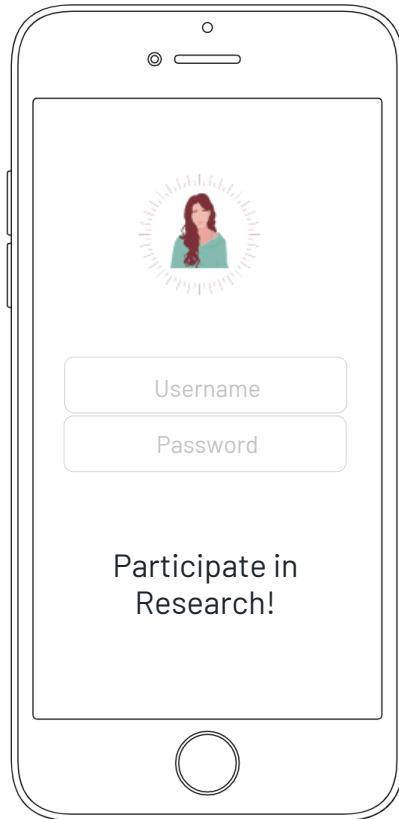

Topic Modeling to Identify Women's Health Concerns from Online Forums

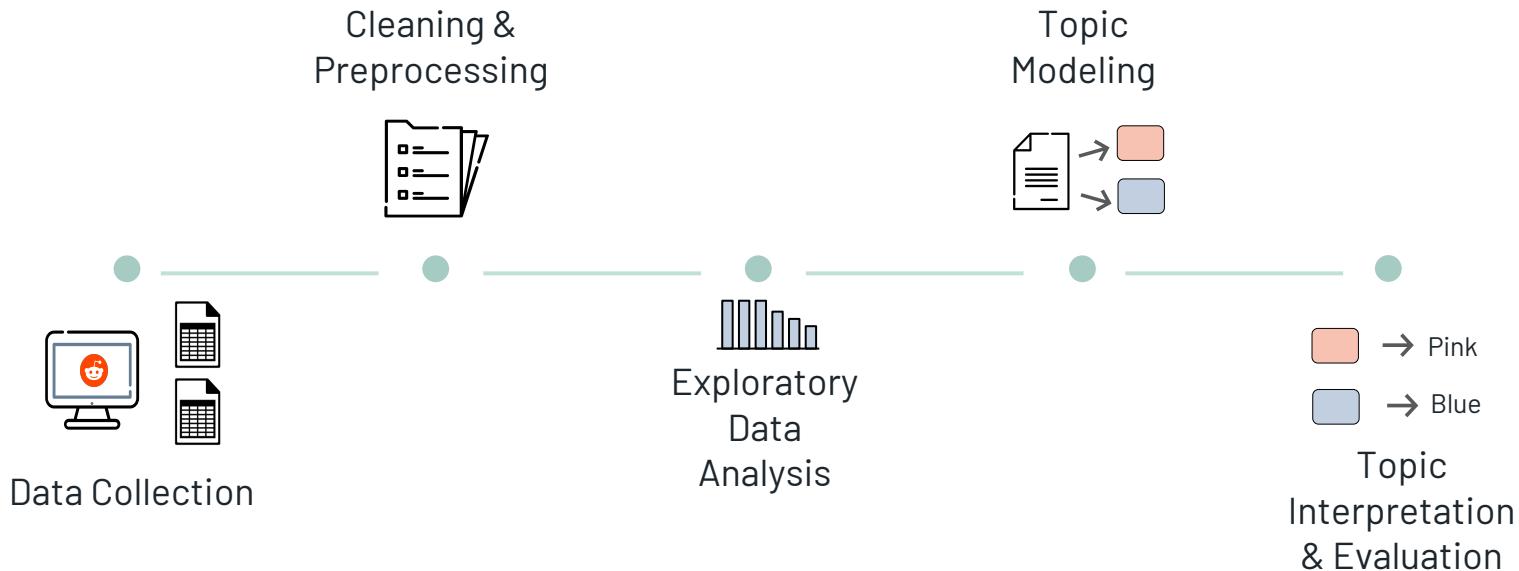
Jocelyn Lutes



Project Background

1. Citizen Science
 2. Maternal and Infant Health Research
 - Current Procedures
 - Mobile Application
 3. Product Development for Incentives
- CURRENT GOAL:
Identify women's health concerns from posts in online forums

