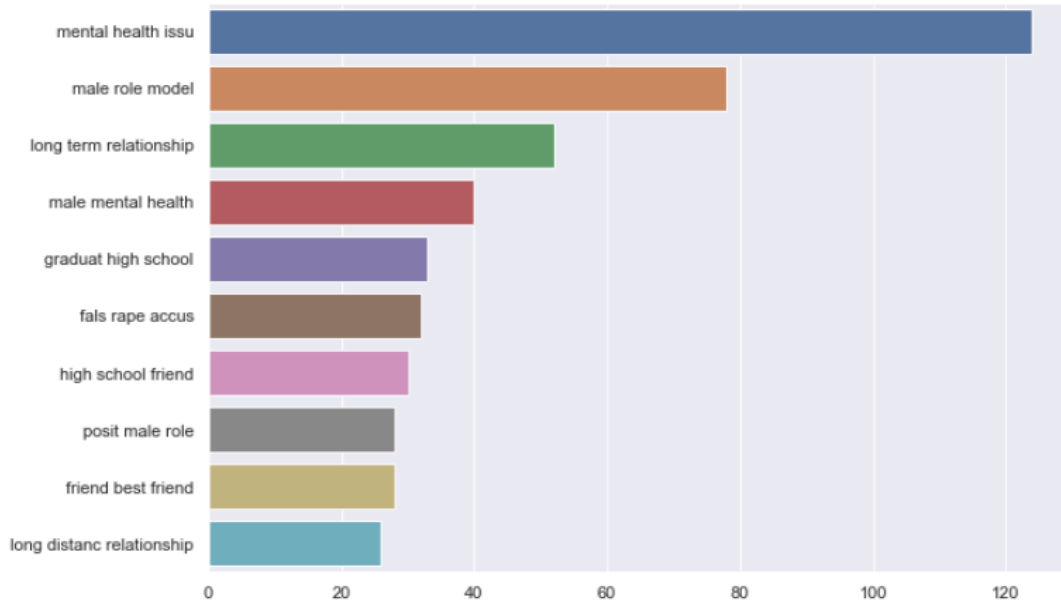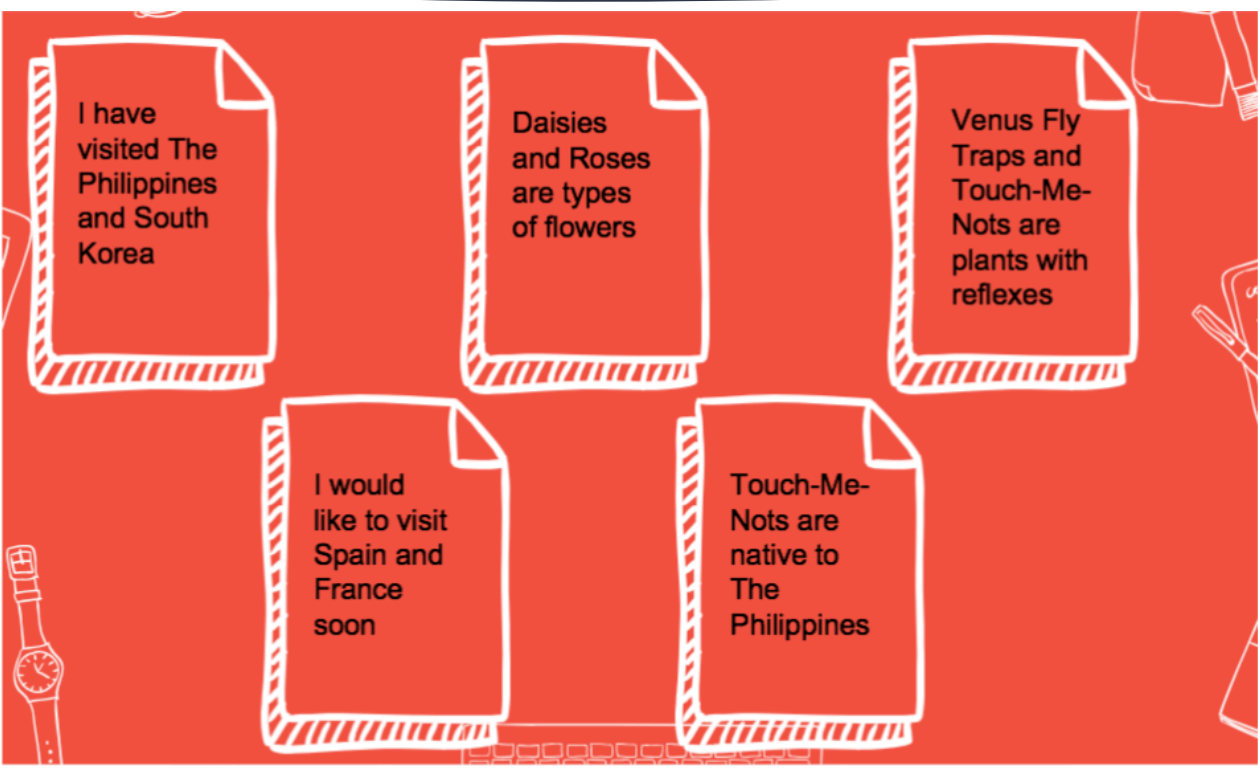
Mental Health, positive male role models, false rape accusations, and long-distance relationships are at top of mind for most men posting in these subreddits

