Spoiler free movies recommender

"DON'T RUIN OTHER PEOPLE'S ENJOYMENT, YOU EVIL PERSON."

—Humanity

01

EDA

"I could do this all day"
-Captain America: Civil War

+ 500,000

Movie reviews dataset
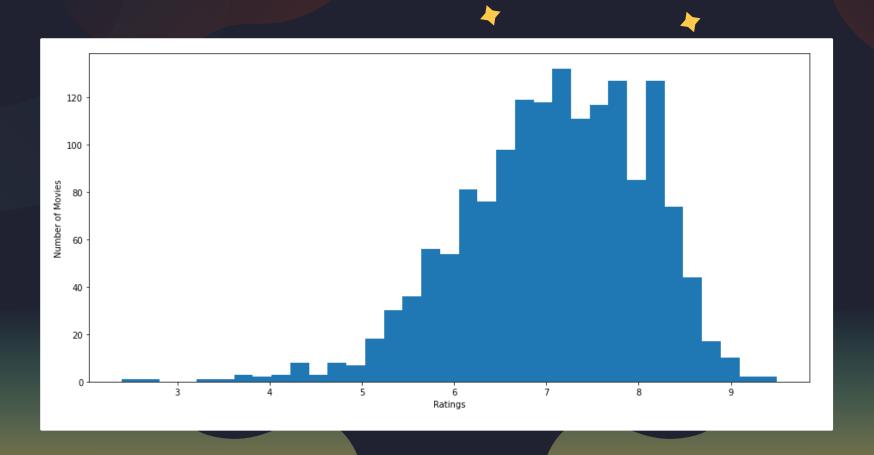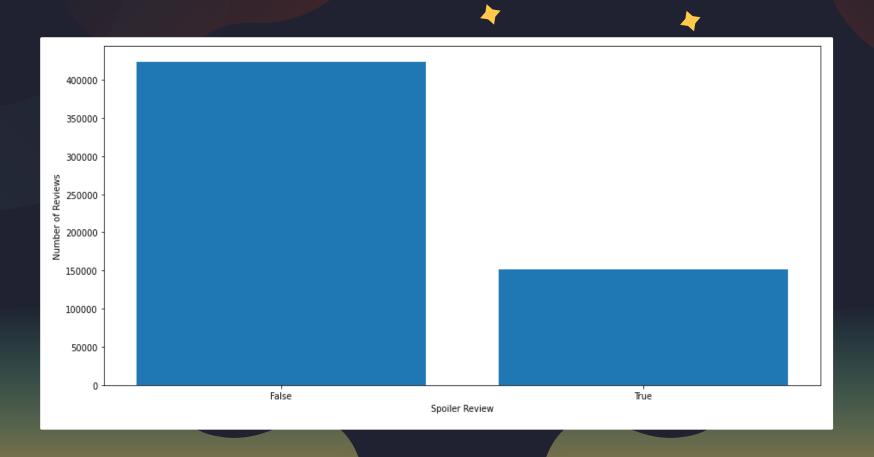(movie_id, user_id, is_spoiler,
review_text, review_summary)

+ 1,500

Movies details
dataset(movie_id,
plot_summary,
plot_synopsis)

API

IMDB API

Words most likely to be not spoilers

Words most likely to be spoilers

Top spoiler	Characteristic
spoilers	overrated
*	underrated
spoiler	masterpiece
if	tarantino
decent	disappointing
disappointing	gripping
your	flick
life	spoilers
way	remake
there	superhero
sequel	mediocre
tale	cinematic
by	sequel
worst	spielberg
	entertaining

Top not spoiler	spoiler
classic	awful
long	disappointment
fantastic	predictable
see	flawed
another	enjoyable
made	boring
movies	pointless
years	pixar
go	thriller
masterpiece	hilarious
much	brilliant
awesome	movie
brilliant	horrible
hell	visually

