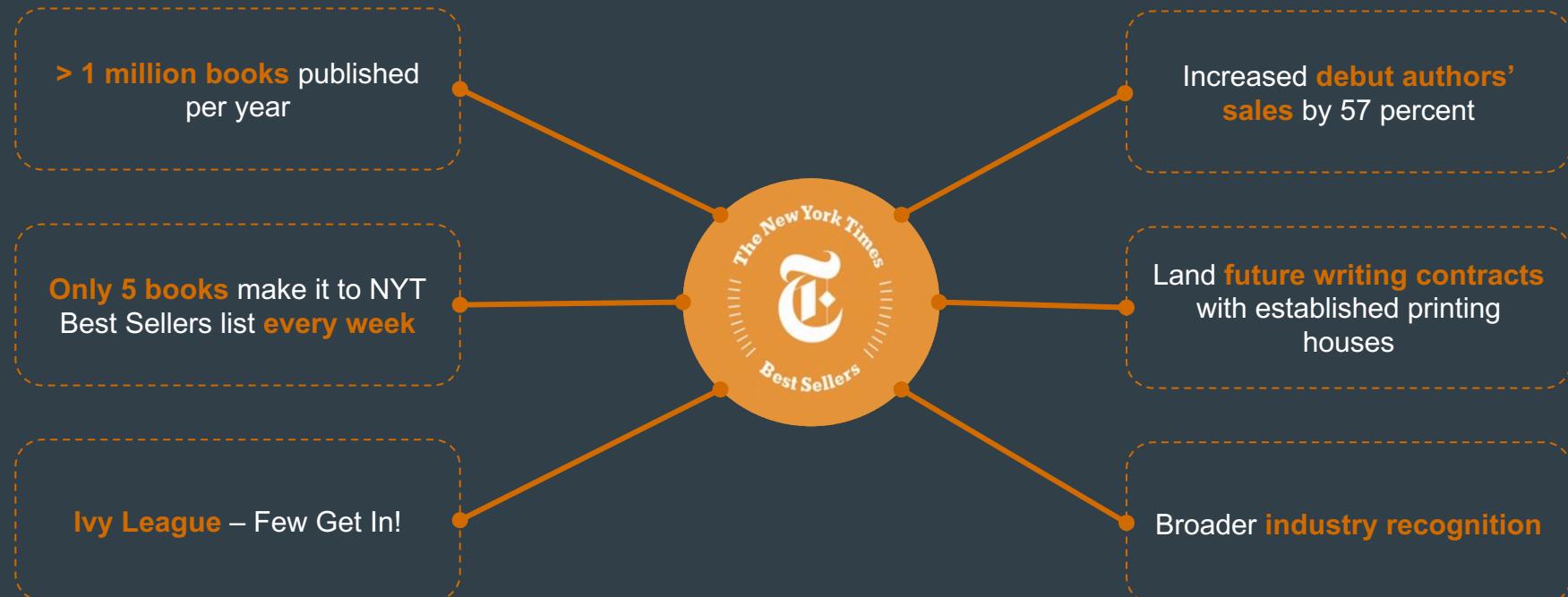




THE ART OF RECOMMENDING & WRITING A NEW YORK TIMES (NYT) BESTSELLING BOOK

Chetna Singhal, David Kinman, Prajval Gupta,
Subhayu Chakravarty, Whitt Hyde

NYT Best Sellers List - A “BIG DEAL”



Project Overview

MOTIVATION



How NYT compiles the bestseller list is a trade secret

Uses a proprietary method that combines sales figures, other data and internal guidelines that are unpublished



