

NLP Machine Learning within the Community

DSI 1213 PROJECT 3

Presented by
K.Kenneth Davis





K. Kenneth Davis

THE TRANS CAPITALIST LLC

- **Emerging Machine Learning Engineer**
- **LGBTQ+ Community Activist**
- **Consultant**

<https://www.linkedin.com/in/kkennethdavis/>

<https://www.Thetranscapitalist.com>

THIS IS OUR MAIN GOAL.

As transgender resources
for advice increase, how
do queer resources
change?

THE TRANS CAPITALIST LLC
2022

Our Roadmap

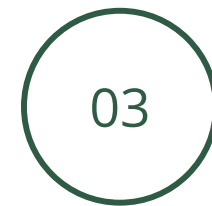
This is my thought process



Collect Subreddits
Data



Clean the
Subreddits



Exploratory Data
Analysis



Model Creation



Conclusion

More data, the better

Our first key is to try



Vote

Posted by u/Kit_Herondale12 **Gay Trans Man, he/him** 51 minutes ago

Looking for advice from people who came out when they were older (30-35+)

I'm a closeted trans guy, 20 years old. I would like some advice regarding social transition from people who came out as trans when they were older - say, past 30.

I live in a very transphobic place. I know that if I come out anytime in the next couple of years, I will find it harder to get a job, my family will not be able to digest it, and it will be



34

Posted by u/bi_dad48 **Bi** 5 days ago

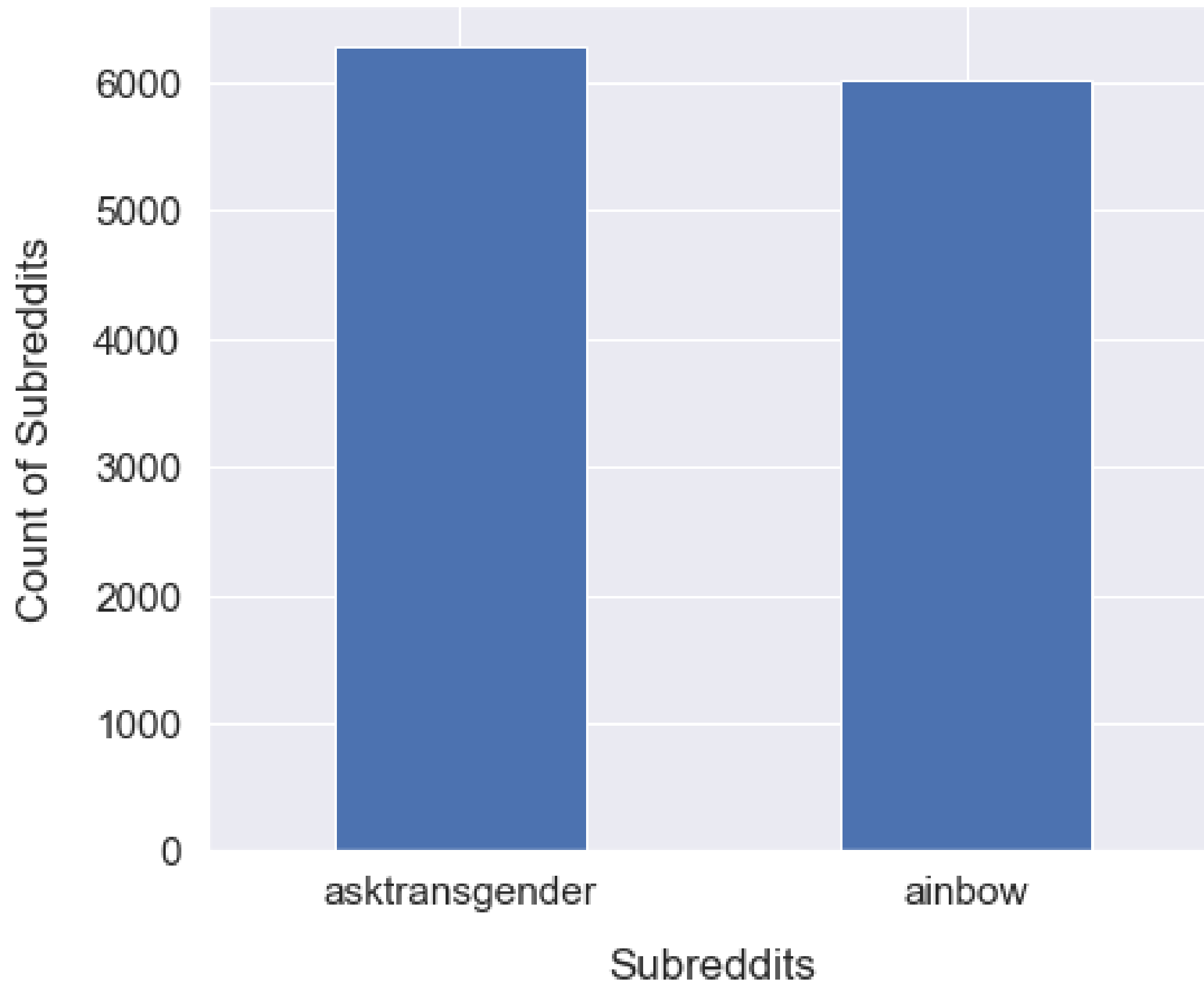
I'm here for you **LGBT Issues**

Hey gang Dad checking in here on Sunday last day of the weekend hoping everybody is doing well. As always remember you are valid you are loved and you are supported.

