Harry Potter Recommendations

By: Aksha

Table of contents

01 **Context**
My Problem. My Hypothesis

02 **Plan**
My Plan of Action

03 **Progress**
What I did.

04 **Improvement**
Places of Improvement

05 **Questions**

01
Context

What Book
should I
Read?

What's your favourite HP Book?
And why?
?

02

Plan

A timeline always works fine

Collect Reviews

Web Scraping

Favourite Book

Sentiment Analysis

Identify Themes

Topic Modelling

More Books!

Recommendation Engine

03

What I did

A timeline always works fine

Collect Reviews

Web Scraping

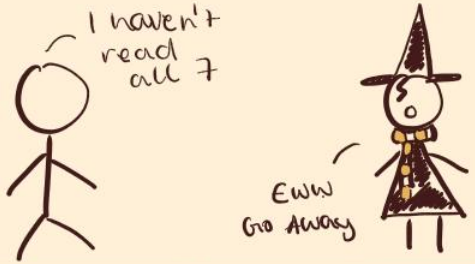
(1) Review Collection

Web Scraping

Book | Reviews | Rating

Assumption

7 Reviews for every person



A timeline ~~always~~ works fine

~~**Favourite Book**~~

How much do you like the book?

Sentiment Analysis

Collect Reviews

Web Scraping

(2) Sentiment Analysis

VADER

C******

While this is a good piece of fiction, the spiritual landscape in which it takes place REALLY has me taking a second look at this series.

★ ★ ☆ ☆ ☆

D*******

Come along with me on my re-read of the epically magical series “Harry Potter”. 🤓⚡

★ ★ ★ ★ ★

T*******

"Nitwit! Blubber! Oddment! Tweak! Thank you!"

★ ★ ★ ★ ★

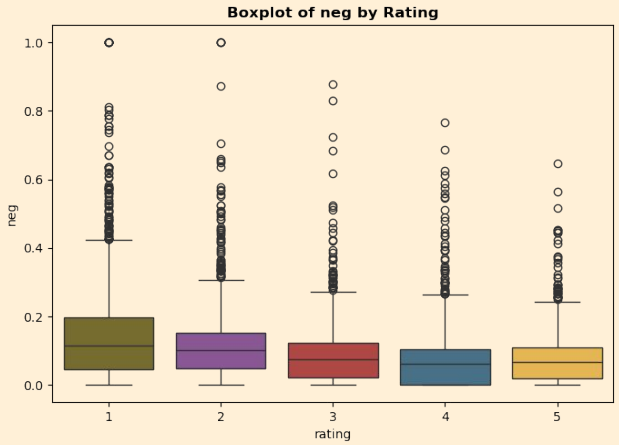
VADER Polarity Scores

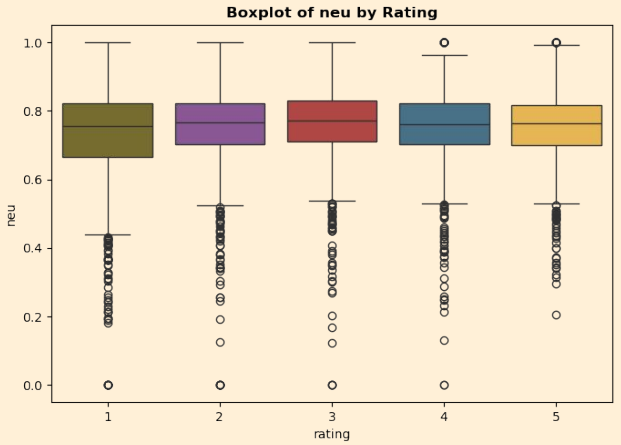
neg → [0,1]

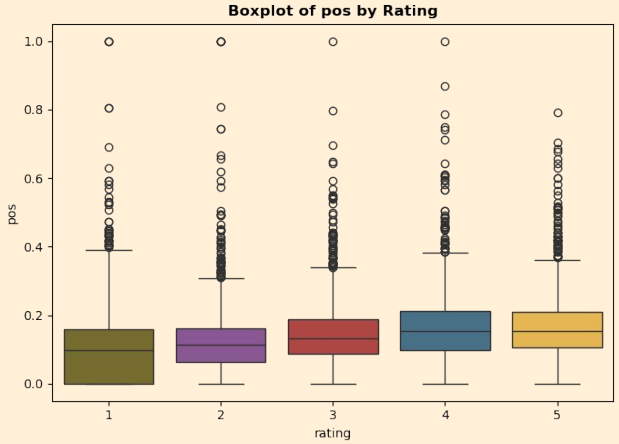
neu → [0,1]

