



GOODREADS SENTIMENT ANALYSIS

MICHAEL BARMADA

INTRODUCTION

- 100 Years of Sci-Fi
 - Bethanie Maples, Srini Kadamtati, & Eric Berlow
 - Interactive clusters of sci-fi books based on keywords generated from Goodreads reviews
- Goals:
 - Train a classifier to properly categorize Goodreads reviews (either by score-level or binary sentiment)
 - Maybe get distracted along the way
- Hypotheses
 - More positive reviews will have more likes
 - Sentiment will be skewed slightly positive for lower-scoring reviews
 - Negative reviews will focus more on one's experience (more personal pronouns)
 - Negative reviews will be longer (?)

GOODREADS IN A NUTSHELI

What is it?

- A website that combines elements of an audience review site with a social media platform
- Not just for reviewing books, but for following what your friends are reading

Why Goodreads?

- High engagement
- Lots of data types
- User-generated data

Share   

Recommend It | Stats | Recent Status Updates

READERS ALSO ENJOYED

See similar books...

GENRES

Fantasy	16,355 users
Fiction	1,996 users
Fantasy > Epic Fantasy	1,209 users
Fantasy > High Fantasy	778 users
Audiobook	473 users
Adult	410 users
Science Fiction Fantasy	395 users
Epic	322 users
Fantasy > Magic	315 users
Adventure	308 users

See top shelves...

ABOUT BRANDON SANDERSON

Brandon Sanderson 9
161,176 followers
Brandon's major books for the second half of 2016 are *The Dark Talent*, the final volume in Alcatraz Smedry's autobiographical account of his battle against the Evil Librarians who secretly rule our world, and *Arkanum Unbound*, the collection of short fiction in the Cosmere universe that includes the Mistborn series and the Stormlight Archive, among others. This collection features *The Emperor's S... more*

OTHER BOOKS IN THE SERIES

The Stormlight Archive (10 books)

The Stormlight Archive (10 books)

BOOKS BY BRANDON SANDERSON

Books by Brandon Sanderson

The Way of Kings
(The Stormlight Archive #1)
by Brandon Sanderson (Goodreads Author)
 4.61 · Rating details · 332,073 ratings · 21,955 reviews

From #1 New York Times bestselling author Brandon Sanderson, *The Way of Kings*, book one of The Stormlight Archive begins an incredible new saga of epic proportion.

Roshar is a world of stone and storms. Uncanny tempests of incredible power sweep across the rocky terrain so frequently that they have shaped ecology and civilization alike. Animals hide in shells, trees pull in ...more

Want to Read

Rate this book

Open Preview

GET A COPY

Kindle Store \$9.99 Amazon Stores Libraries

Hardcover, 1007 pages Published August 31st 2010 by Tor Books More Details... Edit Details

FRIEND REVIEWS

To see what your friends thought of this book, [please sign up](#).

READER Q&A

To ask other readers questions about *The Way of Kings*, [please sign up](#).

Popular Answered Questions

When does the action start? I'm at the part whereupon the guy steals the shard & that's not really that interesting, or scifi.
20 Likes · Like · 6 Years Ago · See All 31 Answers

Stephen Really? I've just started - I'm at 3% and I've just had to come up for air. I feel like I've been reading for ages. Only 3% and I've already been t...more
92 Likes · Like · 6 Years Ago · See All 25 Answers

Yashima Sanderson is one of the most productive writers I know. According to him this is going to be a 10 book series and a book will be published every other...more
See all 96 questions about *The Way of Kings*...

LISTS WITH THIS BOOK

The Best Epic Fantasy (fiction)
3,430 books — 24,576 voters

Most Interesting Magic System
2,342 books — 8,176 voters

WEB SCRAPING: A TEST OF PATIENCE

- Started off with code borrowed from Maria Antoniak's [Goodreads Scraper](#)
 - Ended up heavily modified...
- Ajax requests (the bane of my existence)
 - Update webpage in real time
 - No need to reload page or ***link to another URL***
- Goodreads only shows "top reviews" by default
- Limited to 10 pages of reviews. 30 reviews per page = 300 max
- **Only one filter at a time**

COMMUNITY REVIEWS Showing 1-30

★★★★★ 4.61 · [Rating details](#) · 332,073 ratings · 21,955 reviews

English (19699) More filters | Sort order Search review text

Start your review of **The Way of Kings (The Stormlight Archive, #1)**

Petrik rated it ★★★★★ Dec 10, 2016

Recommends it for: epic/high fantasy fans
Shelves: [favorites](#), [masterpieces](#)

I have a Booktube channel now! Subscribe here: <https://www.youtube.com/petrikleo>

Incredible, impressive or fantastic, all these words are an understatement to the quality this book holds. *The Way of Kings* is the beginning of a masterpiece series in epic fantasy. It is now my life goal and a new addition to my bucket list to obtain and read the entire series of *The Stormlight Archive*, which will probably take at least another 20 years from now to complete.