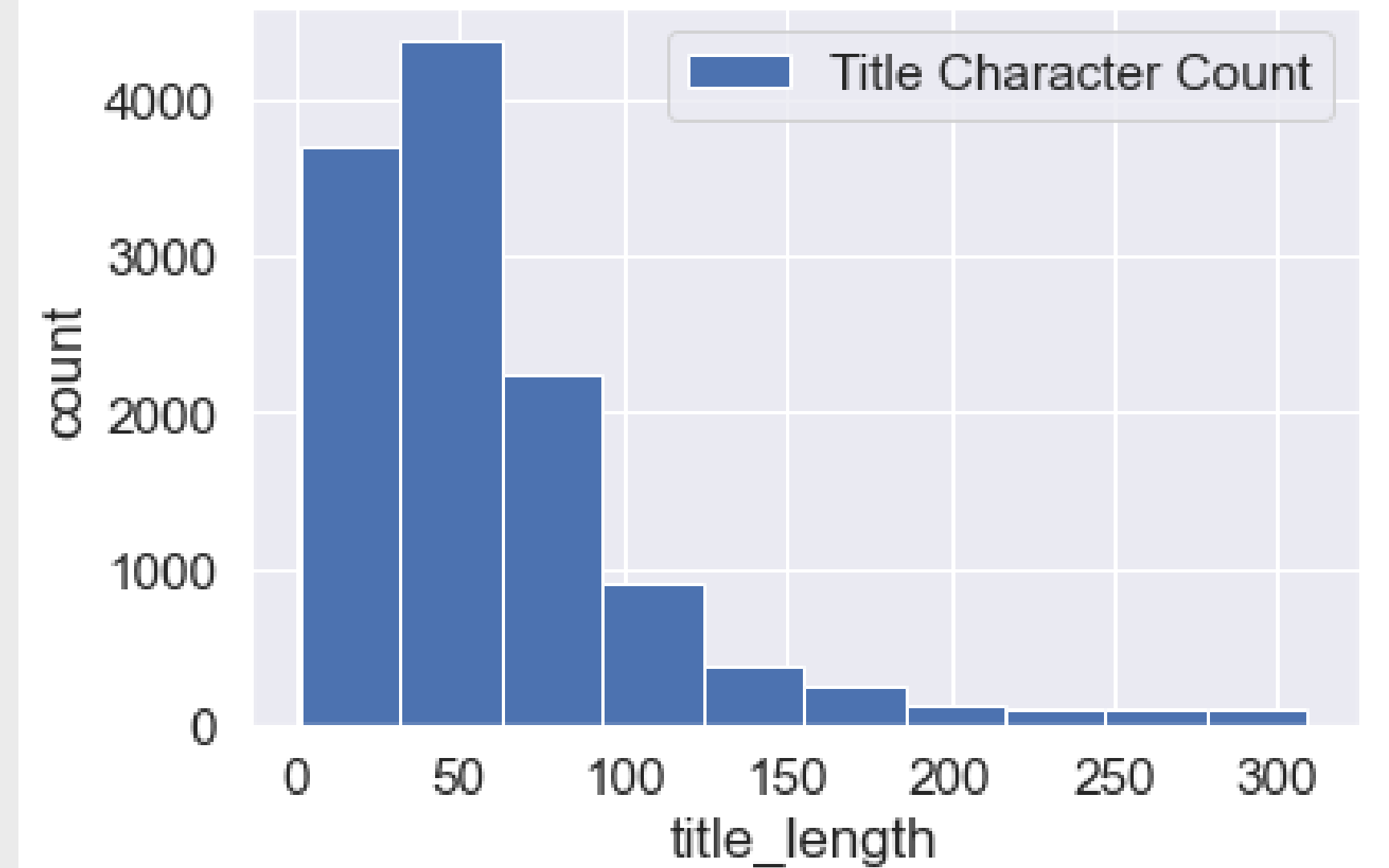
Everyone in life deserves to feel valid and supported for the things that they want to do and for the person that they are. I know as a child growing up my parents did not