Work Flow



Data



General Women's Health Corpus

Subreddits:

- r/WomensHealth
- r/obgyn
- r/TheGirlSurvivalGuide

Number of Posts: 30,616

Unique Authors: 19,753



Fertility and Pregnancy Corpus

Subreddits:

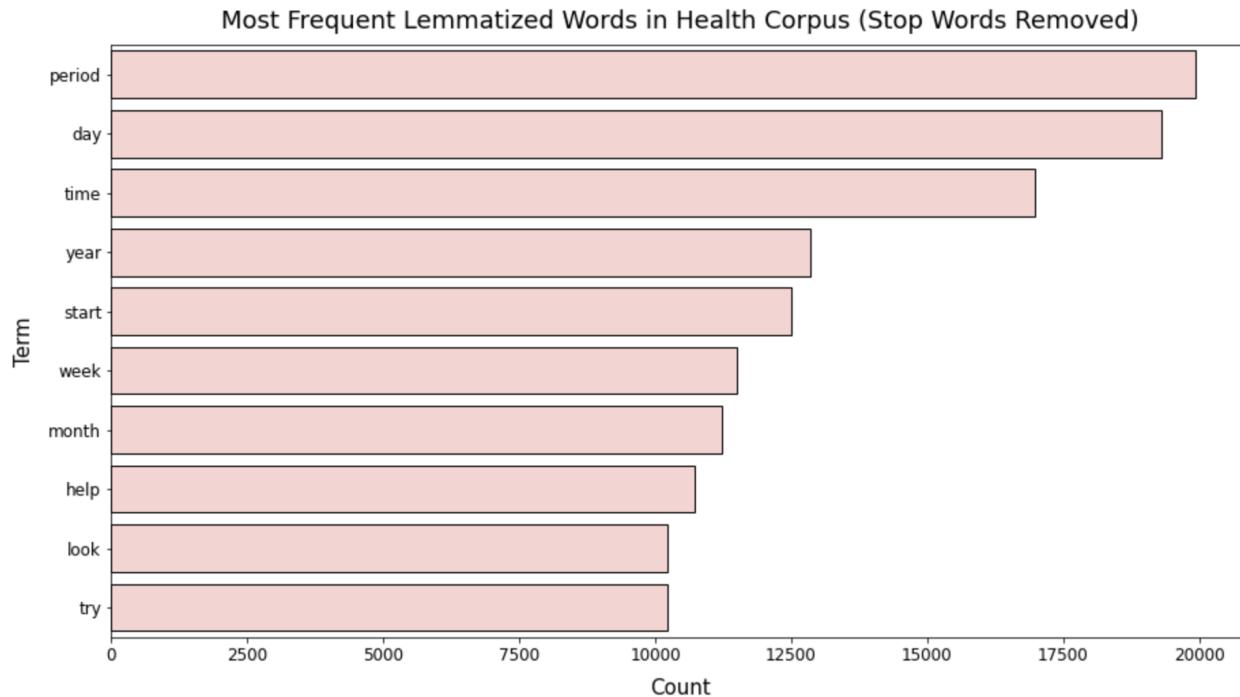
- r/TryingForABaby
- r/pregnant
- r/BabyBumps