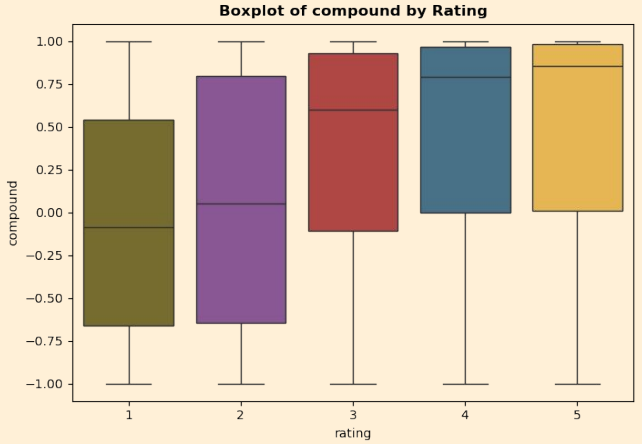
pos → [0,1]

compound → [-1,1]









Voldemort

2 reviews
988 followers

Follow

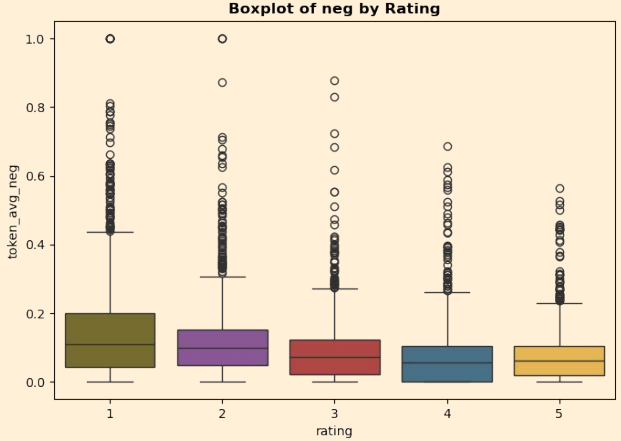
★☆☆☆☆

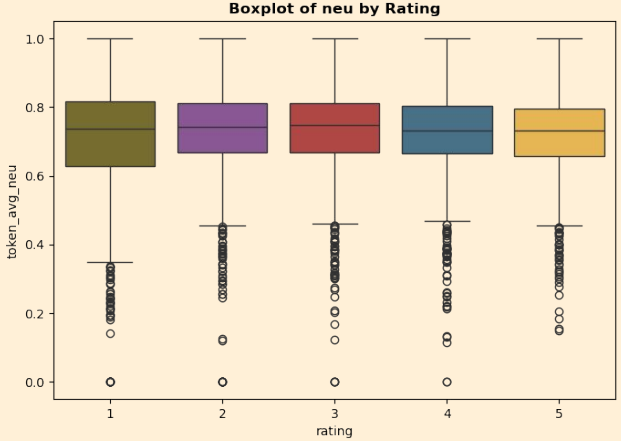
August 15, 2013

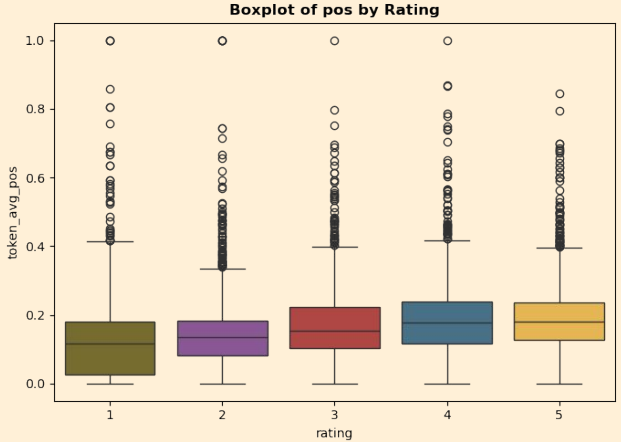
This is a disgrace to all pure bloods. My head looked quite dashing on that strange professor's head.. If only I could have extended my stay. Harry Potter.. This is not over yet.

