

The background of the slide is a composite image. On the left, there is a close-up of a film reel with its sprocket holes visible. On the right, there is a clapperboard with the words 'PRODUCTION', 'DIRECTOR', 'CAMERA', 'SCENE', and 'TAKE' printed on it. A large, faint watermark of a hand is visible across the center of the image.

Movie&TV Reviews NLP

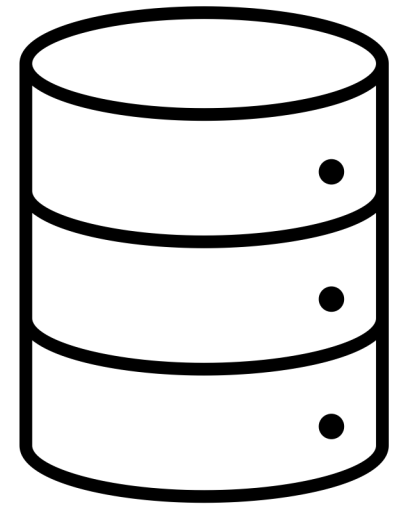
Rui Yuan

Clients and Problem

- Amazon wants to investigate customer reviews and improve products/services
- My focus is on the Movie&TV department
- The NLP goal is to find out the overall customer sentiments and topics customer care about

Data

- Data accessed through UCSD data repository, under the Amazon product session
- 16,905 rows of customer reviews on the Amazon movie&TV
- Features: customer ID, rating (1-5), review text



Work Flow and Algorithms

EDA and Data Visualization:

- Matplotlib, Seaborn

Text Preprocessing

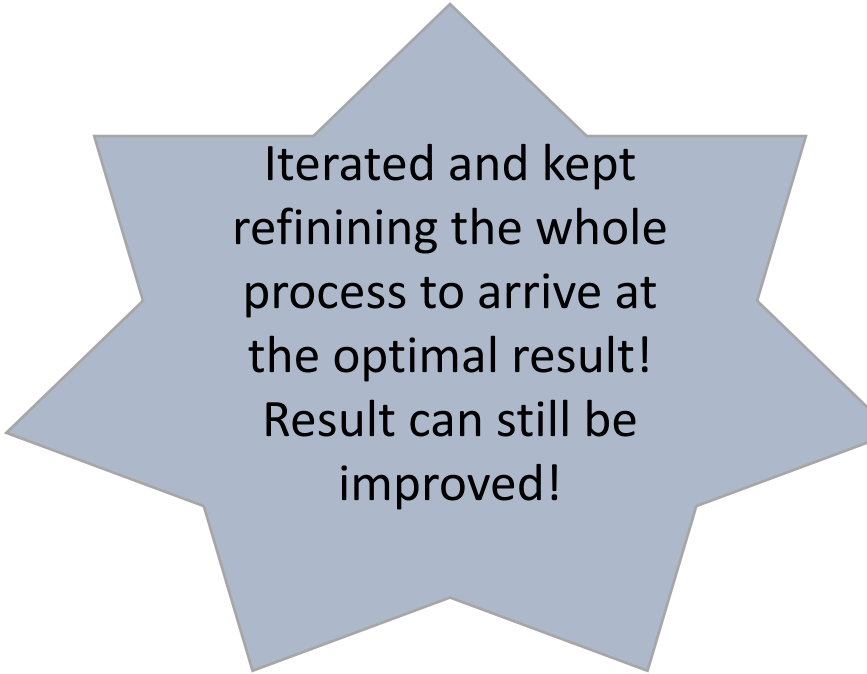
- NLTK used including word tokenization, lemmatization, pos tagging, stop words removal
- Filtered words down to nouns, adjectives, verbs and adverbs

Sentiment Analysis:

- VADER, TextBlob used and settled with VADER

Topic Modeling:

- With only nouns
- LDA, LSA, NMF, CorEx used and settled with NMF



Iterated and kept refining the whole process to arrive at the optimal result! Result can still be improved!