Solving this mystery using a goldmine of user reviews i.e. Goodreads

90 million members
2.6 billion books
90 million reviews

INSIGHTS EXPECTED



For Authors

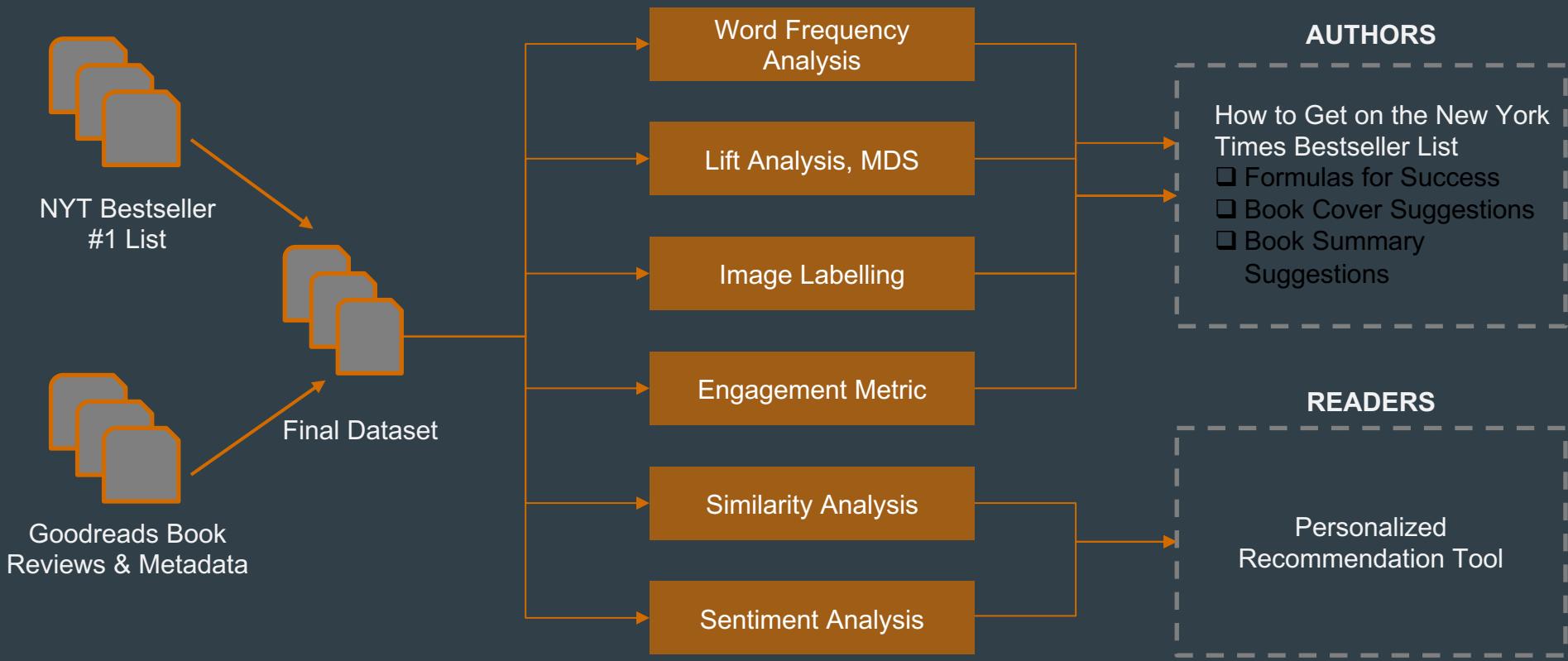
Identify distinguishing characteristics of a bestselling book in Fiction & Nonfiction categories



For Readers

Generate personalized recommendation using a combination of Goodreads overall ratings and user reviews

Analysis Blueprint



Scraping Highlights

50

#1 NYT Hardcover Best Seller **Fiction** Books from 2019 & 2018

50

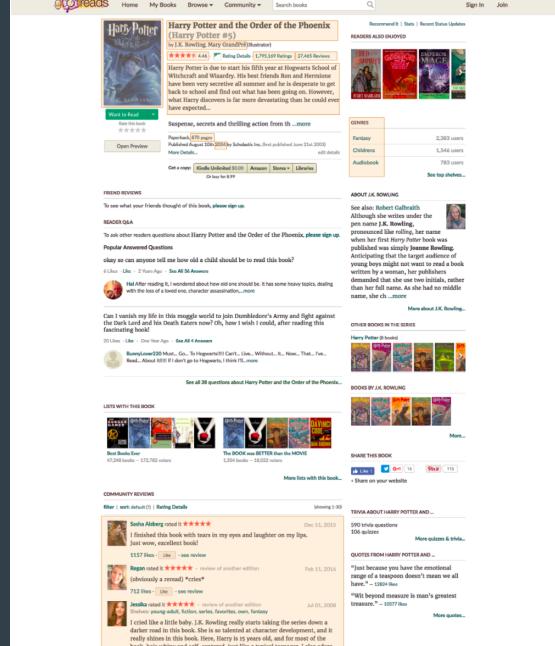
#1 NYT Hardcover Best
Seller **Nonfiction** Books
from 2019 & 2018