Before you started reading this book, let [...more](#)

955 likes · Like · see review

carol. rated it ★★★★★ May 18, 2011

Recommends it for: epic fantasy fans
Recommended to carol. by: virtually everybody
Shelves: [epic-fantasy](#)

A three and a half star read.

THE FANTASY SHELF

- Users can place books on certain “shelves” to categorize them for personal reference
- User-generated
- Goodreads keeps track of the top 100,000 books most frequently shelved as fantasy
- Only looking at first 50
- Issue: Multiple books from the same series

goodreads Home My Books Browse Community Search books Sign In Join

Shelves > Fantasy >

Fantasy Books

Showing 1-50 of 100,000

Harry Potter and the Sorcerer's Stone (Harry Potter #1) by J.K. Rowling (shelved 64073 times as fantasy) avg rating 4.48 — 7,469,089 ratings — published 1997	Want to Read Rate this book ★★★★★
Harry Potter and the Chamber of Secrets (Harry Potter, #2) by J.K. Rowling (shelved 53565 times as fantasy) avg rating 4.43 — 2,891,084 ratings — published 1998	Want to Read Rate this book ★★★★★
The Hobbit, or There and Back Again (Paperback) by J.R.R. Tolkien (shelved 53003 times as fantasy) avg rating 4.28 — 3,032,872 ratings — published 1937	Want to Read Rate this book ★★★★★
Harry Potter and the Prisoner of Azkaban (Harry Potter, #3) by J.K. Rowling (shelved 52611 times as fantasy) avg rating 4.57 — 2,988,156 ratings — published 1999	Want to Read Rate this book ★★★★★
Harry Potter and the Goblet of Fire (Harry Potter, #4) by J.K. Rowling (shelved 51467 times as fantasy) avg rating 4.56 — 2,737,442 ratings — published 2000	Want to Read Rate this book ★★★★★
Harry Potter and the Deathly Hallows (Harry Potter, #7) by J.K. Rowling (shelved 49613 times as fantasy) avg rating 4.62 — 2,945,868 ratings — published 2007	Want to Read Rate this book ★★★★★

LISTS TAGGED "FANTASY"

Best Woman-Author Books
7,078 books — 5,672 voters

Best Science Fiction & Fantasy Books
7,170 books — 21,956 voters

More lists...

QUOTES TAGGED "FANTASY"

“Perhaps the greatest faculty our minds possess is the ability to cope with pain. Classic thinking teaches us of the four doors of the mind, which everyone moves through according to their need.

First is the door of sleep. Sleep offers us a retreat from the world and all its pain. Sleep marks passing time, giving us distance from the things that have hurt us. When a person is wounded they will

THE DATA (FINALLY)

```
A Court of Thorns and Roses
A Darker Shade of Magic
A Game of Thrones
A Wizard of Earthsea
American Gods
Assassin's Apprentice
Eragon
Good Omens: The Nice and Accurate Prophecies of Agnes Nutter, Witch
Graceling
Harry Potter and the Sorcerer's Stone
Neverwhere
Six of Crows
Stardust
The Color of Magic
The Eye of the World
The Fellowship of the Ring
The Final Empire
The Golden Compass
The Hobbit, or There and Back Again
The Lies of Locke Lamora
The Lightning Thief
The Lion, the Witch and the Wardrobe
The Lord of the Rings
The Name of the Wind
The Night Circus
The Ocean at the End of the Lane
The Princess Bride
The Subtle Knife
The Way of Kings
Throne of Glass
Twilight
Uprooted
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3778 entries, 0 to 3777
Data columns (total 8 columns):
 #   Column   Non-Null Count   Dtype  
--- 
 0   Title     3778 non-null    object 
 1   Year      3778 non-null    int64  
 2   Genres    3778 non-null    object 
 3   Rating    3778 non-null    int64  
 4   Likes     3778 non-null    object 
 5   Date      3778 non-null    object 
 6   Shelves   3778 non-null    object 
 7   Text      3778 non-null    object 
dtypes: int64(2), object(6)
memory usage: 236.2+ KB
```