Exploratory Data Analysis

The Distribution Subreddits

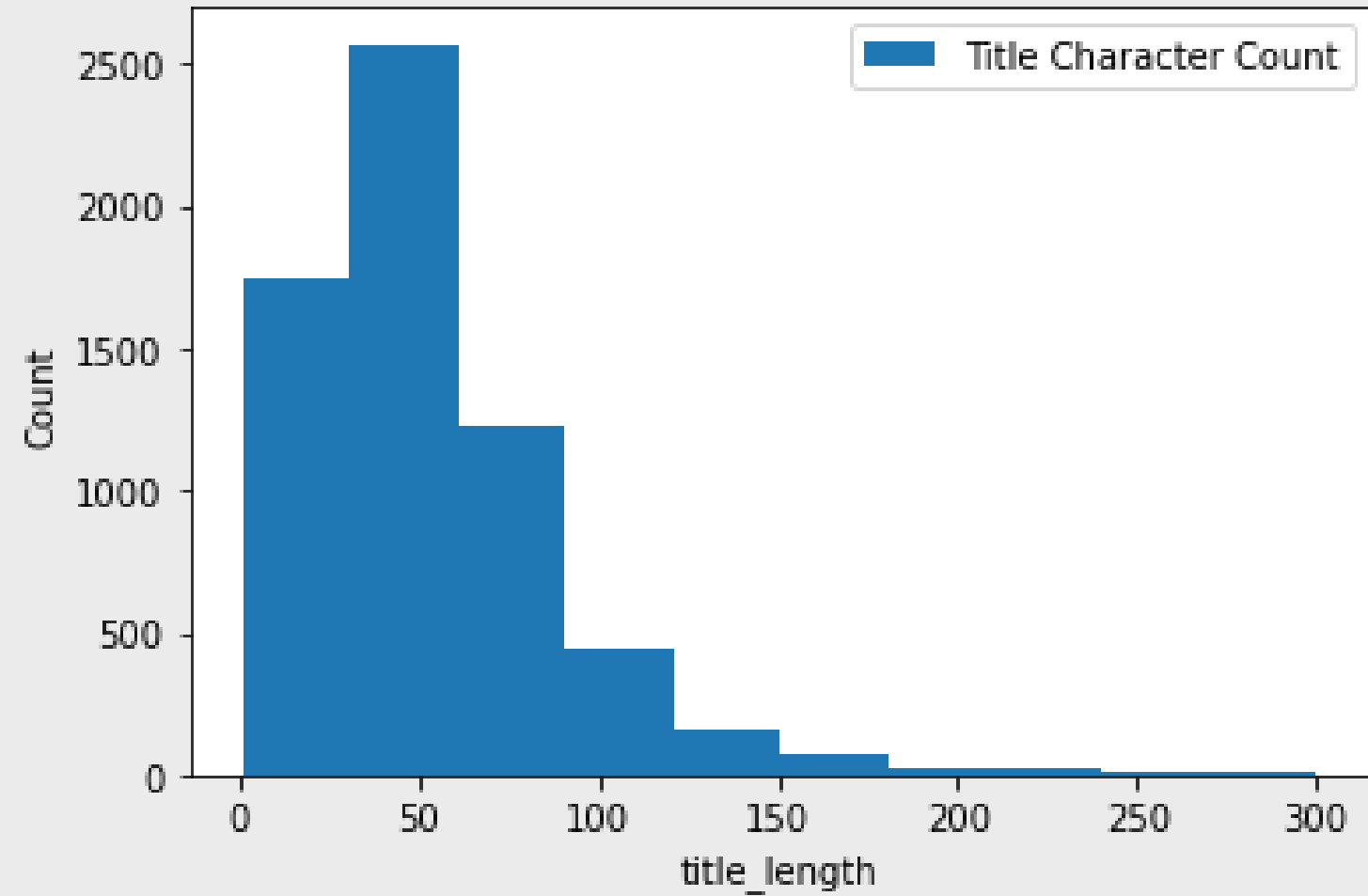