30,000

Goodreads User Reviews

12

Features extracted from Goodreads



User Review, User Rating

Data Preprocessing Highlights



Cloud Vision API

Used Cloud Vision API to gather labels on book cover

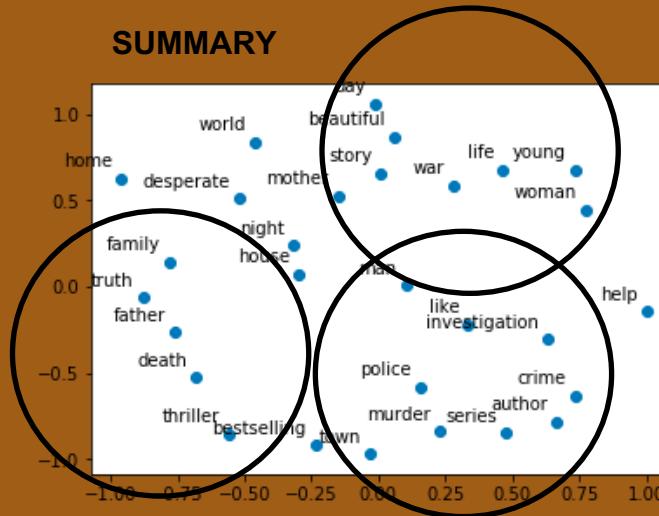
- Sent Book Cover URL
- Object function detected individual objects within a Cover
- Label function predicts what the entire image represents

$$\text{Engagement Score} = .4 \text{ ratingcount} + .6 \text{ reviewcount}$$

We created an engagement metric for all books.

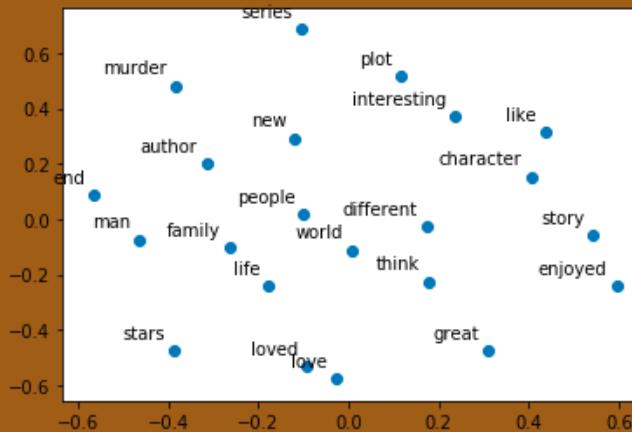
- Weighted reviews higher
- Compared books engagement as high or low based on median engagement

Author Insights For Fiction Books

SUMMARY

Themes that ruled the charts:

- Murder series
- War stories involving young woman
- Family stories (Father-centered)

REVIEWS

 Murder series more likely to succeed than Love series

Murder stories:

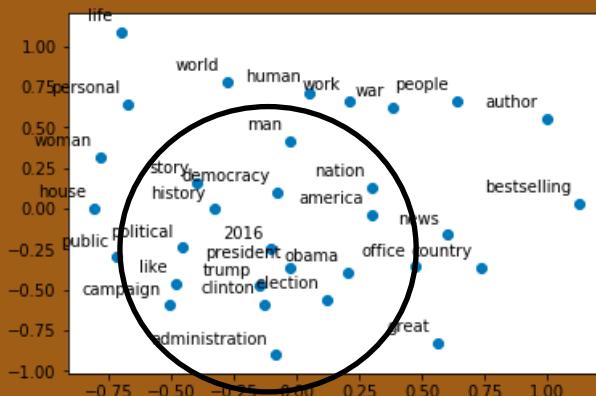
Interesting plot captures more attention than Likable characters