- 3778 entries total from 32 URLs
- Aimed for 120 reviews per book
- 7 columns
 - Title
 - Year
 - Genres
 - Rating
 - Likes
 - Date
 - Shelves
 - Text

Out[4]:

	Title	Year	Genres	Rating	Likes	Date	Shelves	Text
0	Harry Potter and the Sorcerer's Stone	1997	['Fantasy', 'Fiction', 'Young Adult', 'Fantasy...']	5	1532 likes	Nov 11, 2019	['absolute-favorites', 'audiobook']	\nStuck at home? Got some time on your hands?...
1	Harry Potter and the Sorcerer's Stone	1997	['Fantasy', 'Fiction', 'Young Adult', 'Fantasy...']	5	1509 likes	Jun 03, 2011	['favorites', 'book-to-movie', 'lib-read', 'mi...']	I'm going to keep this brief since there isn't...
2	Harry Potter and the Sorcerer's Stone	1997	['Fantasy', 'Fiction', 'Young Adult', 'Fantasy...']	5	802 likes	Dec 10, 2009	['books-i-own', '5-star-excellent', 'favorites...']	My original review was a comparison of sorts b...
2424	The Lord of the Rings	1955	['Fantasy', 'Classics', 'Fiction', 'Adventure']...	3	7 likes	Nov 16, 2016	['audiobooks', 'fantasy', 'fae-creatures', 'ma...']	Before you all kill me for giving this only 3 ...
1170	The Lightning Thief	2005	['Fantasy', 'Young Adult', 'Fantasy', 'Mytholo...']	3	29 likes	Sep 13, 2012	['mythology-folklore', 'middle-grade']	First reread, March 2012: 5 starsSecond reread...
2385	The Lord of the Rings	1955	['Fantasy', 'Classics', 'Fiction', 'Adventure']...	1	2 likes	Feb 06, 2021	[]	Uno dei pochi rari esempi in cui nel film son ...
2782	The Ocean at the End of the Lane	2013	['Fantasy', 'Fiction', 'Magical Realism', 'Hor...']	3	40 likes	Aug 24, 2020	['science-fiction']	This is probably an unpopular opinion regardin...
411	The Fellowship of the Ring	1954	['Fantasy', 'Classics', 'Fiction', 'Adventure']...	1	3 likes	Aug 25, 2012	['1-reviewed-by-me', '6-way-overrated']	The following rant is from an email that I wro...
1261	American Gods	2001	['Fantasy', 'Fiction', 'Fantasy', 'Urban Fanta...']	2	23 likes	Aug 17, 2007	['urban-fantasy', 'religion', 'mythology', 'ho...']	Neil Gaiman is my favorite author, and I have ...
1827	The Eye of the World	1990	['Fantasy', 'Fiction', 'Fantasy', 'Epic Fantas...']	4	72 likes	Mar 12, 2017	[]	4.25/5 "The Wheel weaves as the Wheel wills" The...
504	The Name of the Wind	2007	['Fantasy', 'Fiction', 'Fantasy', 'Epic Fantas...']	5	202 likes	Dec 22, 2019	['2020-reads', 'adult-books', 'fantasy', 'good...']	This Review ↗ Blog ↗ Twitter ↗ Instagram ↗\n...
67	Harry Potter and the Sorcerer's Stone	1997	['Fantasy', 'Fiction', 'Young Adult', 'Fantasy...']	2	8 likes	Sep 03, 2012	['romanzi', 'lett-inglese', 'gran-bretagna']	Non era mia intenzione leggere la saga di Harr...
2302	A Court of Thorns and Roses	2015	['Fantasy', 'Romance', 'Young Adult', 'New Adu...']	3	106 likes	Jan 28, 2017	[]	Aunque Sarah J.Maas y yo no empezamos con buen...