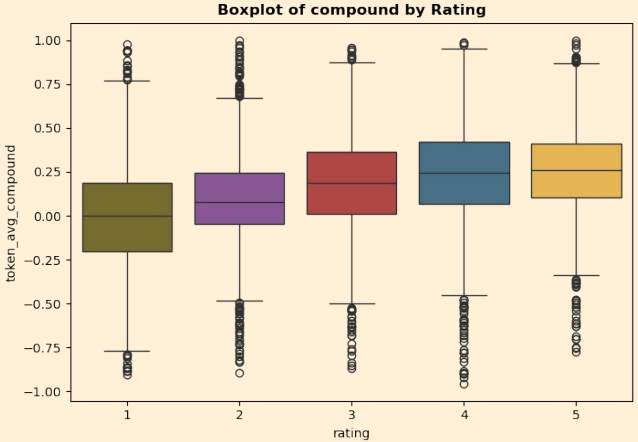
9,266 likes • 34 comments

Like Comment •••









- **Tokenise (Sentences)**
- **Add Words to Lexion**

A timeline works fine

~~**Favourite Book**~~

How much do you like the book?

Sentiment Analysis

Collect Reviews

Web Scraping

Identify Themes

Topic Modelling

(3) Topic Modelling

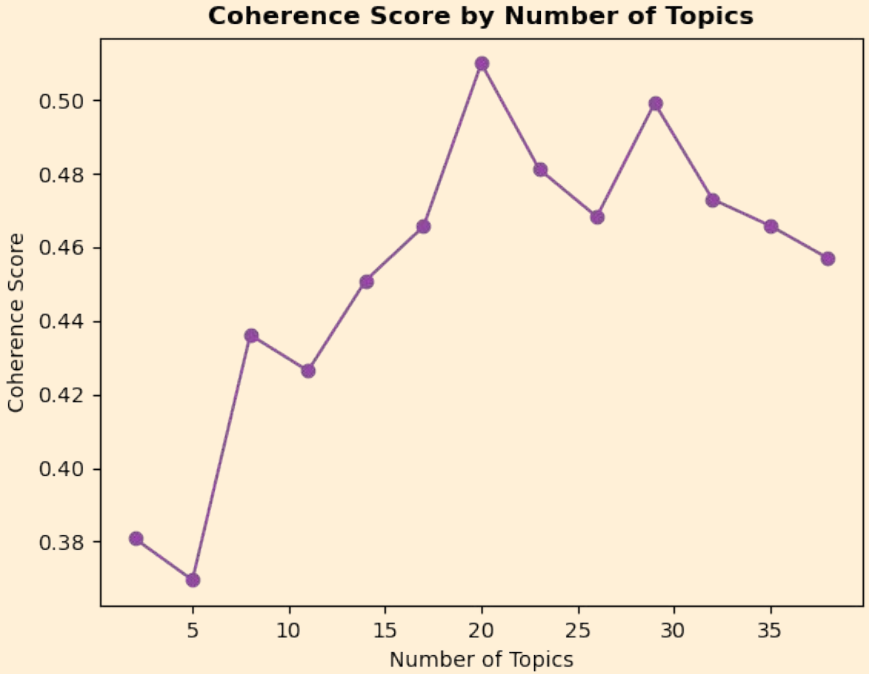
Latent Dirichlet Allocation (LDA)

- Looks at all the documents (reviews)
- Identifies which topics there are
- Identifies how many topics there are

PyLDAvis (Visualise!!!!)

- **Topic Circle:** size = importance
- **Distance:** how closely related
- **Relevance:** helps us what the topic is about