Love stories

Novelty ("think", "different") generates higher engagement

Author Insights For Nonfiction Books

SUMMARY

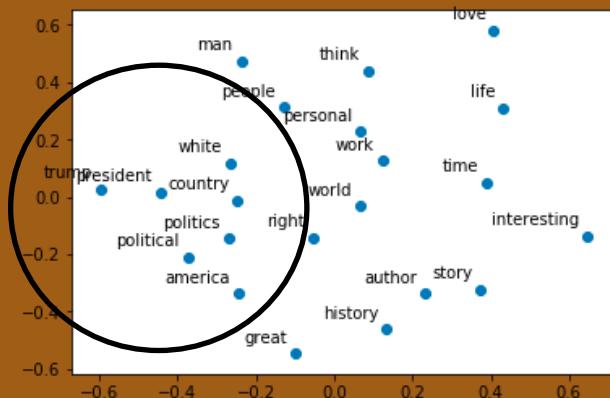


American politics ruled the NYT charts

Personal biographies of "woman" more likely to succeed than personal biography of "man"

Buzz/Interest around **2016 elections** is not going away anytime soon

REVIEWS

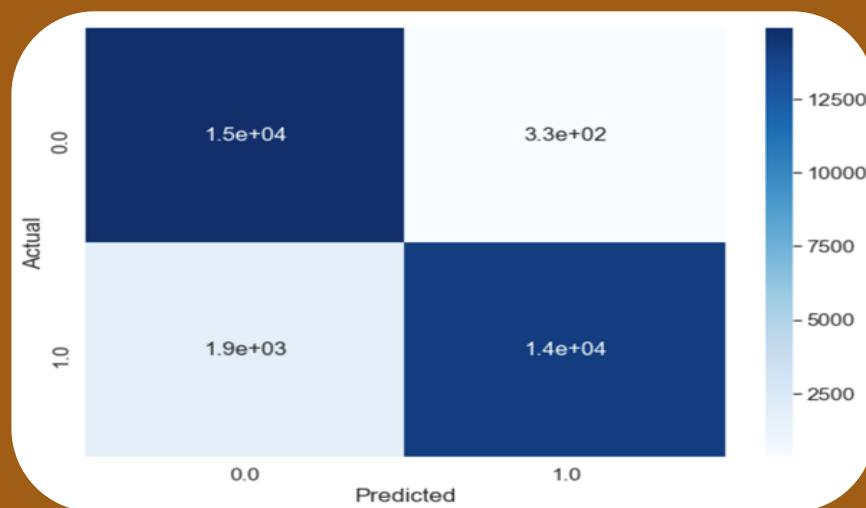


“Politics & Trump” generated more engagement than “Politics & Obama” or “Politics & Clinton”

It's a "man's world" – woman not associated with any particular theme in user reviews

Author Insights For Fiction Books II

Logistic regression model predicted engagement with 84.3% accuracy using cover labels & summaries having high TF-IDF



Features that drove engagement in the model:

- "Nature" drove engagement the highest of any singular cover label
- Cover Labels of weapons, broken glass, and mystery
- Summaries that mention historic, or history do better in engagement



Historical Fiction was a leader in the topic, leader engagement comparison. As well as reviews. War stories in particular.



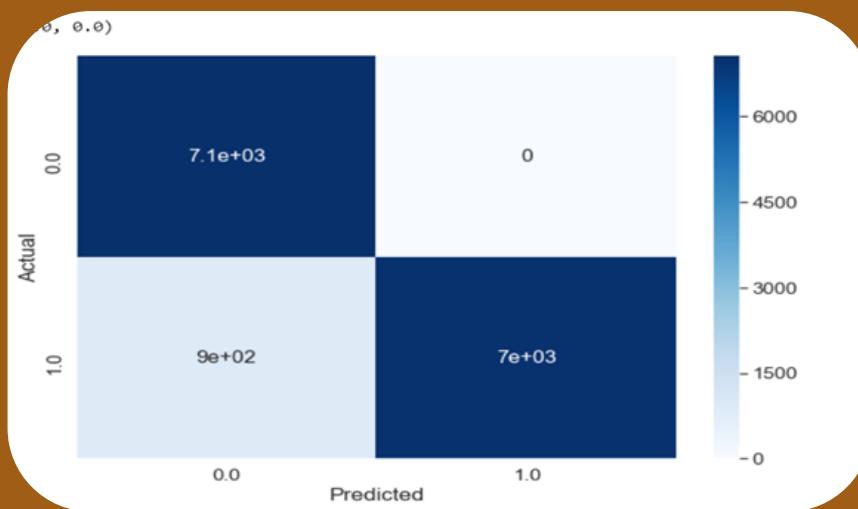
Nature preformed well both in summary and cover analysis .