Distribution of Title Lengths by Combined Data

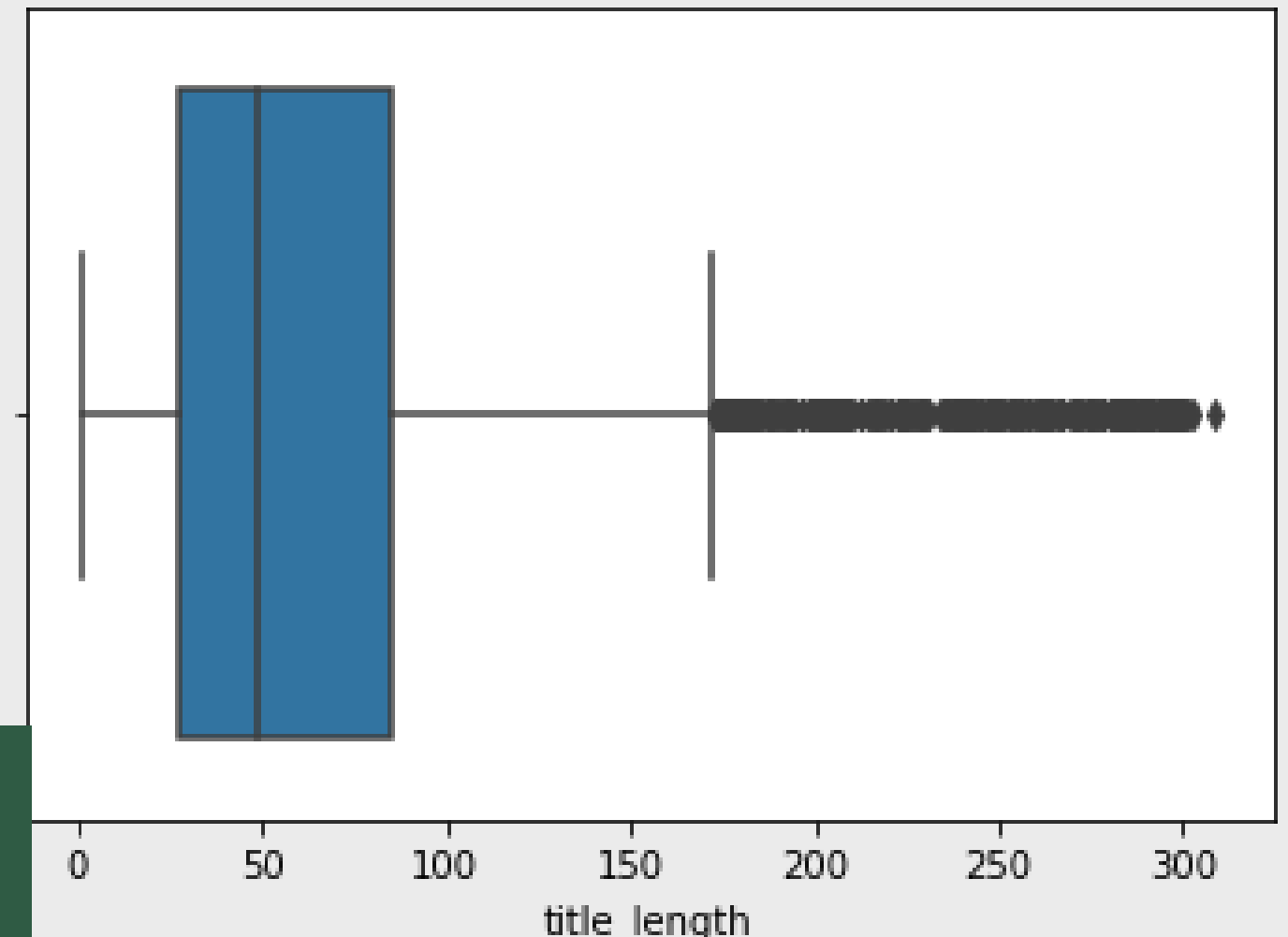