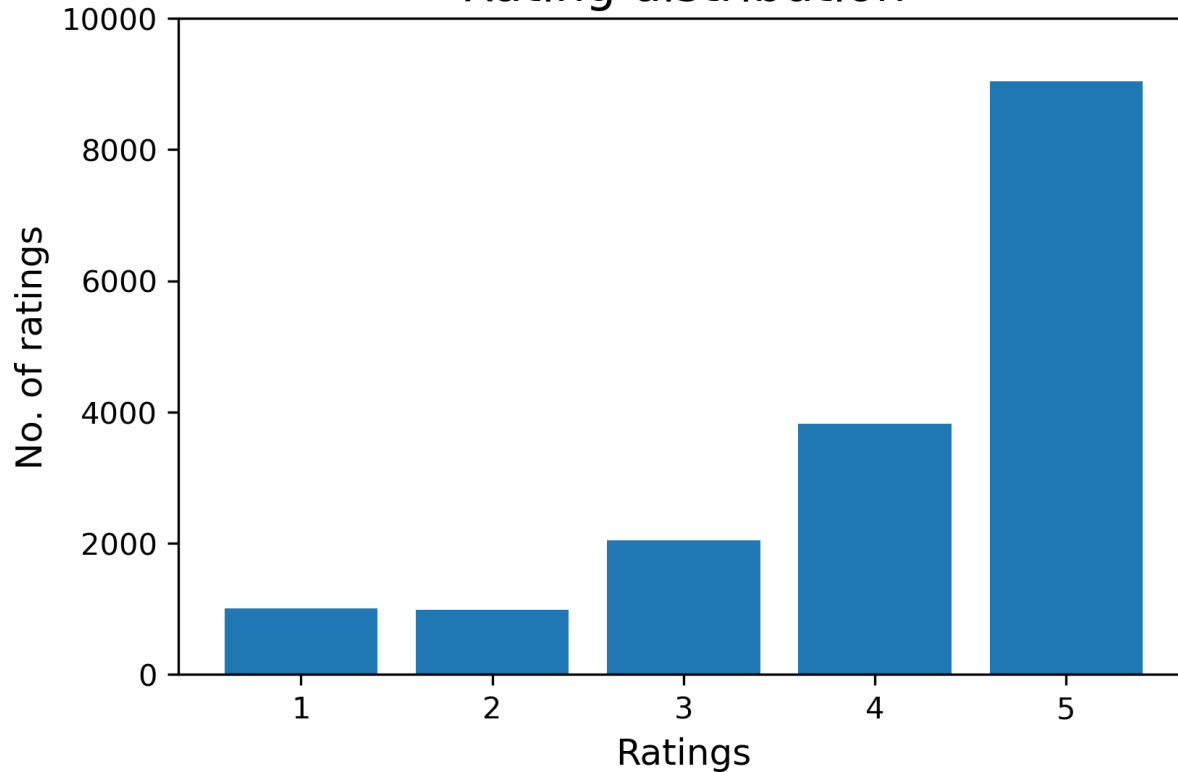
Common Words Visualization: WordCloud



- Positives more than negatives
- Movie elements: story, series, character, action, music, plot, role, performance, scene...

