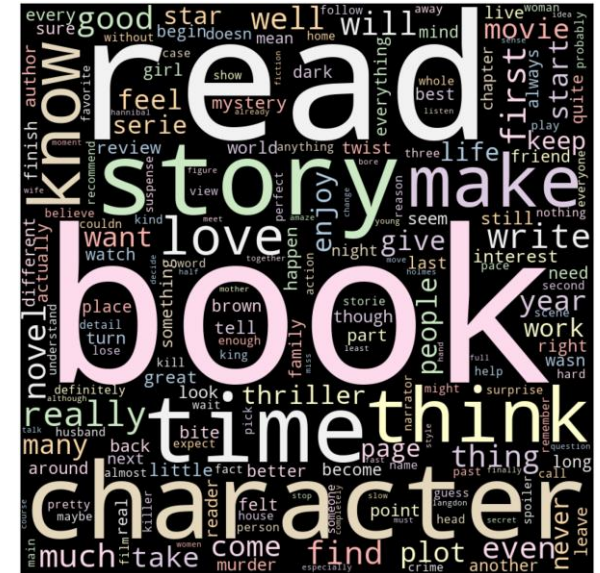


# TEXT SENTIMENT ANALYSIS FOR BOOK RECOMMENDATION

# Sathish Jayaraman



# SUMMARY

---

- Creation of statistical models to provide book recommendations.
- Understand similarity and patterns as well as provide suggestions on similar books to readers
- Existing approaches :
  - One of the approaches towards development of statistical models for book recommendations involves the use of text sentiment analysis (aspect-based sentiment analysis).
  - Aspect level sentiment analysis is the most fine-grained analysis of reviews of texts or social media opinions with respect to specific aspects.
  - Existing published research studies use machine learning techniques such as artificial neural networks, Naïve Bayes classifiers, KNN etc.
- There is a need for better modeling of the sentiment and more accurate estimation of the overall sentence sentiment bearing words.

# SUMMARY

---

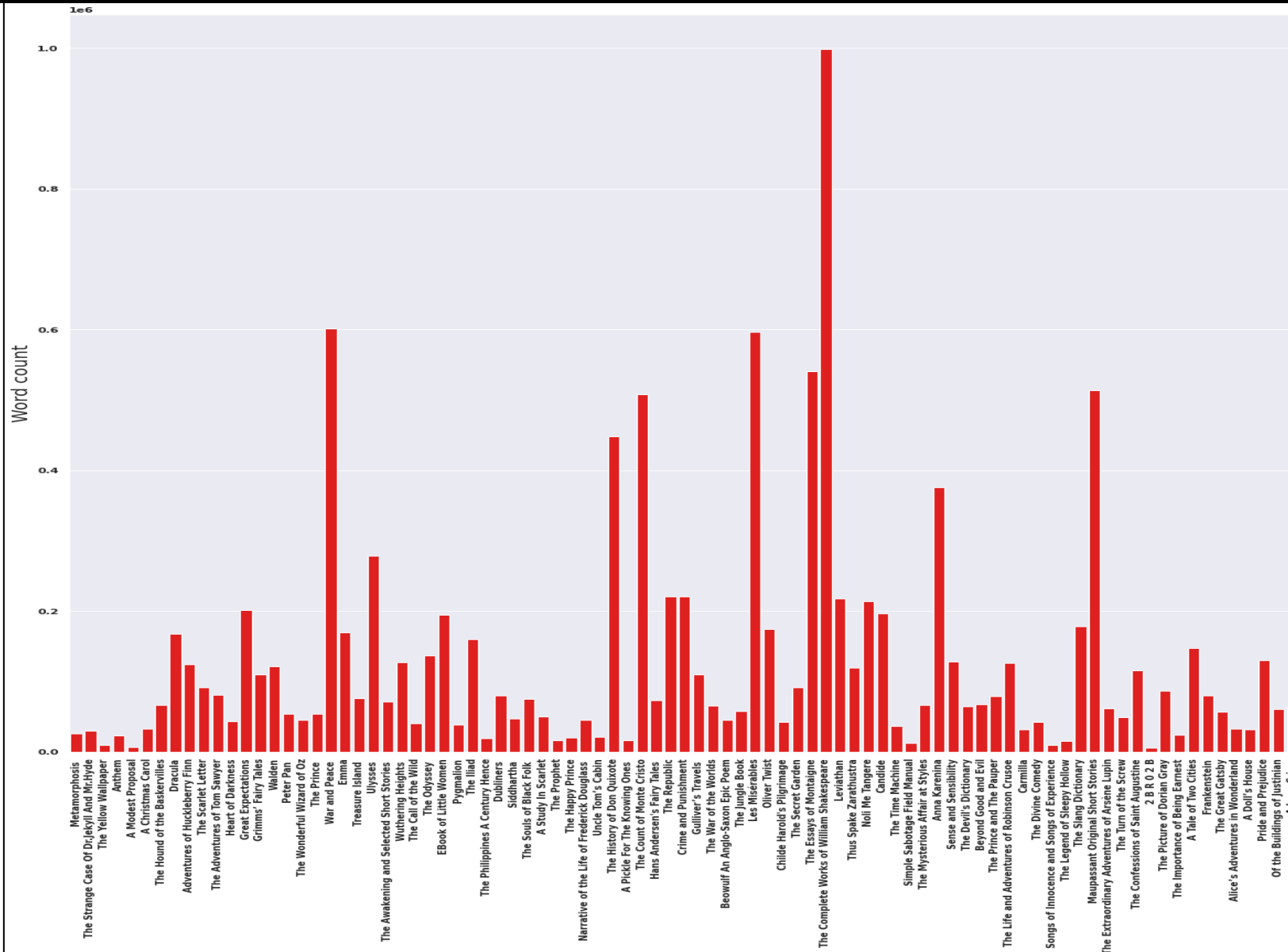
- End user for the project :