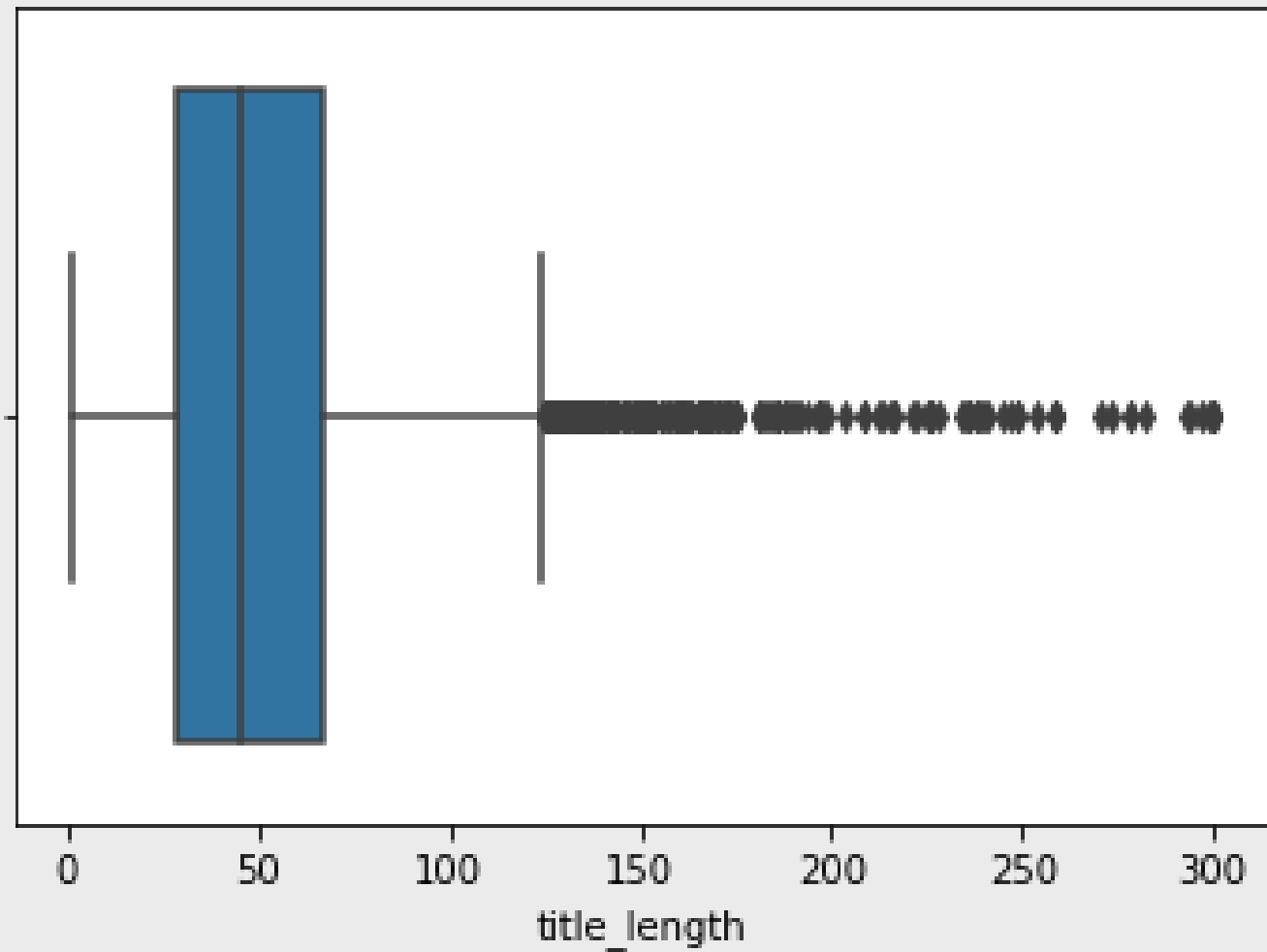
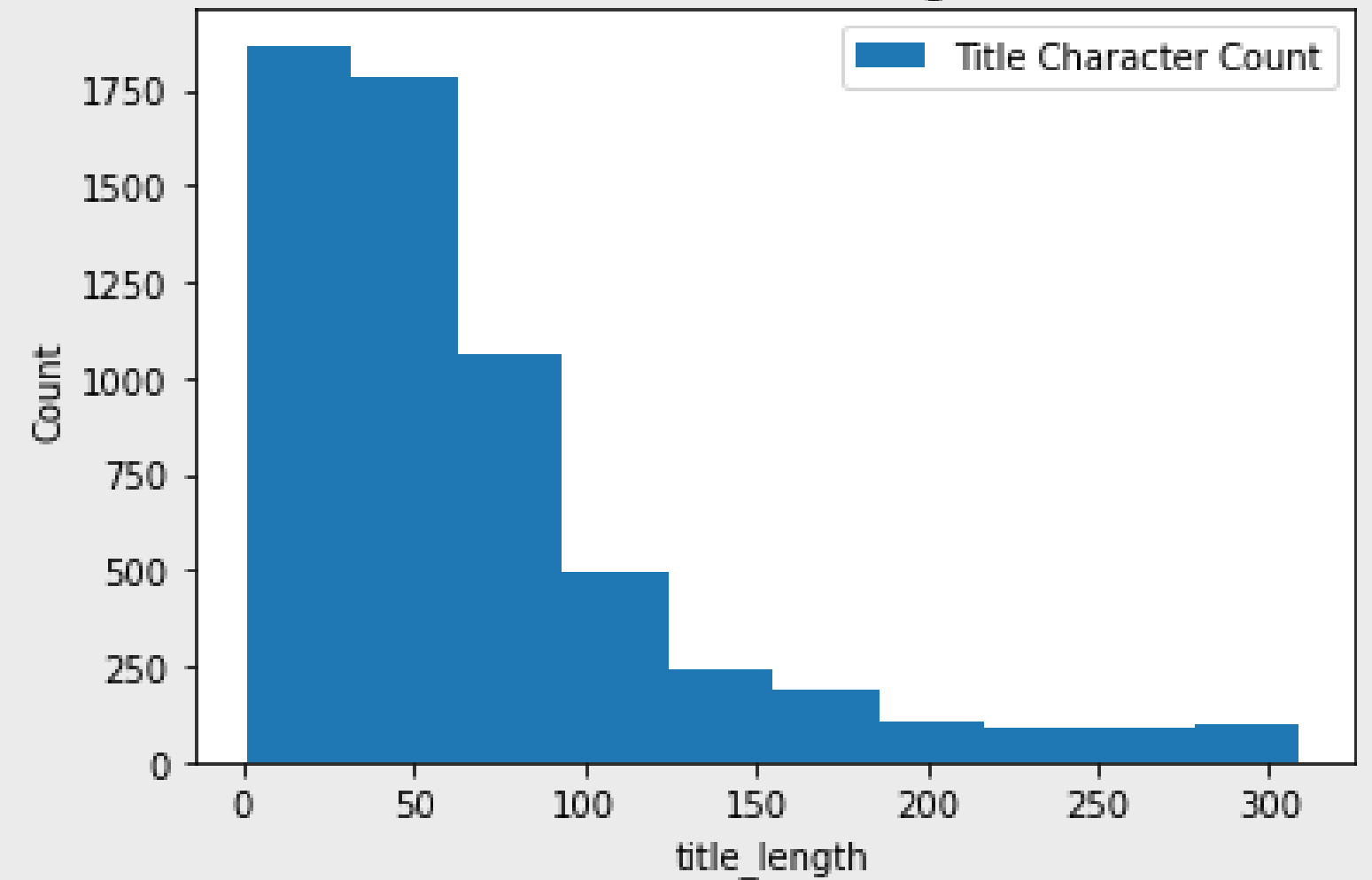