Topic Modeling

Via Latent Dirichlet Allocation

I have visited The Philippines and South Korea

Daisies and Roses are types of flowers

Venus Fly Traps and Touch-Me-Nots are plants with reflexes

I would like to visit Spain and France soon

Touch-Me-Nots are native to The Philippines

Sample Corpus
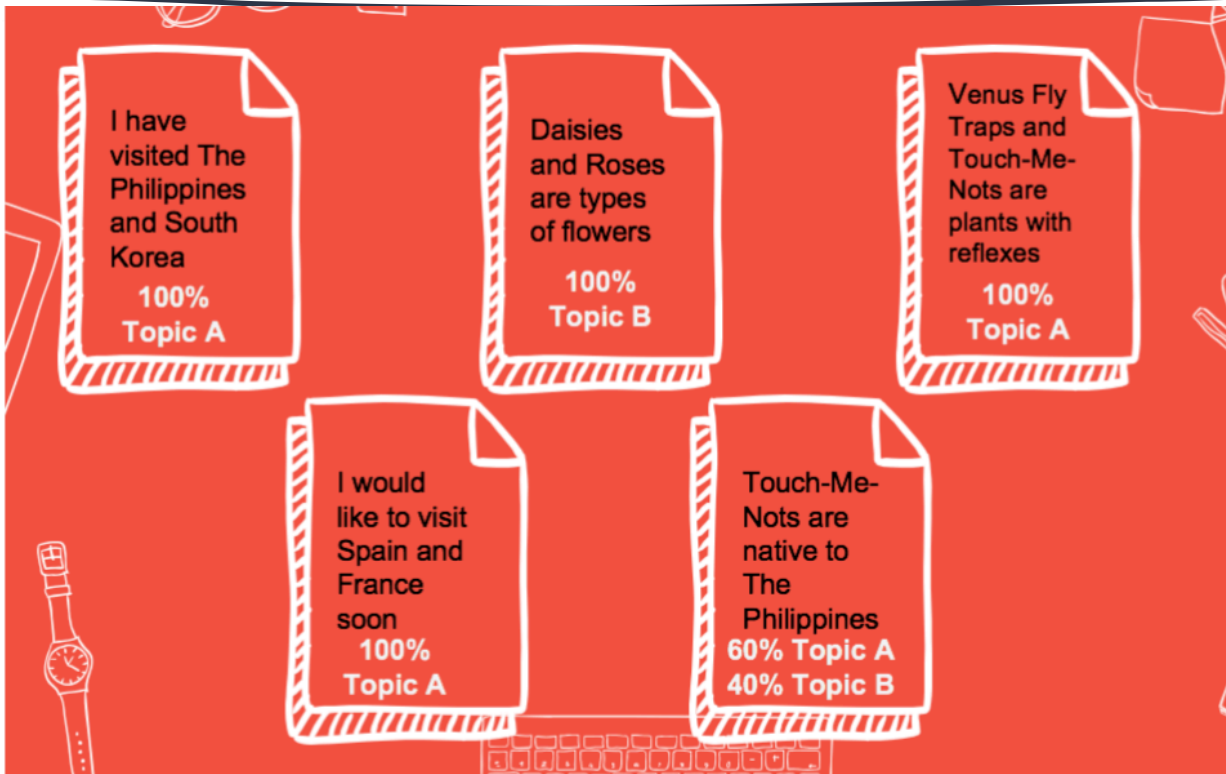
LDA Applied to Documents

Topic A: 40% visit, 20% Philippines, 20% France...
Countries

Topic B: 20% Daisies, 20% flowers, 15% Roses..
Plants

Every Document is a distribution of **topics**

Every Topic consists of a distribution of **words**

Topic Breakdown

Document-Term Matrix

A mathematical matrix that displays the frequency of of terms that occur in a collection of documents.
Each row represents one document
Each column represents one term (word)
Each value typically contains the number of appearances of that term in that document

	adul	adult	adultfriendfind	adulthood	adulti	adultsgirlsfemalesi	adultsthough	advanc	advantag	advantagesbut	advantagesdisadvantag	advent	adventist
5	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	1	0	0	0	0	0

Latent Distribution Allocation

- Choose the number of topics you believe are in the corpus (k=2)

- Randomly assign each word in each document to one of 2 topics.

- Go through every word and its topic assignment in each document. Look at how often the topic occurs in the document and how often the word occurs in the topic overall.

- Go through multiple iterations of this. It will make sense!

Topic Modeling via LDA – All Words