Number of Posts: 92,943

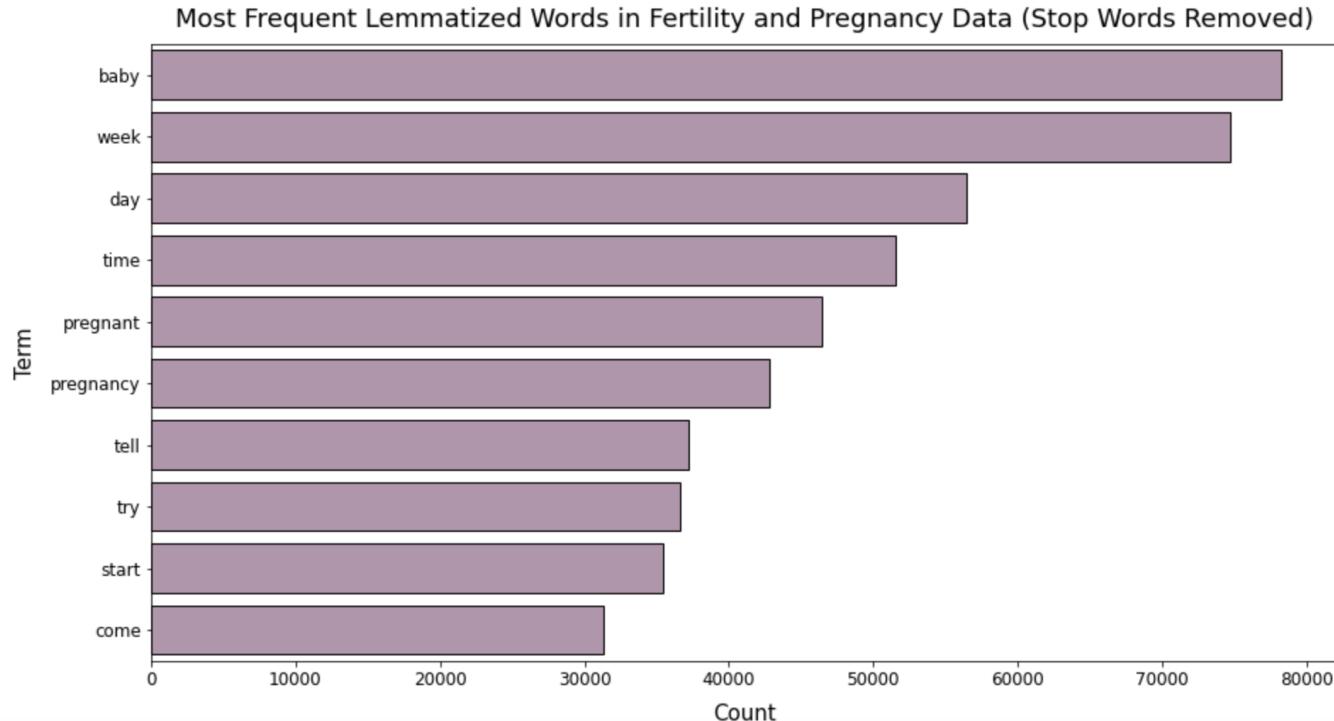
Unique Authors: 35,127

Prior to modeling, the text was cleaned, lemmatized, and vectorized using TF-IDF.

