# Ask

Men vs Women

# Ask Men

— — —

1,634,981 subscribers

230 posts per day

**Recent Posts:**

What was the dumbest thing you have done simply because a female was around?

Have you ever been scared for your life? If yes what was the situation I am very interested!

# Ask Women

— — —

1,263,704 subscribers

75 posts per day

**Recent Posts:**

How does it make you feel when someone who don't find attractive has a crush on you?

What is a food that a lot of people like that you think isn't flavortown?

# Data Cleaning

— — —

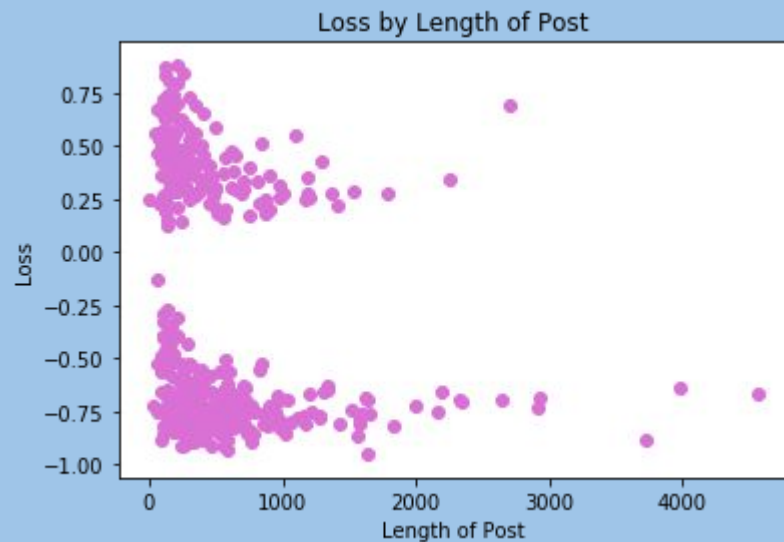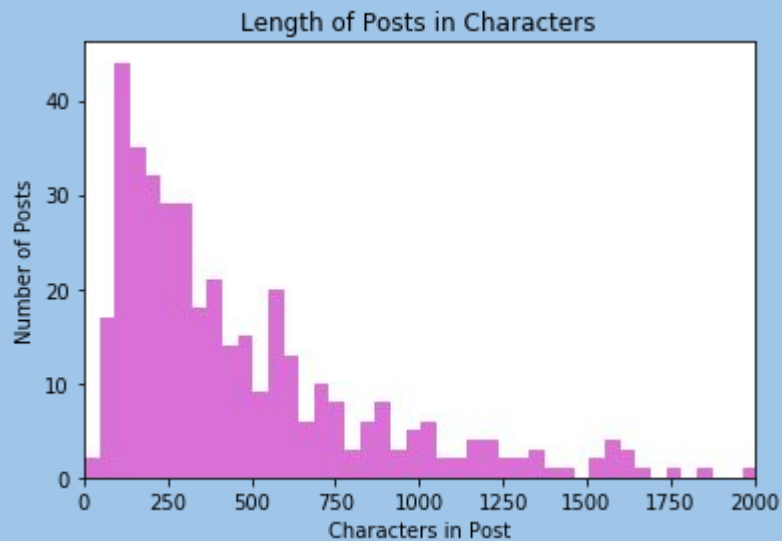Cleaning

Lemmatizing

Vectorization

Removed posts

Titles, Text

# Models Fit

— — —

- Logistic Regression
- Gaussian Bayes
- Multinodal Bayes
- Random Forest
- Support Vector Machine

# Challenges

— — —

### Length of Posts in Characters



### Loss by Length of Post

# Top Words

| Ask Women | |
|---|---|
| subreddit | 740 |
| like | 254 |
| just | 243 |
| very | 180 |
| time | 147 |
| don't | 147 |
| **women** | **133** |
| know | 121 |
| **men** | **121** |
| feel | 120 |

| Ask Men | |
|---|---|
| like | 613 |
| just | 513 |
| very | 362 |
| don't | 362 |
| feel | 340 |
| know | 337 |
| want | 311 |
| really | 310 |
| **women** | **296** |
| time | 275 |

# Topic Modeling using LDA

— — —

Latent Dirichlet Allocation (LDA): represents documents as mixtures of topics that spit out words with certain probabilities

Generated 10 Topics:

Topic 1: Shampoo, product, perfume, cons, what, curiosity, urinal, boxer, red, mothers

Topic 2: League, breaker, fart, russian, muscles, views, intentionally, polish

# Final Model

— — —

**TFID Vectorizer and Logistic Regression**

**Accuracy 68.5%**

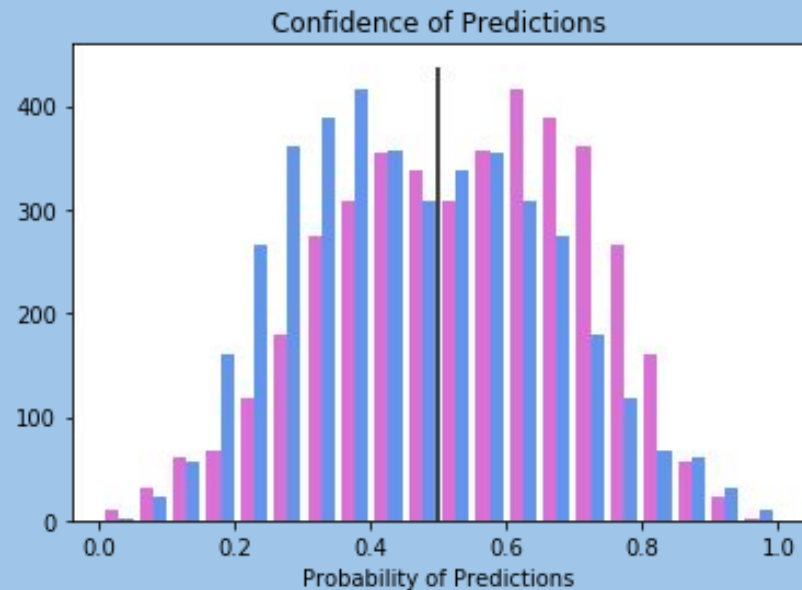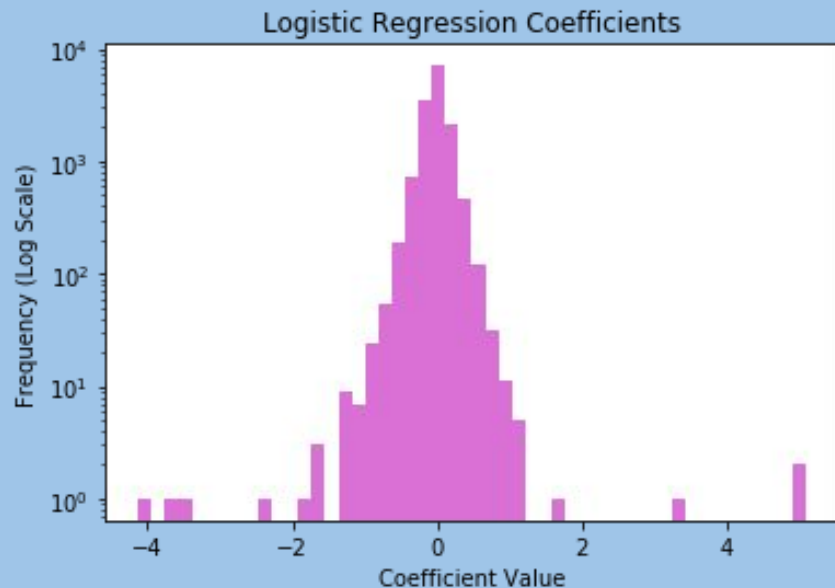Mischaracterized Posts with High Model Confidence:

'What makes a girl marriage material?'

'Would you go to this coffee "date"? 'm 23 and he's 30, and doesn't seem that thrilled about it even though he's the one who asked. I'm afraid I might get it bad since I may really like him whereas he might only be interested in going out with a younger girl to forget about "grown up world" for a while.

How do you deal with finding your girlfriend or wife's hair absolutely everywhere and on everything?

# Result Evaluation

– – –

# Conclusions

———

Maybe men and women aren't so different after all...

**Going Forward:**

Scrape comments matched to posts

Explore possibilities of LDA

# Sources

— — —

http://blog.echen.me/2011/08/22/introduction-to-latent-dirichlet-allocation/

https://towardsdatascience.com/topic-modeling-and-latent-dirichlet-allocation-in-python-9bf156893c24