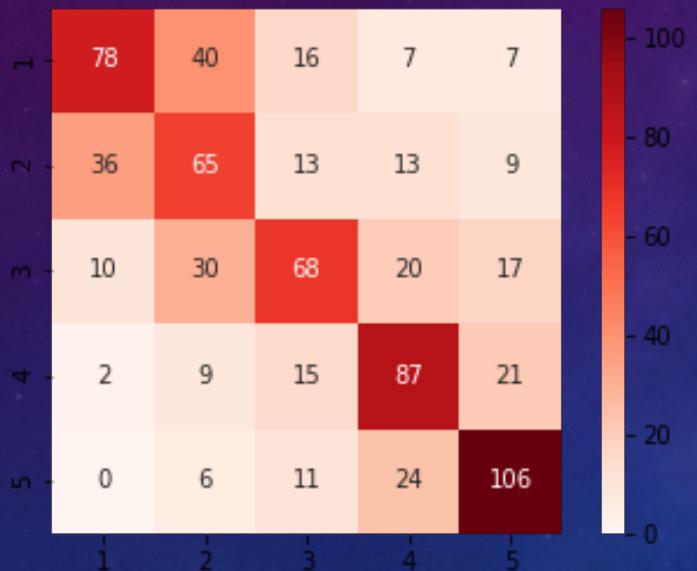
AFTER CLEANUP

- Now 3550 entries
- 228 non-English reviews removed
- Emojis and URLs scrubbed as well
- Added columns...
- Lang
- Toks
- Types
- TTR
- Index by book title

Out[23]:

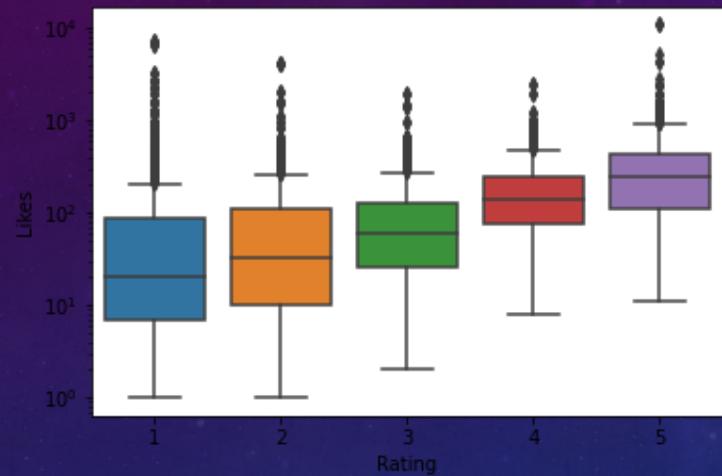
	Year	Genres	Rating	Likes	Date	Shelves	Text	Lang	Toks	Types	TTR
	Title										
Uprooted	2015	['Fantasy', 'Young Adult', 'Fiction', 'Romance...']	5	1456	2015-05-05	['fantasy', '2015', 'arc']	Hey! You there! Please listen. On May 19th thi...	en	638	320	0.501567
Uprooted	2015	['Fantasy', 'Young Adult', 'Fiction', 'Romance...']	5	898	2015-11-26	[]	I love a good fantasy rooted in folklore, and...	en	345	194	0.562319
Uprooted	2015	['Fantasy', 'Young Adult', 'Fiction', 'Romance...']	5	502	2014-09-03	['adult', 'recommended']	This novel effortlessly conjured up the famili...	en	62	49	0.790323
Graceling	2008	['Fantasy', 'Young Adult', 'Romance', 'Fiction...']	5	31	2012-04-19	['ya-series', 'favorites', 'ya-fantasy']	just as fantastic the second time around!UGH I...	en	234	127	0.542735
The Way of Kings	2010	['Fantasy', 'Fiction', 'Fantasy', 'Epic Fantas...']	5	955	2016-12-10	['favorites', 'masterpieces']	I have a Booktube channel now! Subscribe here:...	en	905	388	0.428729
The Night Circus	2011	['Fantasy', 'Fiction', 'Romance', 'Historical...']	3	240	2014-07-07	['owned', 'beautifully-written', 'favorites-20...']	i posted this review in case anyone needs a re...	en	929	441	0.474704
Neverwhere	1996	['Fantasy', 'Fiction', 'Fantasy', 'Urban Fanta...']	5	193	2019-01-29	['fantasy', 'owned', 'recommend', '5-stars', ...']	this full review is now POSTED on my blog! wha...	en	369	197	0.533875
The Ocean at the End of the Lane	2013	['Fantasy', 'Fiction', 'Magical Realism', 'Hor...']	2	50	2013-08-09	['wtf', 'shitty-parents', 'weird', 'fairies', ...']	2.5 stars - SpoilersThe Ocean at the End of th...	en	532	261	0.490602
The Way of Kings	2010	['Fantasy', 'Fiction', 'Fantasy', 'Epic Fantas...']	3	19	2020-12-29	[]	, ! The way of kings . (4.65). ...	en	36	18	0.500000
The Color of Magic	1983	['Fantasy', 'Fiction', 'Humor', 'Humor', 'Come...']	1	2	2007-08-18	[]	This is the first of a long series of books ab...	en	75	52	0.693333
Neverwhere	1996	['Fantasy', 'Fiction', 'Fantasy', 'Urban Fanta...']	5	168	2018-04-14	[]	Wow that was fantastic!! listened to the audio...	en	288	173	0.600694
A Game of Thrones	1996	['Fantasy', 'Fiction', 'Fantasy', 'Epic Fantas...']	3	43	2011-08-12	['adult', 'fantasy', 'gloomy', 'great-world-bu...']	Long winded doesn't even begin to describe Geo...	en	398	234	0.587940
Throne of Glass	2012	['Fantasy', 'Young Adult', 'Romance', 'Fiction...']	5	1149	2011-09-29	['covergasm', 'all-time-faves', 'heroines-i-ad...']	First cover: "Yeah, don't I look badass in thi...	en	192	118	0.614583

