

How I Analyzed Your Mother

Sara Robinson



Introduction

- Purpose
- Implications



The Process

Obtain

- Forever Dreaming Transcripts
- Web Scraping



Scrub

- Text Preprocessing
- SQL Preparation

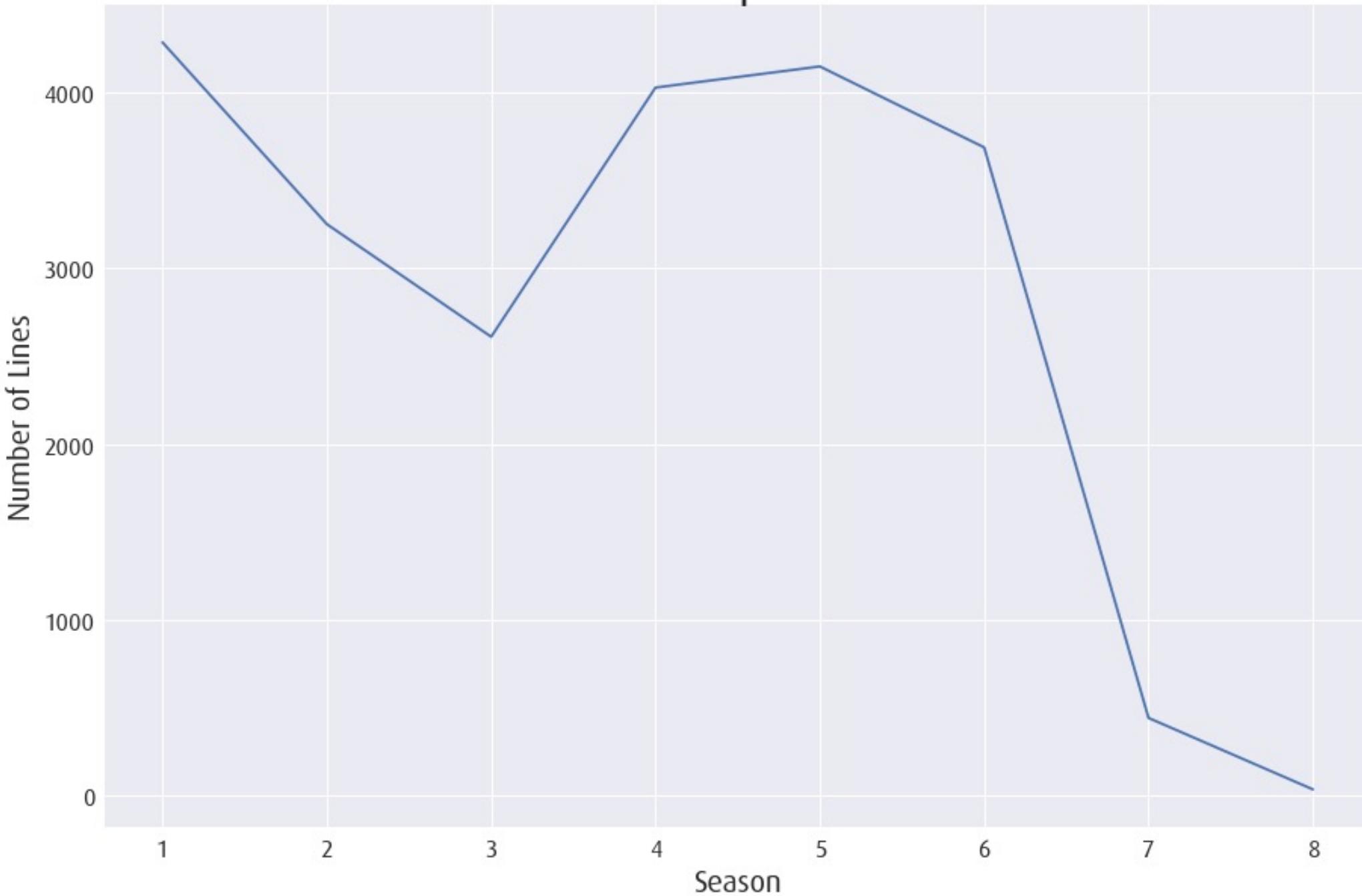


Explore

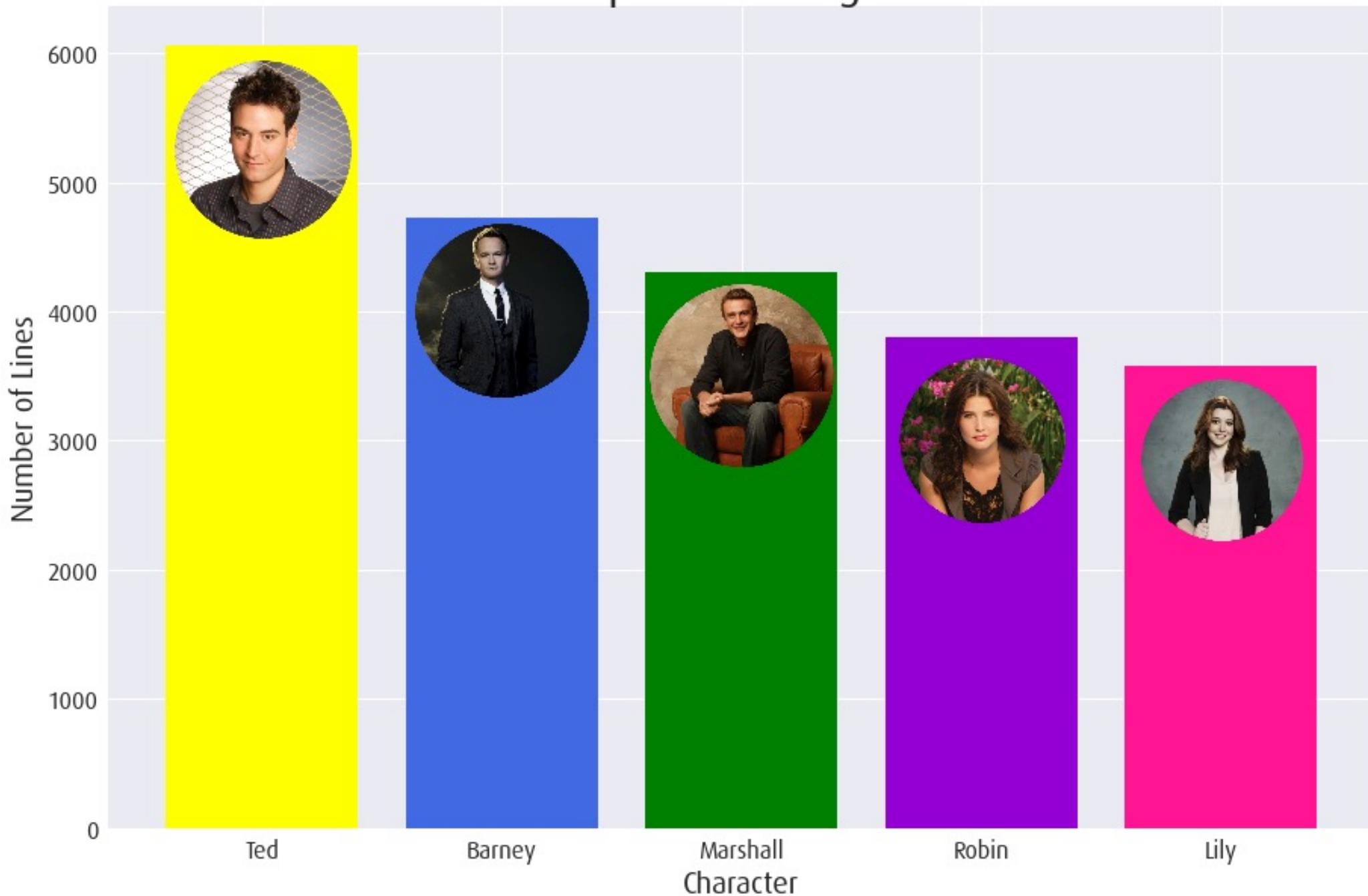
- Total Lines
- Largest Vocabulary
- Most Used Terms
- Common Phrases