Bright Colors did poorly in terms of driving engagement. Darker labels stole the show

Author Insights For Nonfiction Books II

Logistic regression model predicted engagement with 93.8% accuracy **using cover labels & summaries having high TF-IDF**



Features of Highest importance for the model:

- Trump was the singular most important feature
- Summaries involving politics were the most engaging
- Other than politics and biographies, health was the only other real influence on engagement
- Note: The model never falsely predicted engagement



Trump good or bad opinion of the man, drives engagement.



Politics at the forefront of the topic engagement, perhaps due to upcoming election, this may be temporal.

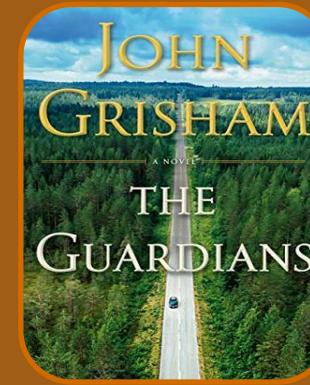


Very hard to break into NYT non-fiction list without being a biography or Politics-based

So, how do You become a NYT Bestseller in 2020?

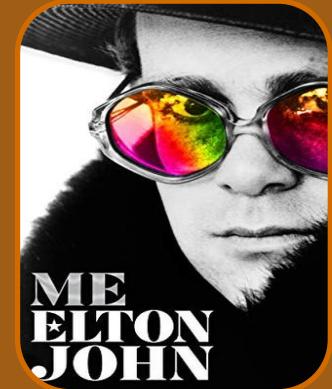
FICTION

- ✍ For any genre in Fiction, try to include a cover with Greenery, Mountains or Nature.
- ✍ No Flowers, because bright Colors are poorly associated with Bestseller Performance



NONFICTION

- ✍ Write about Donald Trump, or a Famous Person
- ✍ If applicable, have a Face On Your Cover



Reader Recommendations For Fiction Books - I

User Attributes:

MAGIC, FANTASY, ADVENTURE

Based on average overall rating

| Title | Author | Overall_Rating |
|-------------------------|-------------------|----------------|
| Oathbringer | Brandon Sanderson | 4.62 |
| Where the Crawdads Sing | Delia Owens | 4.50 |
| The Night Fire | Michael Connelly | 4.41 |
| A Better Man | Louise Penny | 4.39 |
| Leverage in Death | J.D. Robb | 4.35 |

Based on Similarity + Sentiment Score

| Title | Author | Combined | GenRev_Similarity_Score |
|----------------------|-----------------|----------|-------------------------|
| Circe | Madeline Miller | 0.69 | 0.76 |
| Twisted Twenty-Six | Janet Evanovich | 0.45 | 0.74 |
| The Rising Sea | Clive Cussler | 0.53 | 0.74 |
| The Fall of Gondolin | J.R.R. Tolkien | 0.62 | 0.72 |
| Texas Ranger | James Patterson | 0.85 | 0.72 |

Based on Overall rating + Similarity + Sentiment scores

| Title | Author | Sim_Sent_Overall_Rating |
|-------------------------|-----------------|-------------------------|
| The Woman in the Window | A.J. Finn | 12.23 |
| The Cast | Danielle Steel | 12.04 |
| Texas Ranger | James Patterson | 11.70 |
| Circe | Madeline Miller | 11.53 |
| The New Girl | Daniel Silva | 11.36 |

Reader Recommendations For Nonfiction Books - I

User Attributes:

PERSONAL, PRESIDENT, TRUMP

Based on average overall rating

Based on Similarity + Sentiment Score

Based on Overall rating + Similarity + Sentiment scores

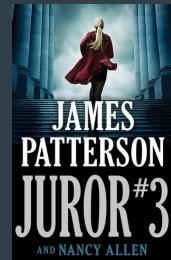
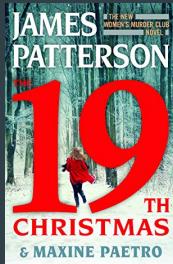
| | Title | Author | Overall_Rating |
|---|-----------------|-------------|----------------|
| | Grant | Ron Chernow | 4.54 |
| Call Sign Chaos: Learning to Lead | Jim Mattis | | 4.43 |
| Blowout: Corrupted Democracy, Rogue State Russ... | Rachel Maddow | | 4.45 |
| The Russia Hoax: The Illicit Scheme to Clear H... | Gregg Jarrett | | 4.39 |
| Beastie Boys Book | Michael Diamond | | 4.56 |

| | Title | Author | Combined | GenRev_Similarity_Score |
|--|---|------------------|----------|-------------------------|
| | Rediscovering Americanism: And the Tyranny of ... | Mark R. Levin | 0.87 | 0.72 |
| | Triggered: How the Left Thrives on Hate and Wa... | Donald Trump Jr. | 0.80 | 0.69 |
| | The Soul of America: The Battle for Our Better... | Jon Meacham | 0.64 | 0.73 |
| | I Can't Make This Up: Life Lessons | Kevin Hart | 0.60 | 0.64 |
| | Old School: Life in the Sane Lane | Bill O'Reilly | 0.57 | 0.68 |

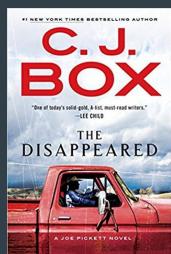
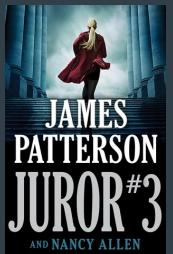
| | Title | Author | Sim_Sent_Overall_Rating |
|--|---|------------------|-------------------------|
| | Rediscovering Americanism: And the Tyranny of ... | Mark R. Levin | 12.05 |
| | The Soul of America: The Battle for Our Better... | Jon Meacham | 11.10 |
| | Becoming | Michelle Obama | 10.88 |
| | Triggered: How the Left Thrives on Hate and Wa... | Donald Trump Jr. | 10.87 |
| | Portraits of Courage: A Commander in Chief's T... | George W. Bush | 10.74 |

Reader Recommendations For Fiction Books - II

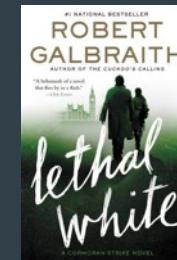
LOVE



MURDER



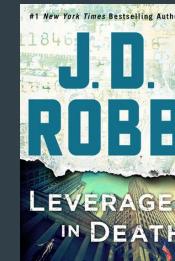
INTERESTING



User Attributes

MAGIC
FANTASY
ADVENTURE

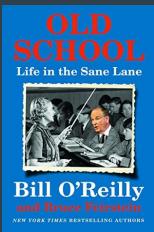
DIFFERENT



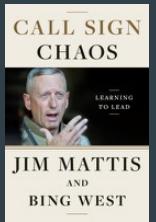


Reader Recommendations For Nonfiction Books - II

LIFE



FAMILY



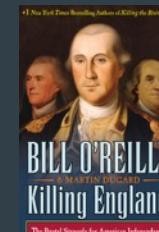
User Attributes

PERSONAL
PRESIDENT
TRUMP

INTERESTING



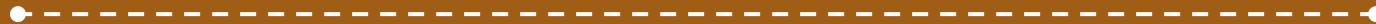
DIFFERENT



Reader Recommendations – Explanation



Avid readers do not select a book to buy/read based only on overall rating of the book on Goodreads because they are not the **RIGHT “FIT”**. They end up reading as many user reviews as possible for that book(s) before they are convinced that they would like it.



Therefore, we **developed this tool** which would help such readers identify the best book to read by using 3 user-given attributes/themes (simply words) and performing sentiment analysis on all the user reviews that are most similar to these attributes. Next, we **refined the recommendation engine** by using preset words like interesting, love, story etc. to find sub-clusters of user interests. This gives us even better results for each theme and combination of above preset words.



So **our final recommendation** is based on Similarity score, Sentiment score around preset words - book, story, series. This gives us which “story” or “book” users have written most positively about and **shortens the reader’s search through all reviews for his/her next read**.

THANK YOU



The right book in the right hands at the right time
can change the world.