Labels within the plot (spoiler Frequency vs not spoiler Frequency):

spoilers, the, of, not, film, a, ever, was, very, to, from, about, s, t, great, all, time, *, disappointing, way, life, if, out, by, your, at, has, up, made, movies, much, masterpiece, tale, sequel, series, do, there, will, films, decent, spoiler, who, some, flick, how, we, or, horror, get, brilliant, see, every, people, wonderful, 10, fi, think, again, original, special, world, ending, into, did, his, year, new, classic, disappointed, **, then, where, hard, here, disney, day, quite, thriller, waste, emotional, he, bit, incredible, dark, makes, two, star, d, comic, heart, watching, disappointment, let, m, acting, awesome, garbage, re, without, full, gripping, done, hilarious, old, stunning, far, almost, simply, long, magic, superb, ll, other, ok, feel, poor, true, off, say, must, western, crap, powerful, high, does, american, fantastic, journey, genre, different, wrong, last, remake, go, though, visually, am, need, little, scary, her, god, start, expected, thing, modern, got, solid, war, sad, saw, had, once, flawed, oscar, being, real, mind, top, hollywood, another, message, want, effects, surprisingly, franchise, didn, sweet, violent, come, trek, laugh, romantic, would, nice, game, before, moving, kids, oh, critics, piece, money, now, script, look, many, 1, script, cinematic, 3, after, any, okay, absolutely, truly, family, short, side, whole, were, black, everything, overall, know, character, king, lot, works, which, since, years, jones, art, boy, possibly, ride, hell, cast, flaws, marvel, half, pixar, always, audience, second, work, clever, girl

Axis labels: spoiler Frequency (vertical: Frequent, Average, Infrequent); not spoiler Frequency (horizontal: Infrequent, Average, Frequent)
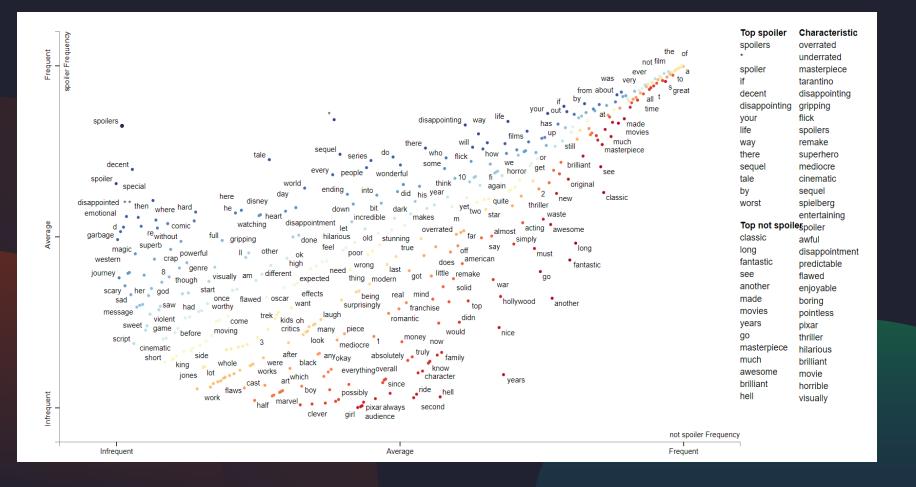
02

Sentiment Analysis

"That's my secret, Captain: I'm always angry."
-The Avengers

Sentiment Analysis

Flair

NLP library pretrained sentiment analysis model on IMDB dataset.

VADER

lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media.
It is used for sentiment analysis of text which has both the polarities

03 Data Cleaning

"You shall not pass!"
- Lord of the Rings

Data Cleaning

TOKENIZATION

breaking up a piece of text into smaller parts

STOP WORDS

drop the words that are present in text but do not contribute to the meaning of a sentence

LEMMATIZATION

extract the base form of the words by using WordNetLemmatizer

05

Recommendation Systems

"May the odds ever be in your favor."
-The Hunger Games

Recommendation Systems

Two recommendation approaches:

Get movies similar to a movie

Get personal recommendations

06

EXTRA Work
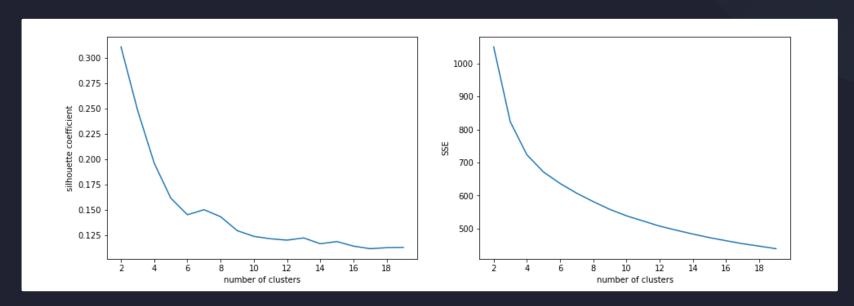
"Are you not entertained?"
- Gladiator

Topic Modeling

Clustering

Topic Modeling

LDA

```
[(0,
 '0.002*"family" + 0.002*"school" + 0.002*"life" + 0.002*"man" + 0.002*"young" + 0.001*"team" + 0.001*"agent" + 0.001*"high" + 0.001*"mission" + 0.001*"time"'
(1,
 '0.002*"life" + 0.002*"man" + 0.002*"new" + 0.002*"friend" + 0.002*"meet" + 0.002*"time" + 0.001*"try" + 0.001*"know" + 0.001*"house" + 0.001*"love"'),
(2,
 '0.003*"life" + 0.002*"new" + 0.002*"know" + 0.002*"work" + 0.001*"try" + 0.001*"time" + 0.001*"world" + 0.001*"make" + 0.001*"year" + 0.001*"plan"'),
(3,
 '0.002*"father" + 0.002*"life" + 0.002*"new" + 0.002*"man" + 0.001*"year" + 0.001*"make" + 0.001*"world" + 0.001*"meet" + 0.001*"begin" + 0.001*"live"'),
(4,
 '0.003*"life" + 0.002*"year" + 0.002*"friend" + 0.002*"old" + 0.002*"meet" + 0.002*"want" + 0.002*"help" + 0.001*"day" + 0.001*"man" + 0.001*"world"'),
(5,
 '0.004*"life" + 0.002*"friend" + 0.002*"year" + 0.002*"new" + 0.002*"time" + 0.002*"want" + 0.002*"way" + 0.001*"leave" + 0.001*"work" + 0.001*"love"'),
(6,
 '0.002*"life" + 0.002*"new" + 0.002*"set" + 0.002*"turn" + 0.002*"world" + 0.001*"day" + 0.001*"make" + 0.001*"love" + 0.001*"child" + 0.001*"young"'),
(7,
 '0.002*"life" + 0.002*"love" + 0.002*"father" + 0.002*"woman" + 0.001*"try" + 0.001*"want" + 0.001*"young" + 0.001*"friend" + 0.001*"world" + 0.001*"year"')
(8,
 '0.004*"life" + 0.003*"year" + 0.002*"new" + 0.002*"love" + 0.002*"friend" + 0.002*"family" + 0.002*"try" + 0.001*"young" + 0.001*"meet" + 0.001*"know"'),
(9,
 '0.002*"life" + 0.002*"year" + 0.002*"young" + 0.001*"make" + 0.001*"day" + 0.001*"agent" + 0.001*"world" + 0.001*"new" + 0.001*"friend" + 0.001*"work"'),
(10,
 '0.003*"new" + 0.003*"life" + 0.003*"world" + 0.002*"friend" + 0.002*"try" + 0.002*"year" + 0.002*"man" + 0.002*"young" + 0.002*"kill" + 0.002*"family"'),
(11,
 '0.003*"life" + 0.002*"make" + 0.001*"friend" + 0.001*"work" + 0.001*"world" + 0.001*"decide" + 0.001*"plan" + 0.001*"year" + 0.001*"try" + 0.001*"leave"'),
(12,
```

Rcommendations

Using Flair instead of Vader

For 100,000 row, Vader=1H, flair=13H

Find different methods for Topic Modeling

Use wights on the most mentioned words in spoilers

THANK YOU!

"Chewie, we're home."
- Star Wars