- Feature based or aspect-based sentiment analysis would be helpful to understand feedback and reviews from a website such as Goodreads.
- Someone in the early stages of buying a book may look for opinions and reviews online. These snippets of texts could be a gold mine for companies/authors that would want to monitor their reputation as well as get real-time feedback about the products / texts
- The direct beneficiaries of text sentiment analysis could be readers, marketing managers, public relations officials, authors of texts and online retailers.
- Such an analysis for a book recommendation system may be transferrable to other problems to provide movie recommendations or music recommendations (not in current scope).

- Data sources :

- Project Gutenberg – opensource access to eBooks, book reviews from Goodreads or online retailers
- Requires being able to access large datasets, have reasonable access to representative data to perform good analysis and to understand how ratings of a book may vary depending on the genre.
- Other data sources include reviews on manning.com etc., to process and run algorithms and large-scale computation as afforded by AWS.

# EARLY RESULTS : WORD COUNT



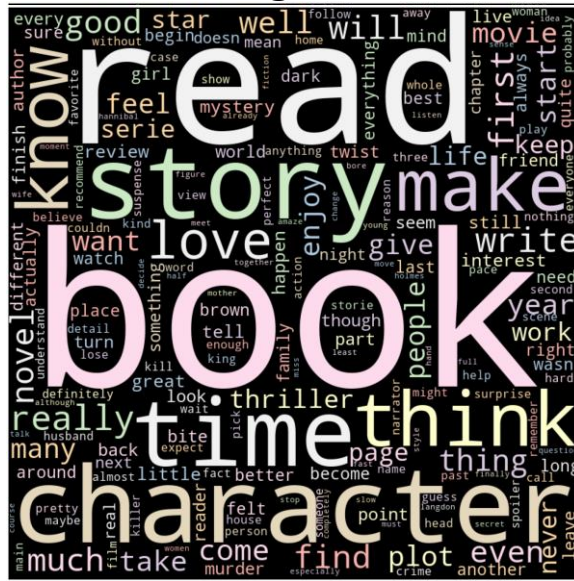
- Data from Gutenberg.org was analyzed to extract the word count of top 86 most downloaded books in the last 30 days.
- 40 MB of data (ASCII Text) was analyzed.
- As part of the analysis / cleaning up the data, Beautiful Soup for web scraping/parsing was used along with RegexpTokenizer.
- Further work is based on understanding correlation between rating or popularity of the book and the word count as well as frequency of words.

# EARLY RESULTS (part of speech tagging)

## Thriller genre

```
In [310]: print(book_reviews['review'])
0      \nI read this book due to the fact that everyo...
1      \nI just sent a 3 page, 3,000 word email about...
2      \nDepressing from start to finish. Most every ...
3      \n(A-) 83% | Very GoodNotes: It's about escapi...
4      \nHalf a "what the hell happened ?" star !! 20...
...
8646   \nIt was very interesting and exciting to fina...
8647   \npersonally, this's the first time i read thi...
8648   \nFavorites: "A Scandal in Bohemia""A Case of ...
8649   \nThis includes 12 short stories, some of whic...
8650   \nThat's how I like my Sherlock Holmes stories...
Name: review, Length: 8651, dtype: object
```