Deeper Exploratory Data Analysis

Ask Transgender Title Length



Rainbow Title Length



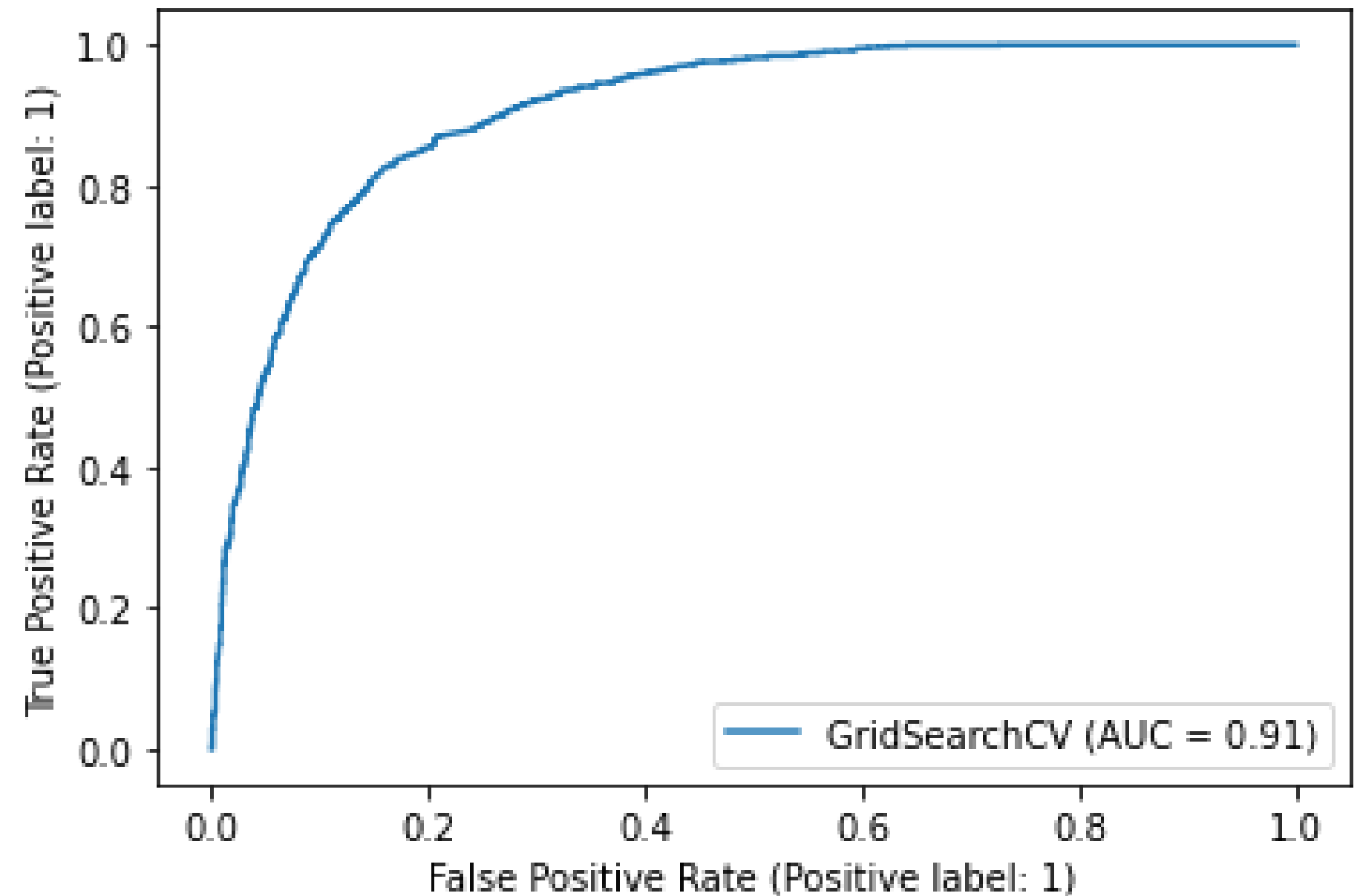
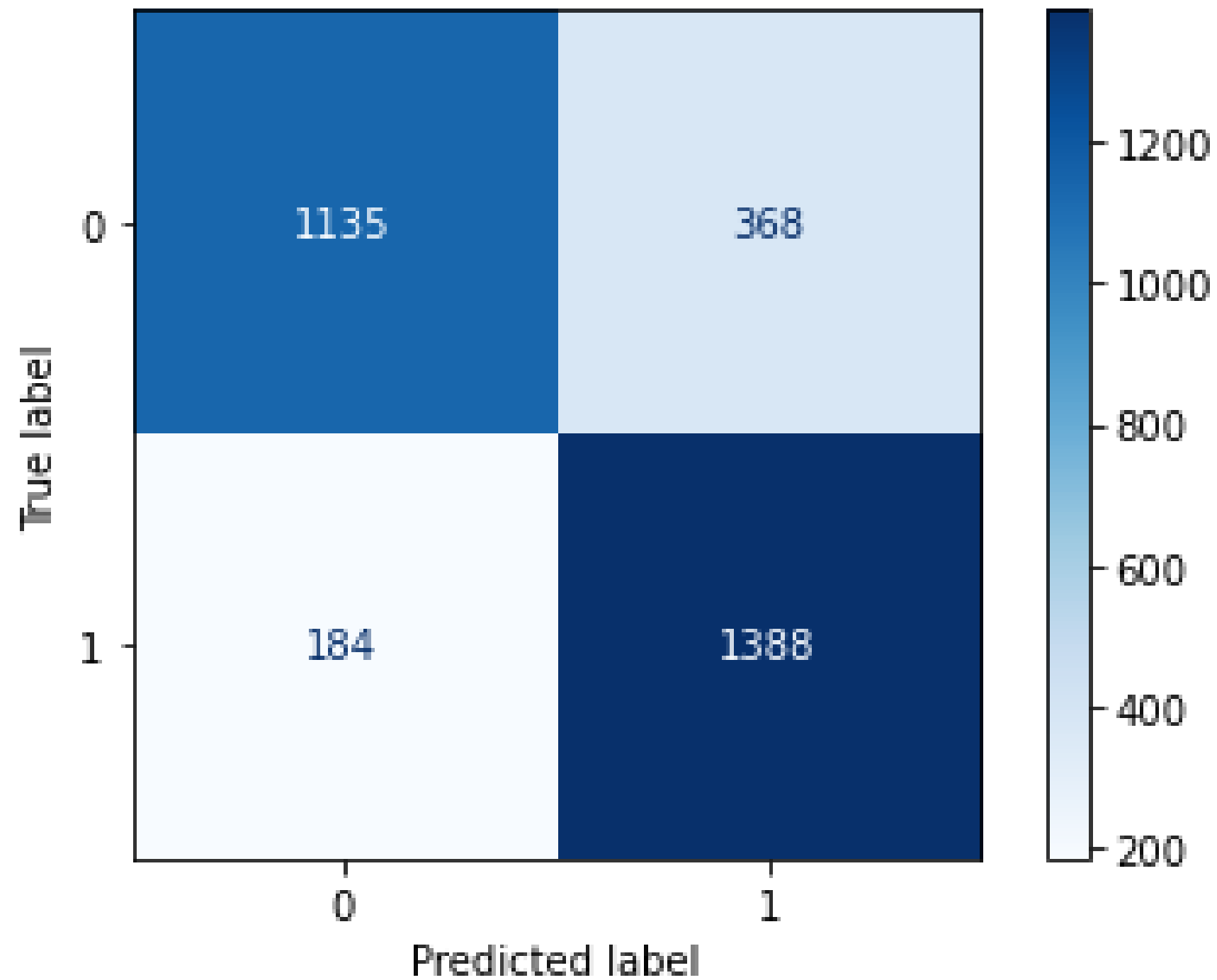
Can you guess which word cloud belongs to
their respective subreddit?