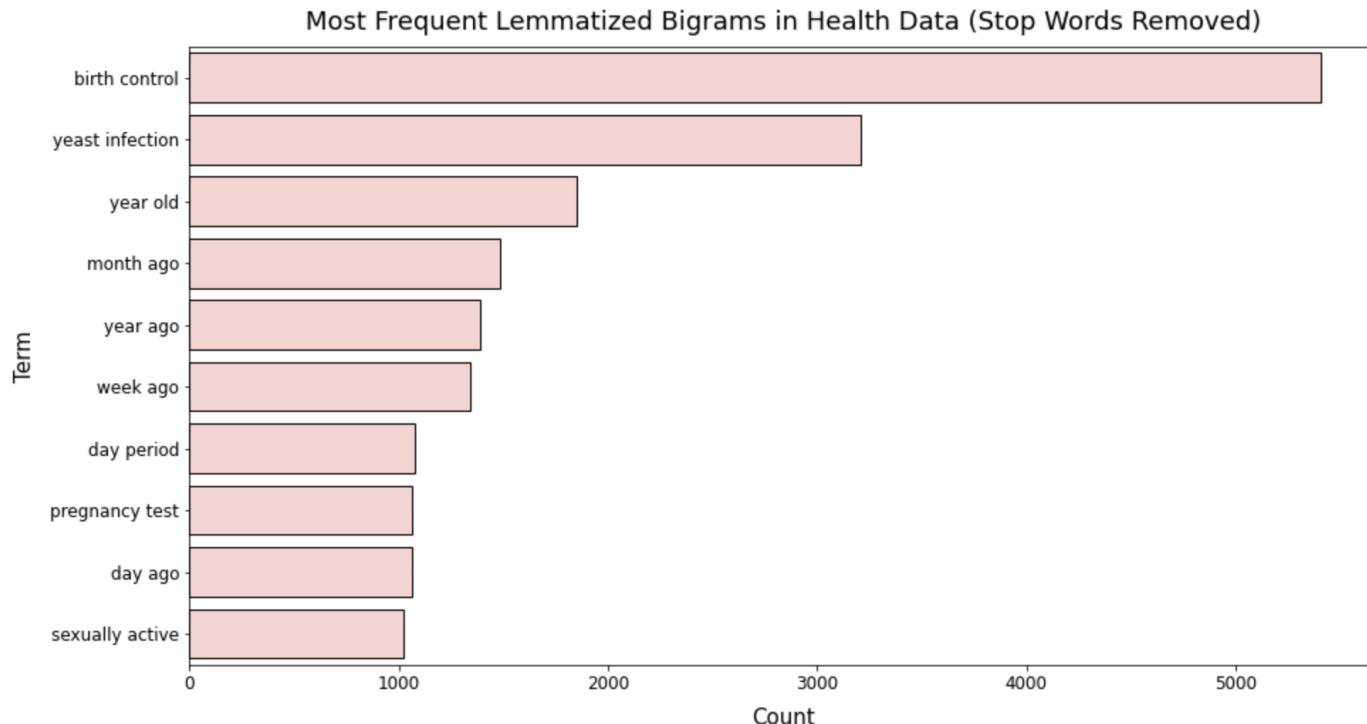
Most Frequent Words by Corpus



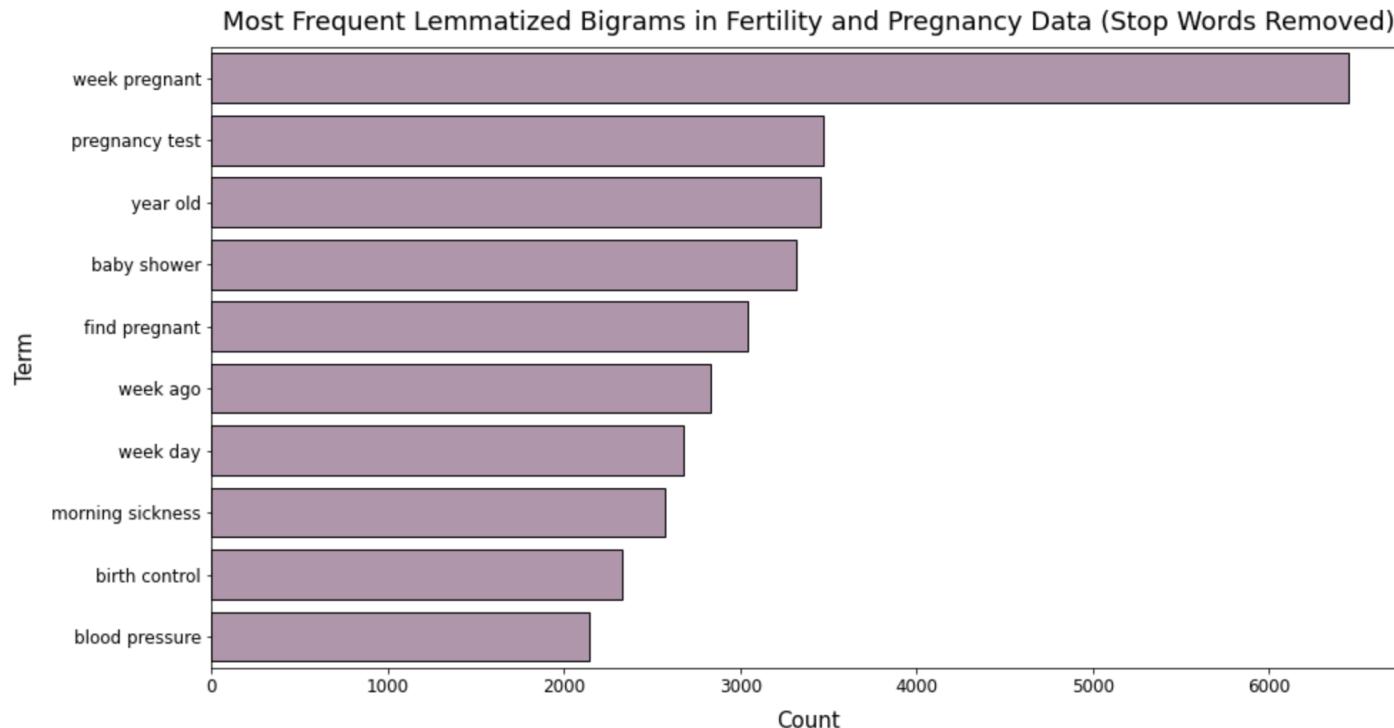
Most Frequent Words by Corpus



Most Frequent Bigrams by Corpus

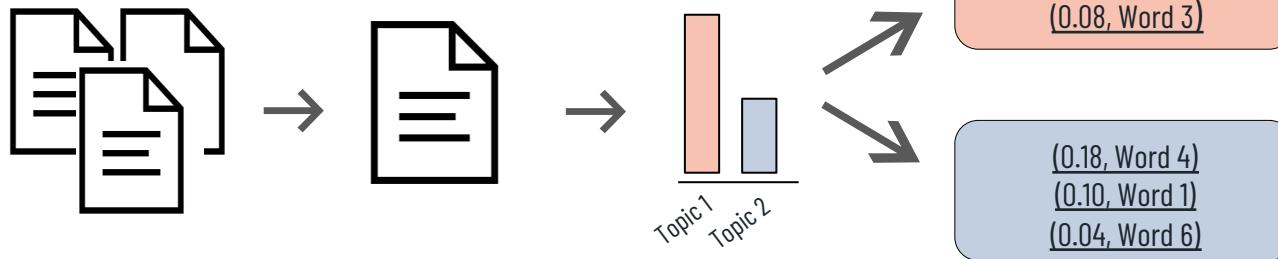


Most Frequent Bigrams by Corpus



Latent Dirichlet Allocation (LDA)

Assumptions



Model Process

1. Select Number of Topics (K)
2. For each document, randomly assign a word to topic (K topics total)
3. Repeat word assignment based on $p(\text{topic} | \text{doc})$ and $p(\text{word} | \text{topic})$