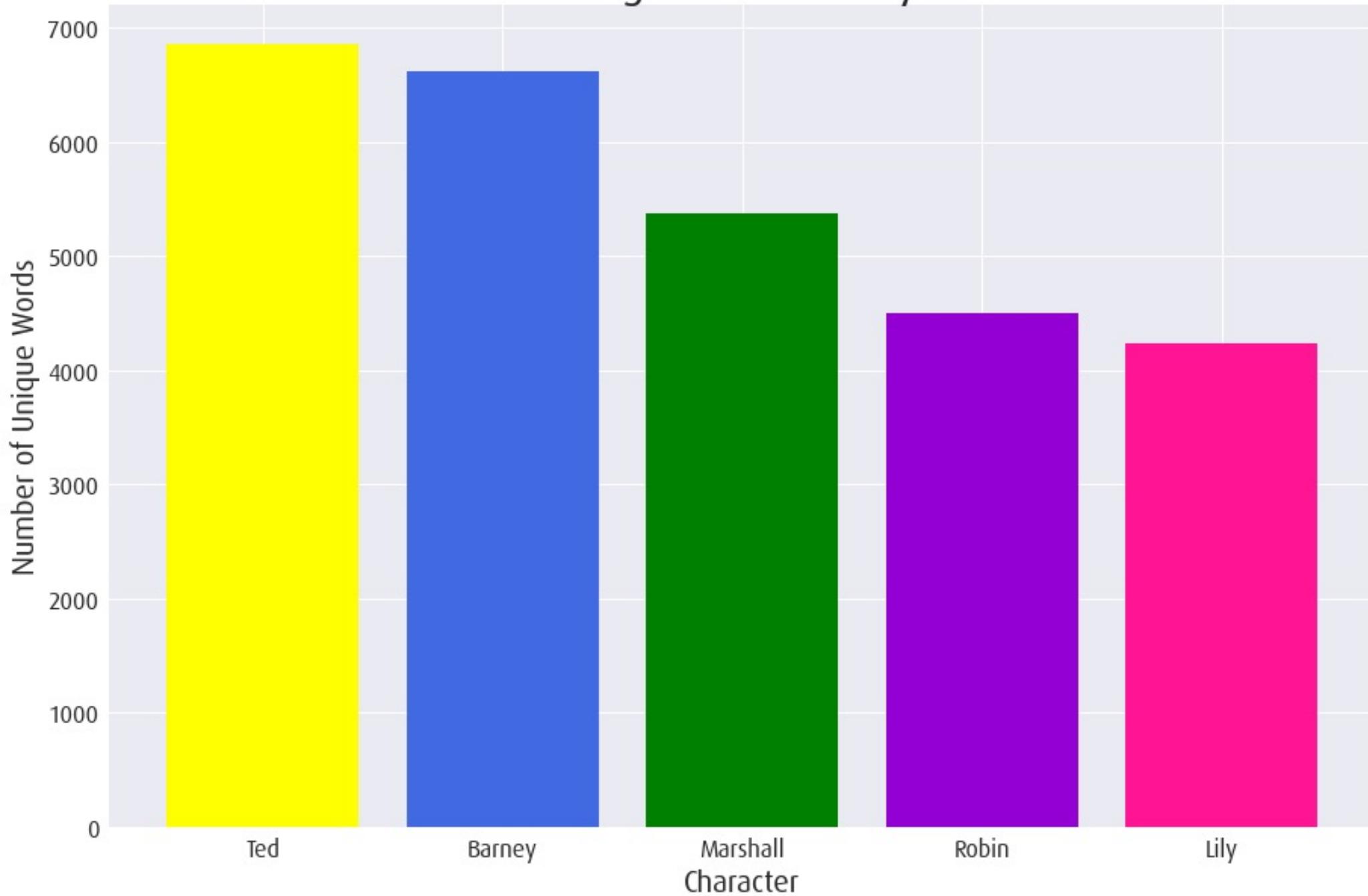
Total Lines per Season



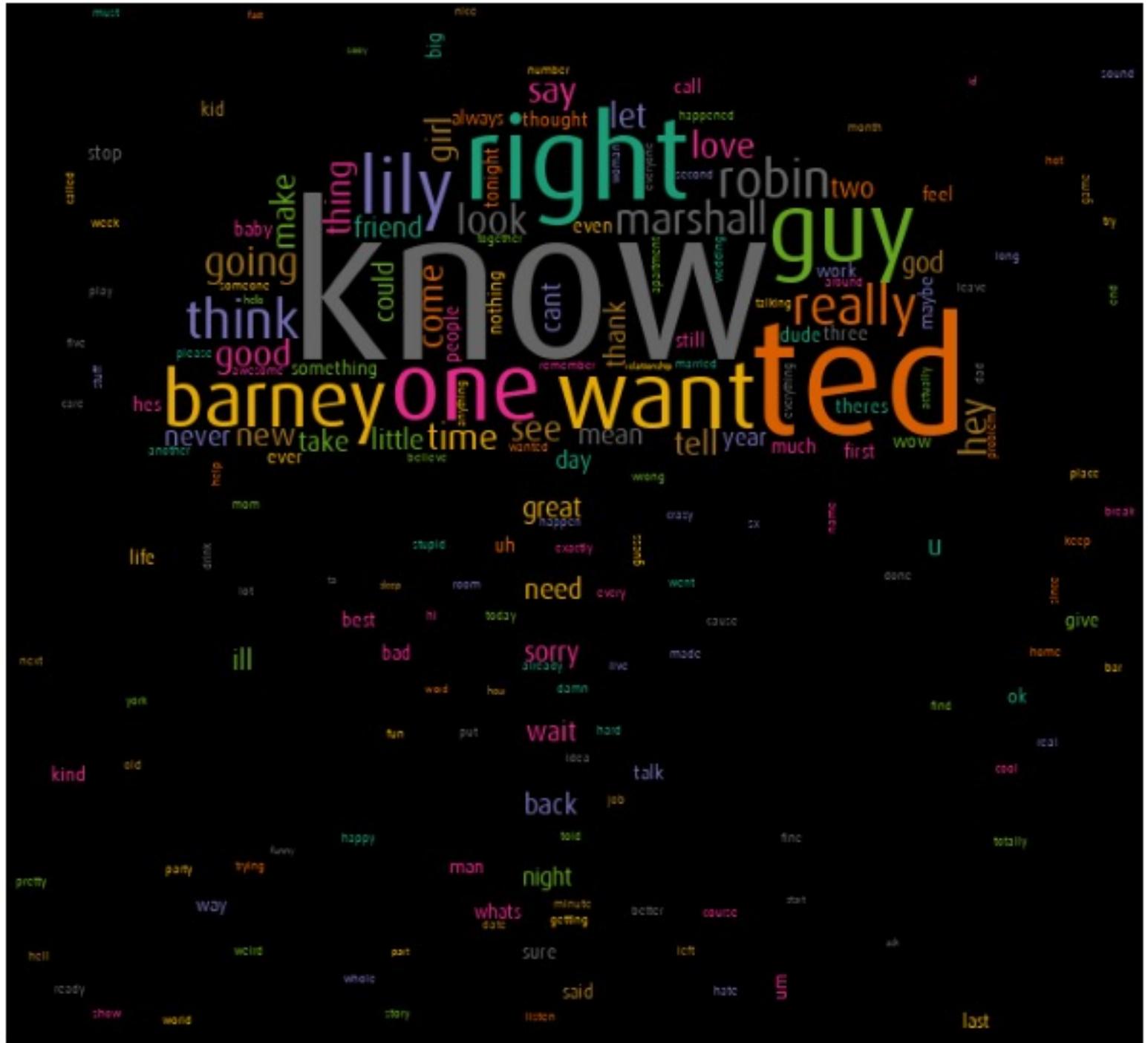
Total Lines Spoken During Entire Series



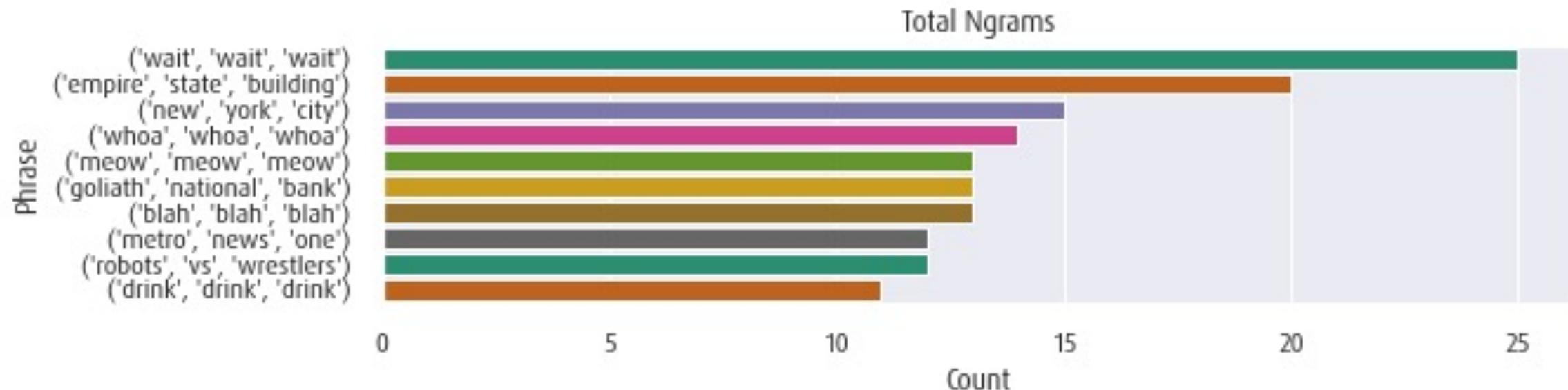
Largest Vocabulary



Most Common Terms



Three Word Phrases



Model

- Brute Force Approach
- Tuned Top Models
- Cohen's Kappa Score

iNterpret

- Logistic Regression using Count Vectorizer with C = 0.5
- Cohen's Kappa Score = 0.125
- Accuracy = 0.32

Conclusions

- Final Model
- Imbalanced Data
- Pitfalls

Further Explorations

- More EDA
- Different Modeling



Thank You

Image Citations

- [Title Slide Image](#)
- [Blue French Horn](#)
- [Ducky Tie](#)
- [Red Cowboy Boots](#)
- [Yellow Umbrella](#)
- [GNB Building](#)
- [Thank You Slide Image](#)