RATING CLASSIFIER (5-LEVELS)



- Multinomial Naïve Bayes Classifier
- Accuracy: 0.5478879559575024
- Best Parameters
 - `clf_alpha: 0.1`
 - `tfidf_max_features: 10000`
 - `tfidf_ngram_range: (1, 2)`
- Best at 5-star reviews
- Weaker w/ 2 & 3 star reviews
- Overall, not super great. We'll do binary sentiment analysis instead...

(DETOUR) LIKES / RATING CORRELATION



- Trends upwards
- Seems to support hypothesis about likes and rating being correlated
- Not *all* groups are significantly different, however...
- Trouble with (1, 4) group and (2, 3) group

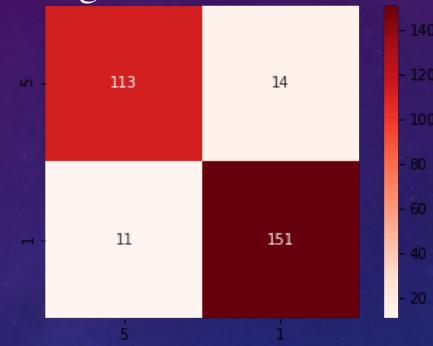
```
ttest group 1, 2:          t = 2.0045188884683998 p = 0.045258901223099994
ttest group 1, 3:          t = 2.54511713452734 p = 0.011106940802698004
ttest group 1, 4:          t = -1.0038992260412052 p = 0.31570412413906984
ttest group 1, 5:          t = -6.156791939965312 p = 9.608351714781413e-10
ttest group 2, 3:          t = 0.5378464285981069 p = 0.5907990088107831
ttest group 2, 4:          t = -4.970925215750165 p = 7.651938463269331e-07
ttest group 2, 5:          t = -9.345662246726858 p = 4.6687939299204386e-20
ttest group 3, 4:          t = -7.864747766691318 p = 7.384665692987975e-15
ttest group 3, 5:          t = -10.524198595265732 p = 1.863059175961704e-24
ttest group 4, 5:          t = -7.034155147049798 p = 3.904914711889499e-12
```