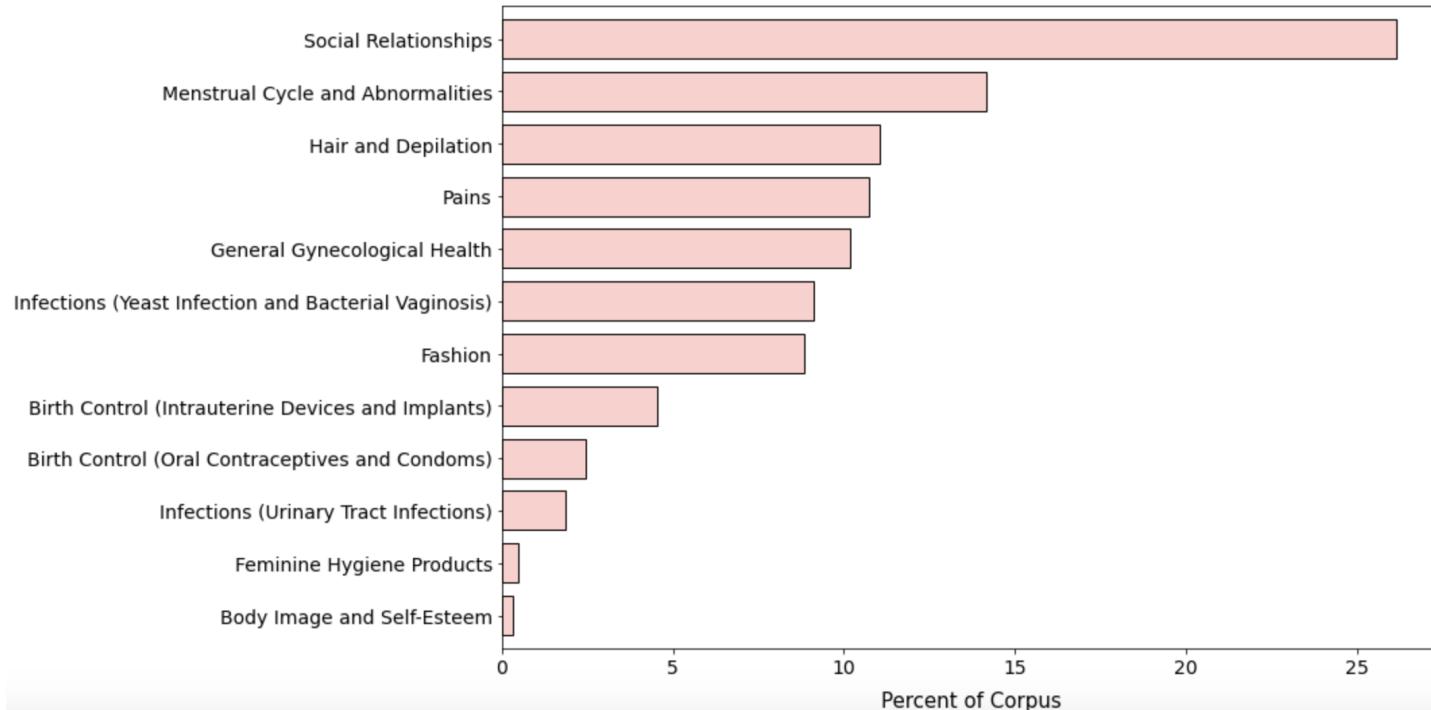
Topics Identified in General Women's Health Corpus

Topic Name	Top 5 Terms
Birth Control (IUDs and Implants)	IUD (intrauterine device), control, birth, birth control, pill
Birth Control (Pills and Condoms)	Pill, birth control, birth, control, pack
Infections (YI and BV)	Infection, yeast, yeast infection, discharge, smell
Infections (UTI)	UTI (urinary tract infection), pee, drink, antibiotic, orgasm
Menstrual Cycle & Abnormalities	Period, bleed, start, test, pregnancy
Feminine Hygiene Products	Tampon, cup, pad, menstrual, menstrual cup
General Gynecological Health	Vagina, breast, sex, pap, hurt
Pains	Pain, period, cyst, cramp, doctor
Body Image and Self-Esteem	Eat, food, self esteem, esteem, diet