Coherence Score

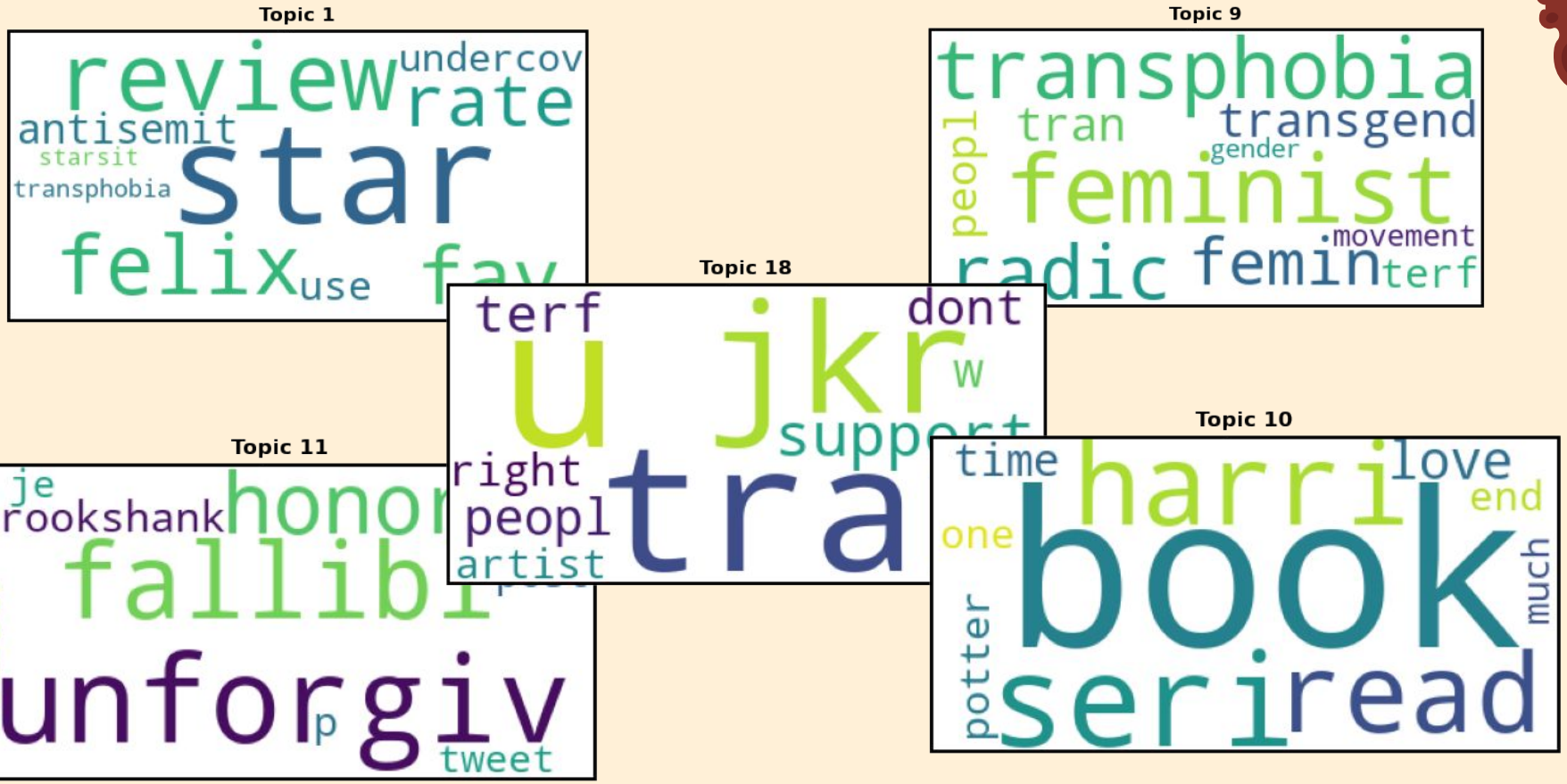


- Measures Interpretability of Topics
- Many different ways to calculate
- High score = Good

0.381 5 Topics

0.510 20 Topics

Note: Tokens = Words, not Sentences here


Topic 1
review
undercov
rate
antisemit
starsit
star
transphobia
felix
use
Topic 9
transphobia
tran
transgend
gender
peopl
feminist
movement
radic
femin
terf
Topic 18
terf
dont
u
jkr
w
support
right
tra
peopl
artist
Topic 11
je
honor
crookshank
fallibl
scabber
unforgiv
p
tweet
Topic 10
time
harri
love
end
one
book
potter
much
seri
read

What do I have?

Review	**Web Scraped**	Data Collected From GoodReads
Sentiment Label	**Sentiment Analysis**	Used VADER to get a polarity score, then used the score to label positive, negative, neutral
Topics	**Topic Modelling**	Used LDA to get topics for each review
Other	**Web Scraped**	More data collected from GoodReads: Ratings|Likes|Images...

A timeline works fine

~~Favourite Book~~

How much do you like the book?

Sentiment Analysis

More Books!

Recommendation Engine

Collect Reviews

Web Scraping

Identify Themes

Topic Modelling

04

Improvements

(4) Improvements

Feature Engineering

Other Languages; Typos; Contractions

More Data

Use more data for better results.

Bigrams/Trigrams

Use bigrams or trigrams instead of word tokenisation for modelling.

Questions

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon** and infographics & images by **Freepik**