seems
non someone
person school
hot though way gay
honest
wanna
mean sex appreciate

help middle
need MtF
name

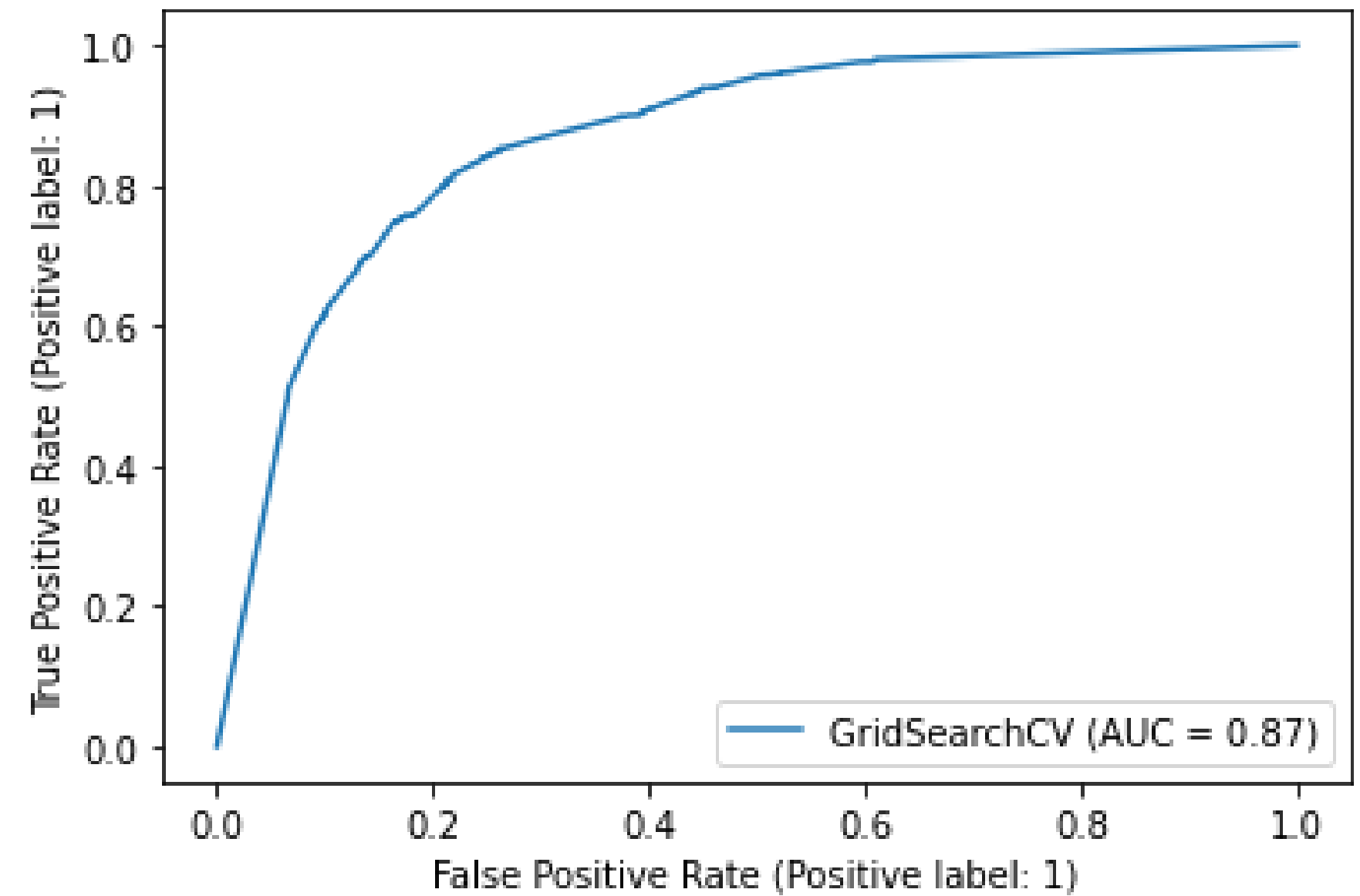
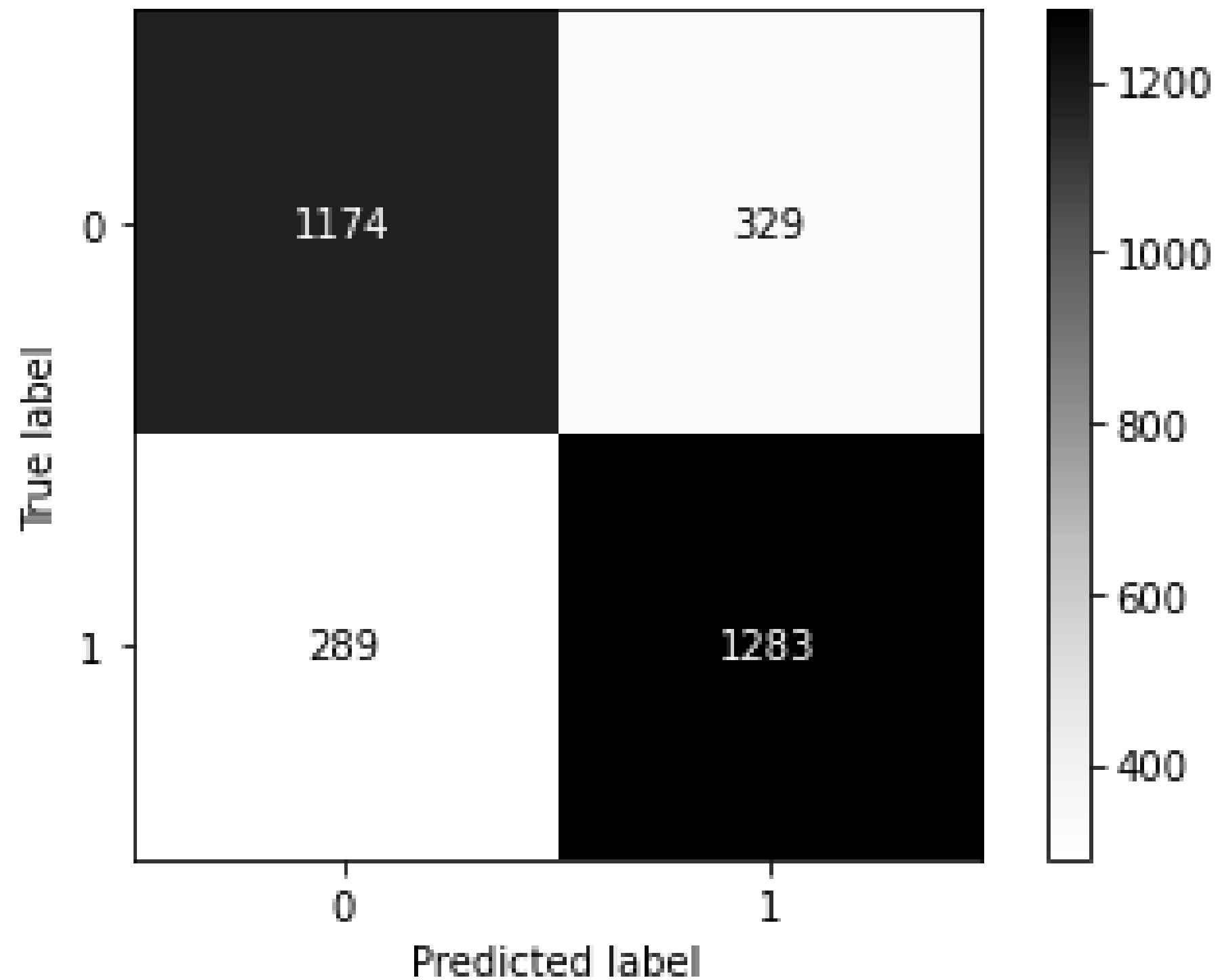
Model Type	Training Score (accuracy)	Testing Score (accuracy)
Null Model	N/A	51.1%
Multinomial Naive Bayes with Countvectorizer	83.1%	82.0%
Logistic Regression with Tfidf (No hyperparameters)	83.0%	82.0%
Multinomial Naive Bayes with Tfidf	82.3%	81.2%
Tfidf with Bagging Classifier	79.9%	79.9%

Multinomial Naive Bayes with Countvectorizer



1 = asktransgender, 0 = r/ainbow

Tfidf with Bagging Classifier



1 = asktransgender, 0 = r/ainbow

Conclusion:

Queer resources increase too. Since the two subreddits are very similar, the model can help people increasingly direct their questions to the appropriate subreddit



As the queer population increases, the resources needed to help people with their new transitional lifestyle increase.



Using a multinomial Naive Bayes with countvectorization has proved the most robust model. It predicted which subreddit the title text belongs to.



Thank you!



Questions?

***If you would like to see a deeper analysis within jupyter notebooks, please request after the presentation.**

Sources: [reddit.com](https://www.reddit.com)