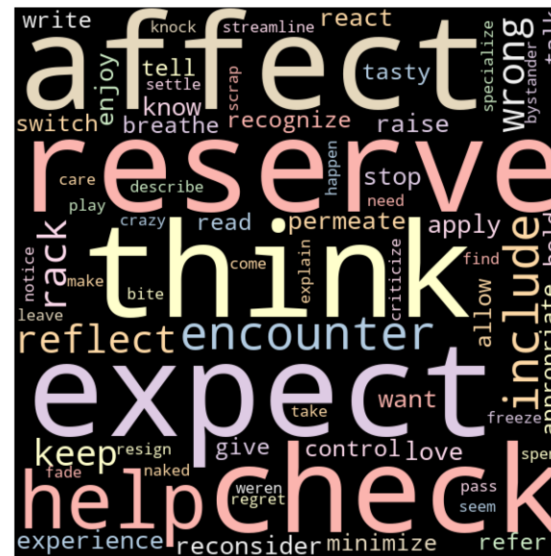
## Tag cloud



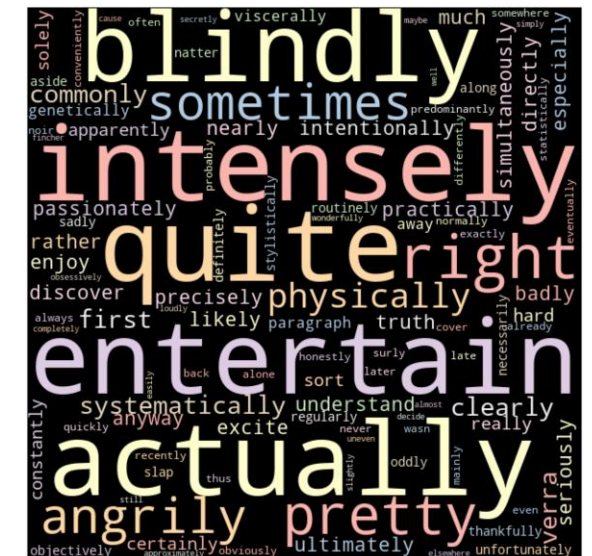
## The girl with the dragon tattoo

```
In [458]: print(book_reviews['review'])
0      \nSwedish people are nuts! I realize that's a ...
1      \n(A) 85% | ExtraordinaryNotes: It's a logical...
2      \nWomen are raised to routinely fear rape."If ...
3      \nAfter having leaped onto the bandwagon with ...
4      \nI HATED this book. I don't understand all of...
5      \nHe slapped her hard. Salander opened her eye...
6      \nThanks to the slew of Swedish and Hollywood ...
7      \nNordic noir of the highest order and very en...
8      \nThe Girl with the Dragon Tattoo ranks high a...
Name: review, dtype: object
```

## Tag cloud (nlTK pos tagger used to extract verb)



## Tag cloud (nlTK pos tagger used to extract adverb)





# EARLY RESULTS (part of speech tagging)

## Comics

```
In [18]: print(book_reviews['review'])
0      \nTom King's Vision is literary sci-fi, dark a...
1      \nI don't usually review any of the series I f...
2      \n"I am the astro-creep, a demolition-style he...
3      \nHas the feel of a horror or thriller movie u...
4      \nA lot darker in nature than I thought, but a...
...
3944   \nThis is a though provoking exploration of th...
3945   \nDNF @ like 45%I hated this so effing much. E...
3946   \nM an P are watching this ..... what a fab so...
3947   \nI thought I has already read this, fifteen o...
3948   \nI'm left speechless. Really enjoyed it mostl...
Name: review, Length: 3949, dtype: object
```