BINARY SENTIMENT ANALYSIS: TESTING CLASSIFIERS

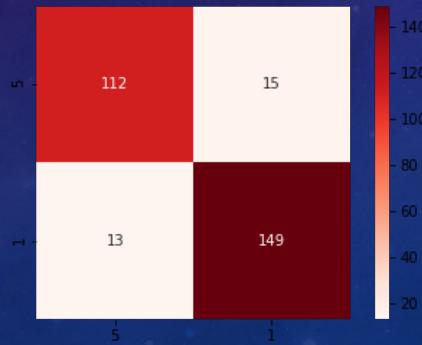
MultinomialNB: 0.9134 acc



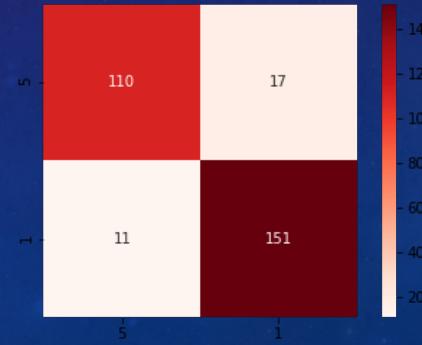
LogisticRegressor: 0.9134
acc



GridSearch + LR: 0.9003
acc



LSA + SVC: 0.9031
acc



SENTIMENT AT DIFFERENT SCORE LEVELS

```
In [53]: mid_df[mid_df.Rating== 2].Sentiment.value_counts()  
mid_df[mid_df.Rating== 3].Sentiment.value_counts()  
mid_df[mid_df.Rating== 4].Sentiment.value_counts()
```

```
Out[53]: 1    552  
5    132  
Name: Sentiment, dtype: int64
```

```
Out[53]: 1    343  
5    340  
Name: Sentiment, dtype: int64
```

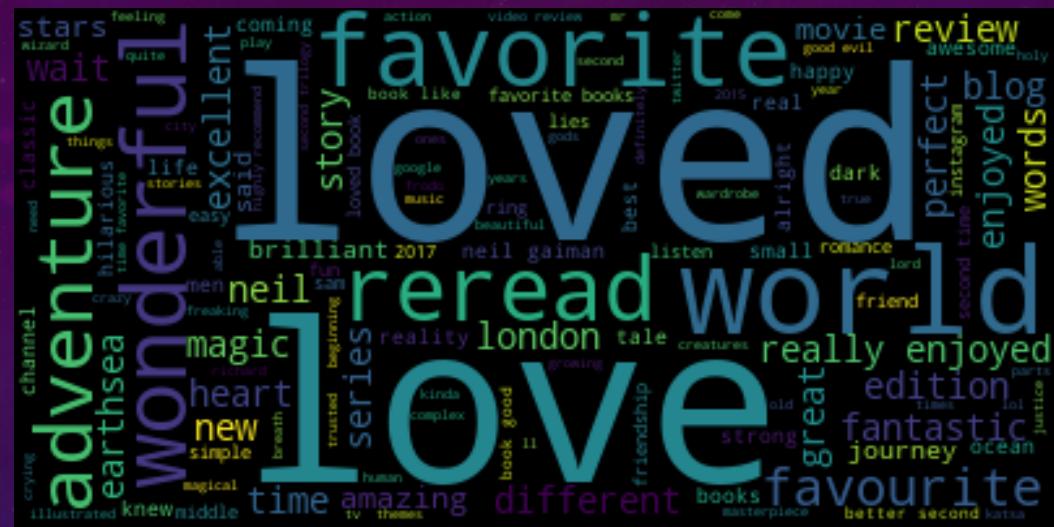
```
Out[53]: 5    601  
1    139  
Name: Sentiment, dtype: int64
```

```
In [54]: mid_df.groupby(['Rating', 'Sentiment']).Toks.mean()
```

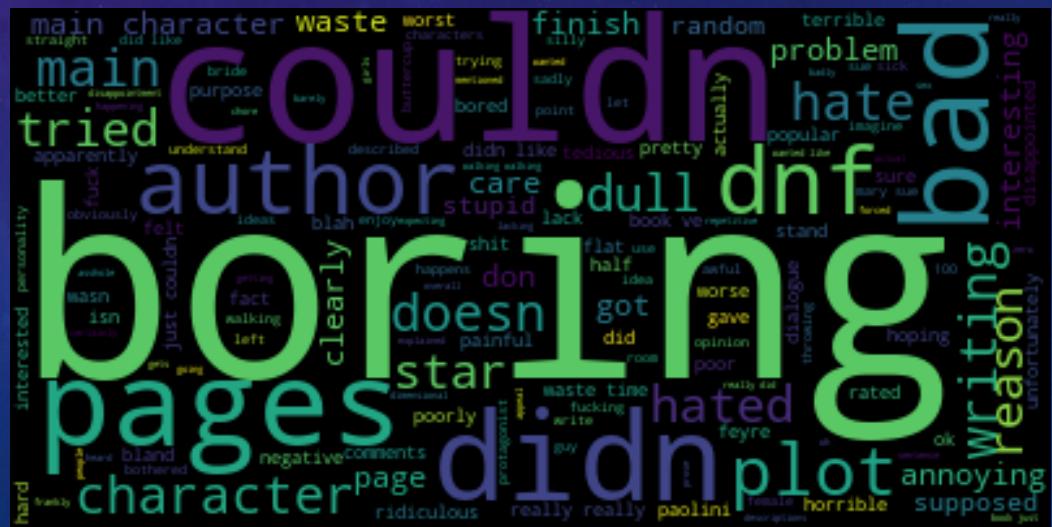
```
Out[54]: Rating  Sentiment  
          2        1            489.228261  
                  5            469.916667  
          3        1            466.781341  
                  5            470.105882  
          4        1            433.316547  
                  5            535.184692  
Name: Toks, dtype: float64
```