Model: LDA (12 topics, 50 iterations); UMass Coherence: -2.3

Topics Identified in General Women's Health Corpus

Percentage of Posts in Women's Health Corpus by Topic



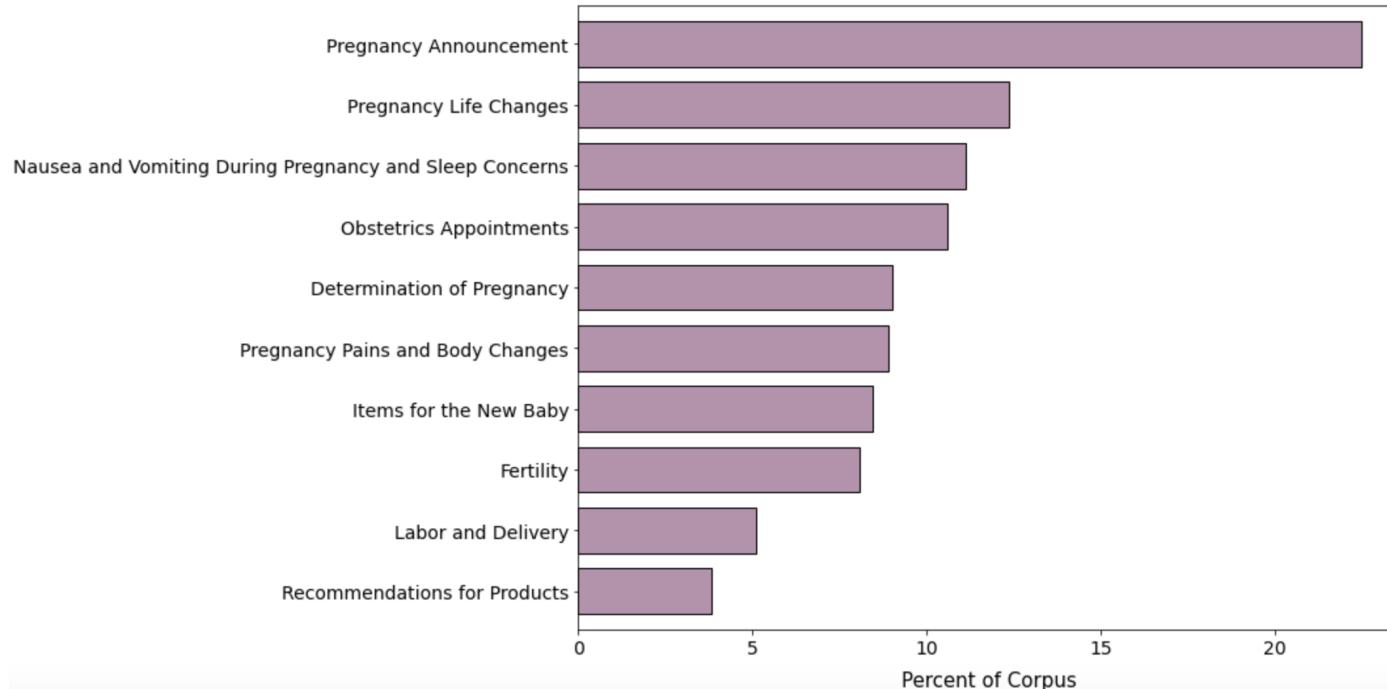
Topics Identified in Fertility and Pregnancy Corpus

Topic Name	Top 5 Terms
Pregnancy Life Changes	Birth, work, baby, hospital, leave
Pregnancy Announcement	Baby, pregnant, tell, husband, leave
Labor and Delivery	Contraction, labor, cm, induction, induce
Nausea and Sleep	Eat, sleep, food, nausea, morning
Determination of Pregnancy	Test, period, pregnancy, pregnant, positive
Fertility	Cycle, ovulation, TTC , try, ovulate
Obstetrics Appointment	Ultrasound, doctor, blood, baby, appointment
Pains and Body Changes	Pain, belly, baby, weight, kick

Model: LDA (10 topics, 50 iterations); Umass Coherence: -1.96

Topics Identified in Fertility and Pregnancy Corpus

Percentage of Posts in Fertility and Pregnancy Corpus by Topic



Product Recommendations

For Non-Pregnant Users	For Pregnant Users
Menstrual Cycle Abnormalities	Coping with pregnancy life changes
Feminine Hygiene Products	Dealing with nausea and sleep disruptions
Birth Control Options	Obstetrics Appointments
Symptoms of Infections	What to expect during labor and delivery
Common Gynecological Issues	
Fertility Options	
Determination of pregnancy	

Conclusions

1. Topic Modeling with Latent Dirichlet Allocation (LDA) successfully identified several areas of concern in the women's health domain.
2. In comparison to focus groups or surveys, use of data from online forums likely allowed us to analyze data from more unique users.
3. Next Steps:
 - Topic Modeling of Concerns during Postpartum/Early Parenthood
 - Identification of concerns for pre- and post-COVID-19
 - Topic Modeling within previously-identified topics

Thank You!

Do you have any questions?

Email: jocelyn.lutes@gmail.com

LinkedIn: jocelyn.lutes

GitHub: jlu90

CREDITS: This presentation template was created by
Slidesgo, including icons by **Flaticon**, infographics &
images by **Freepik**