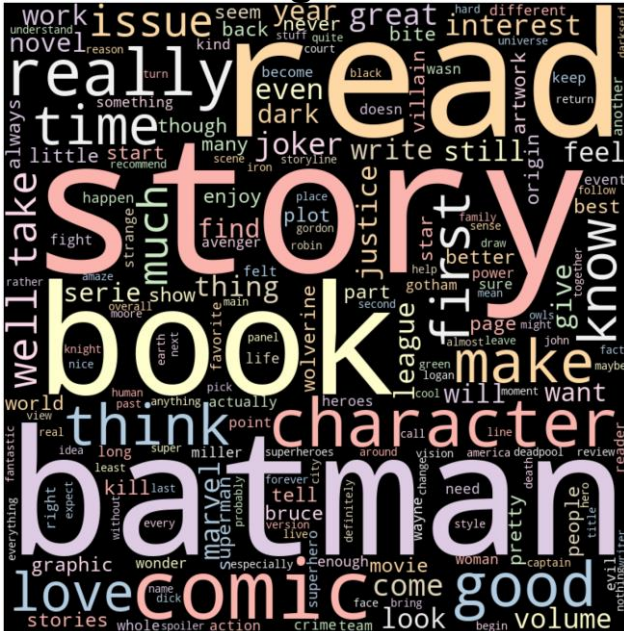
# Avengers, Captain Marvel

```

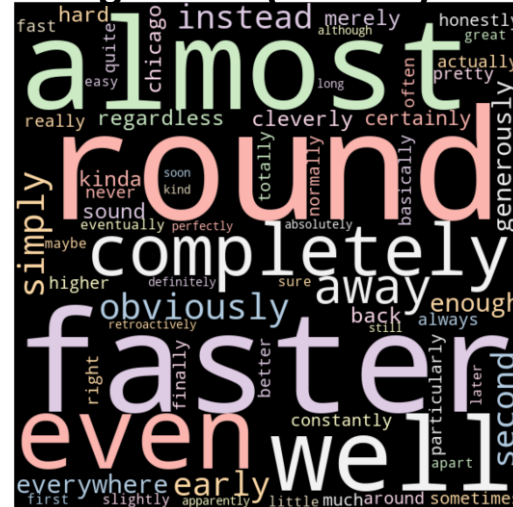
6         \nThis comic is very fantastic.\n
7         \n"Have you ever seen a little girl run so fas...
8         \nLike most comics, it just assumes you know w...
9         \nAWESOME! Love the crossover with Guardians o...
10        \nBullet Review:Wow, I apparently read this *1...
11        \nCaptain Marvel is kinda boring in my opinion...
12        \nWhat the hell did she need the Guardians of ...
13        \nCarol Danvers is badass and I can't wait to ...
14        \nDefinitely fun, but not really for beginners...
15        \nAww yeah. This was crazy fun and my first im...
16        \noverall 2.5 stars. in 32yrs iv read 8 comic...
17        \nI'm embarrassed to say the first time I actu...
18        \nFirst time reading any Marvel graphic novels...
19        \nI honestly thought this was very cute with a...
20        \nIt's a free and short comic, so complaints a...
21        \nDon't know wether to laugh or cry with this ...
22        \n*yawn*\n
23        \nPoignant AnswersGreat for clarifying certain...
Name: review, dtype: object

```

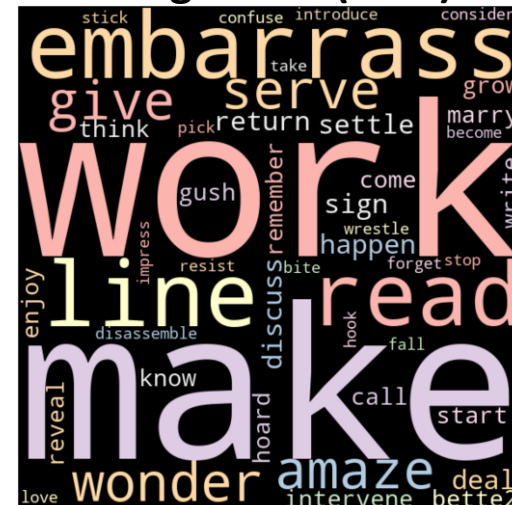
## Tag cloud



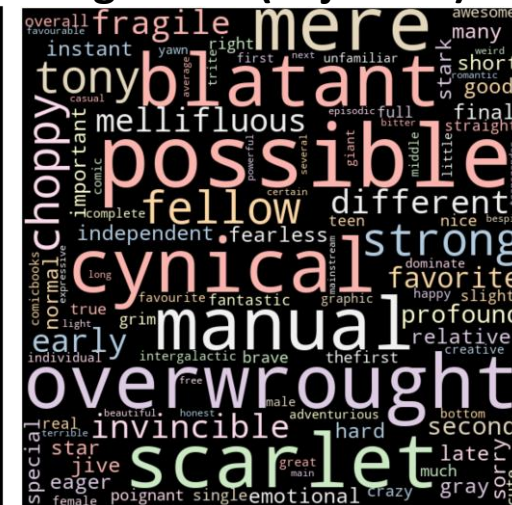
## Tag cloud (adverb)



### Tag cloud (verb)



### Tag cloud (adjective)



**Stanford pos tagger was used for the analysis**

# TAKEAWAYS

---

- This project proposes the use of text sentiment analysis to develop book recommendations.
- Existing approaches for sentiment analysis could be compared to better understand and quantify the accuracy as well as regression-based approaches could be explored as part of further work.
- The analysis will be helpful to understand how readers enjoy a book as well as determine their reaction, pick up on any aspects on sentiment or emotion that needs immediate attention, understand social impact of books.
- Continued analysis on user acceptance or A/B testing validation tools to understand possible improvement (to the application) could be explored.