- Surprisingly balanced
- More positive 4-star reviews
- More negative 2-star reviews
- Almost equal at 3-stars
- Still, why the outliers?
- Length doesn't seem to be a factor. They are all fairly equal

TOP NEGATIVE FEATURES

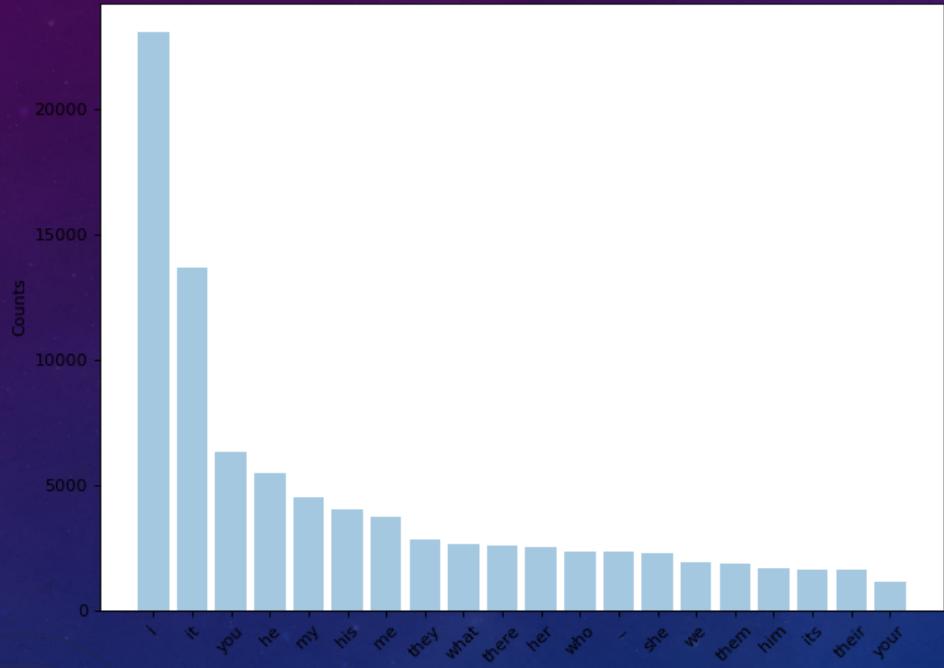


TOP POSITIVE FEATURES

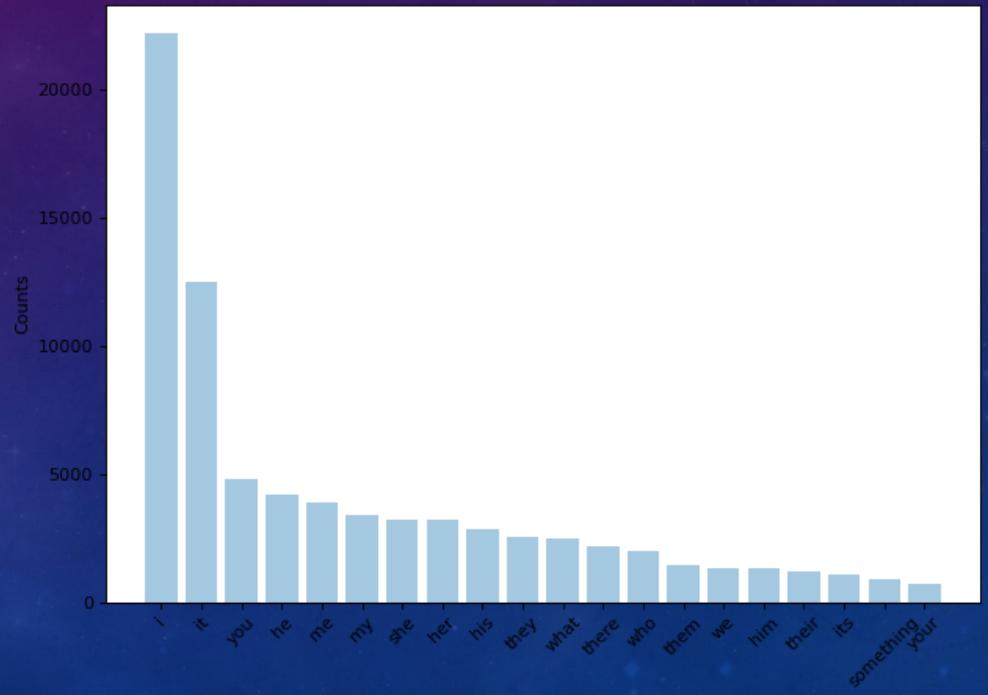


PRONOUN USE: POSITIVE VS NEGATIVE REVIEWS

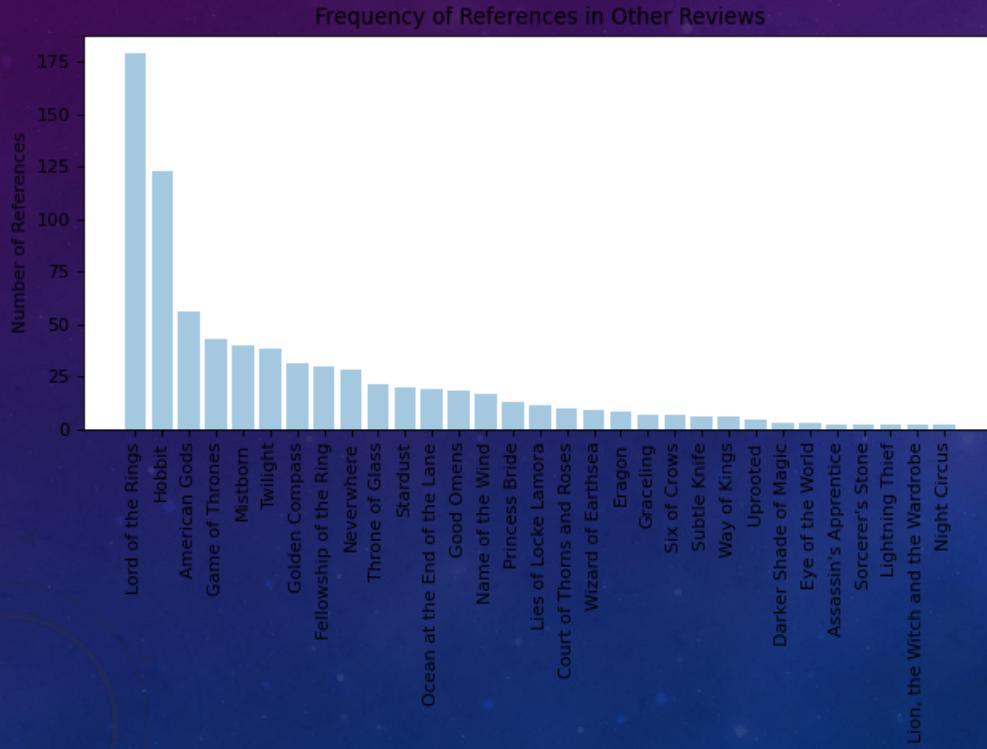
Frequency of Pronouns (positive)



Frequency of Pronouns (negative)



BONUS ROUND: HOW OFTEN DOES A BOOK'S TITLE OCCUR IN ANOTHER BOOK'S REVIEW?



- Had to drop leading determiners
- Certain titles had to be adjusted by hand
- Tolkien at the top (hardly surprising)
- Use of acronyms may throw off some of this data

TO-DO

- **NON-LEXICAL DATA**
- Explore how we can improve accuracy w/ spacy
- Collocation of top features / book titles
- Explore outliers (e.g. negative 4-star reviews and positive 2-star reviews)