```
lda = models.LdaModel(corpus=corpus, id2word=id2word, num_topics=2, passes=10)
lda.print_topics()
```

- 1st attempt

```
[(0,
 '0.005*"autoerotic" + 0.004*"catastoph" + 0.004*"coattail" + 0.004*"ablat" + 0.004*"battlefront" + 0.003*"chequ" +
0.003*"coars" + 0.003*"child" + 0.003*"afghanistan" + 0.003*"bigfram"'),
 (1,
 '0.003*"gail" + 0.003*"bugaloo" + 0.002*"gona" + 0.002*"hatefil" + 0.002*"burr" + 0.002*"girl" + 0.001*"bezo" + 0.0
01*"garth" + 0.001*"electromagnet" + 0.001*"growingup"')]
```

- 2nd attempt

```
[(0,
 '0.008*"coars" + 0.006*"battlefront" + 0.003*"gamedev" + 0.003*"haphazardli" + 0.003*"gail" + 0.003*"finsta" + 0.00
2*"gona" + 0.002*"excusedelus" + 0.002*"dwarfism" + 0.002*"garth"'),
 (1,
 '0.011*"fancentro" + 0.007*"banish" + 0.005*"burr" + 0.004*"chin" + 0.004*"argumentfight" + 0.003*"goddamn" + 0.003
*"ccvictimsskycom" + 0.003*"monthsold" + 0.003*"betray" + 0.003*"beholden"'),
 (2,
 '0.005*"bugaloo" + 0.004*"gail" + 0.002*"hatefil" + 0.002*"finest" + 0.002*"electromagnet" + 0.002*"homework" + 0.0
02*"girl" + 0.002*"friendso" + 0.002*"foolhardi" + 0.002*"burr"'),
 (3,
 '0.011*"bezo" + 0.005*"comichttpimgurcomaxm" + 0.004*"betasi" + 0.004*"centauri" + 0.004*"cuddli" + 0.004*"bloodbat
h" + 0.003*"fatigu" + 0.003*"afsp" + 0.003*"growingup" + 0.003*"bathroom"'),
 (4,
 '0.006*"autoerotic" + 0.005*"catastoph" + 0.005*"coattail" + 0.005*"ablat" + 0.004*"chequ" + 0.004*"child" + 0.004
*"afghanistan" + 0.003*"cismenmen" + 0.003*"bigfram" + 0.003*"blackpil"')]
```