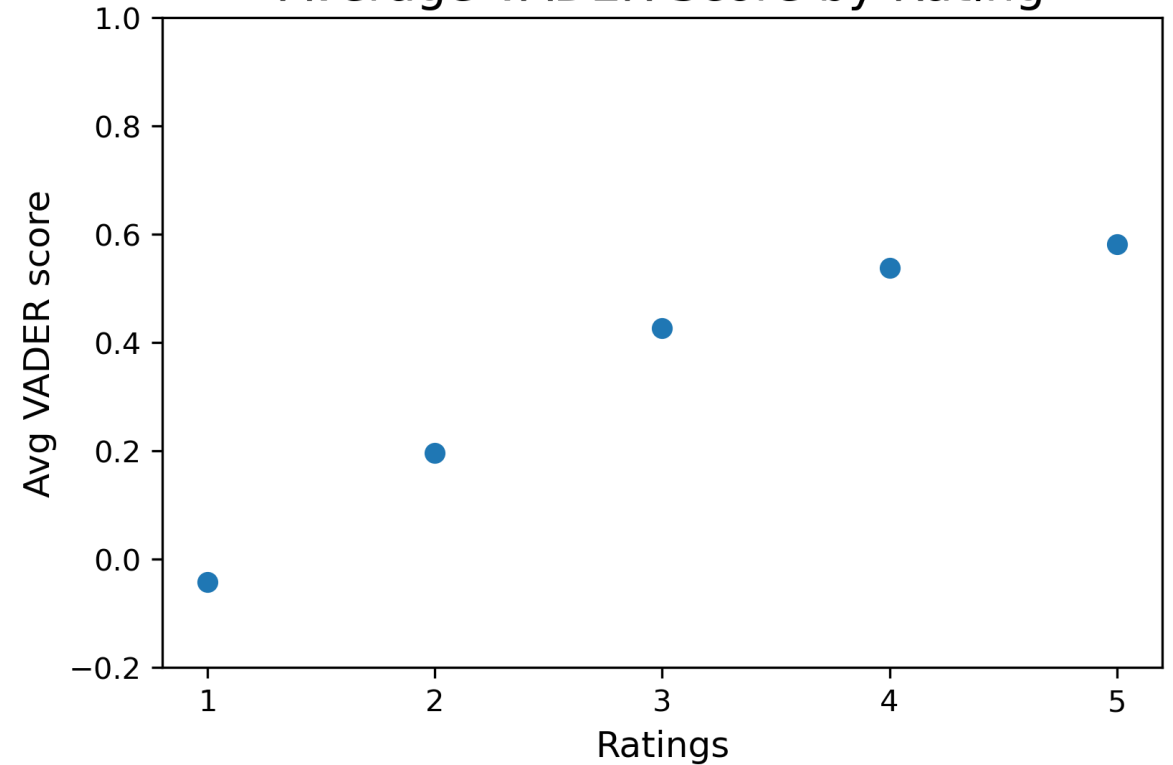
☹️ Review Sentiment: VADER

Rating distribution



- Many high ratings
- Customers are relatively happy with the products

Average VADER Score by Rating



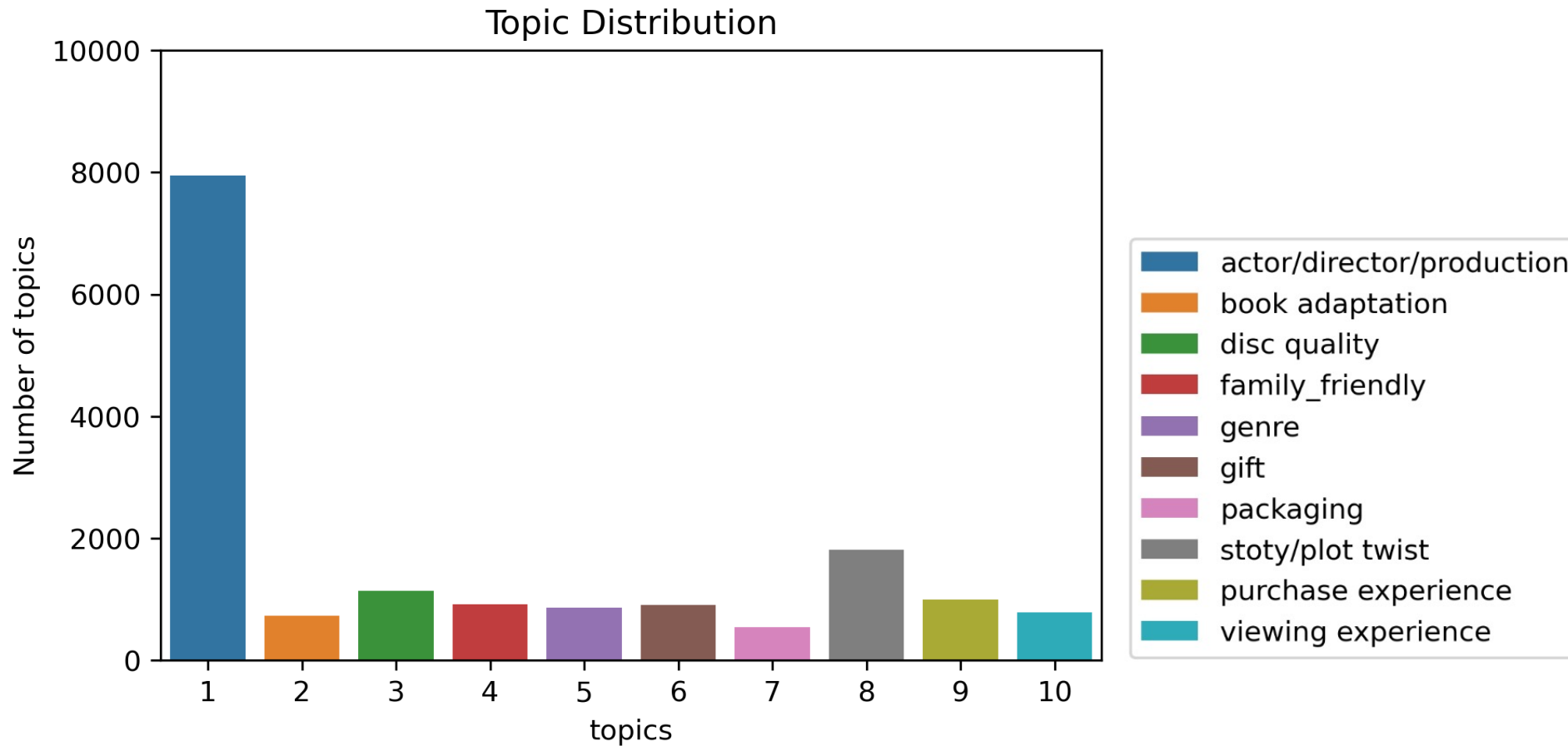
- Higher ratings, higher VADER score
- VADER seems to correctly detect the sentiments

Before Topic Modeling:
Continue refining words; Filter words down to only NOUNs



- Filter out emotions
- Zoom in movie elements
 - Story, character, series, plot, performance, actor, footage...

Topic Modeling: NMF



Actor/director performamce and movie production are mostly mentioned;
Followed by plot twist

Topic Modeling: go back to review text

Original Review Text	NLP Text	NMF Topic
Killdozer is an enjoyable scarey movies not gorey but the movie keeps you involvedThe movie has a good storey line great actorsI would recommend Killdozer to anyone who enjoys old classic movies	scarey line recommend	story/plot twist
When I was a child and first saw Peter Pan in the theaterI stayed to see it three times in a row	child theaterI row	actor/director/production
Our kids loved this and I wanted it on DVD for grandchildren Color sound everything is lovely A family favorite	sound everything family favorite	family_friendly

Future Work

- Continue refining topic modeling and exploring more NLP methods and tools such as Spacy, ScatterText, etc.
- Build Content and Collaborative Recommender System