Topic Modeling Via LDA – Nouns Only

```
[(0,
  '0.053*"girl" + 0.053*"friend" + 0.051*"date" + 0.025*"relationship" + 0.013*"look" + 0.012*"girlfriend" + 0.010*"s
tart" + 0.009*"text" + 0.008*"work" + 0.007*"meet"'),
 (1,
  '0.078*"women" + 0.031*"woman" + 0.013*"masculin" + 0.010*"male" + 0.010*"gender" + 0.008*"gener" + 0.008*"issu" +
0.007*"discuss" + 0.007*"attract" + 0.006*"commun"'),
 (2,
  '0.022*"help" + 0.018*"parent" + 0.012*"hair" + 0.009*"bodi" + 0.007*"money" + 0.007*"father" + 0.007*"mother" + 0.
007*"famili" + 0.006*"deal" + 0.006*"need"'),
 (3,
  '0.029*"life" + 0.026*"work" + 0.013*"relationship" + 0.012*"help" + 0.010*"love" + 0.009*"start" + 0.009*"school"
+ 0.007*"point" + 0.007*"chang" + 0.007*"care"'),
 (4,
  '0.053*"delet" + 0.014*"masturb" + 0.014*"health" + 0.012*"help" + 0.012*"porn" + 0.012*"watch" + 0.010*"studi" +
0.009*"peni" + 0.008*"video" + 0.007*"research"')]
```

Topic Modeling Via LDA – Nouns and Adjectives

```
[(0,
  '0.015*"partner" + 0.012*"health" + 0.010*"help" + 0.010*"support" + 0.010*"issu" + 0.010*"male" + 0.009*"rape" +
0.007*"victim" + 0.006*"assault" + 0.006*"research"'),
 (1,
  '0.082*"women" + 0.034*"woman" + 0.018*"masculin" + 0.013*"date" + 0.013*"attract" + 0.012*"gender" + 0.012*"male"
+ 0.008*"gener" + 0.007*"toxic" + 0.006*"group"'),
 (2,
  '0.025*"delet" + 0.019*"girlfriend" + 0.012*"wife" + 0.010*"mother" + 0.010*"watch" + 0.010*"father" + 0.009*"marr
i" + 0.009*"parent" + 0.009*"brother" + 0.008*"cheat"'),
 (3,
  '0.024*"friend" + 0.023*"work" + 0.022*"girl" + 0.022*"relationship" + 0.020*"date" + 0.015*"help" + 0.011*"start"
+ 0.010*"school" + 0.007*"problem" + 0.007*"point"'),
 (4,
  '0.019*"bodi" + 0.018*"hair" + 0.010*"dick" + 0.008*"watch" + 0.008*"love" + 0.007*"favorit" + 0.007*"size" + 0.007
*"face" + 0.007*"posit" + 0.006*"peni"')]
```

I interpreted the above topics as follows:
1. Sexual Assault/Trauma
2. Gender Identity/Roles
3. Familial Issues
4. Dating Advice/Relationship Issues
5. Body Issues/Positivity

Recommendations for Talkspace

- Hire talented behavioral therapists who specialize and have a deep understanding with the topics found
- Provide training to already employed behavioral therapists who can assist with the aforementioned topics
- Ensure male clients that we have staff that can help with any mental health issues - no matter how frivolous they may be

Next Steps

To Improve on my analysis, I would:

X Include analysis of other metrics towards certain male demographics

X Explore Other Topic Modeling Techniques (LSA, PLSA, and lda2vec)

X Topic Generation App

Thank you for your time!