ANALYZING PRODUCT REVIEWS

Chetana Vyas

WORKFLOW

EDA → **NLP** → **TOPIC MODELLING** → **RECOMMENDATION SYSTEMS**

EDA
- Initial EDA
- Visualization

NLP
- Text Preprocessing

TOPIC MODELLING
- NMF using TF-IDF
- LDA using TF-IDF
- Dimensionality Reduction - PCA

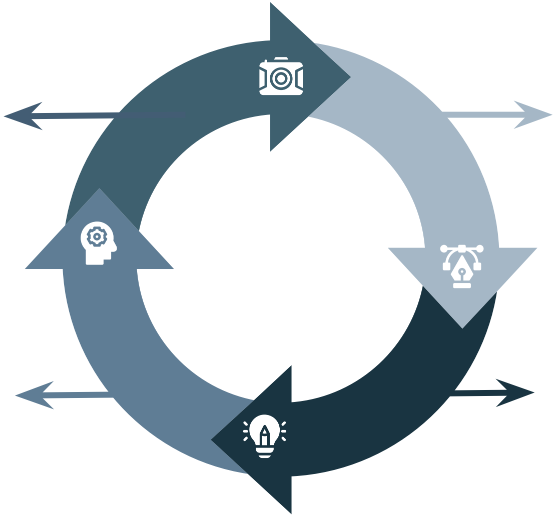
RECOMMENDATION SYSTEMS
- Find similar products

Why Product Reviews?

DATASET Datafiniti's Product Database

71,000+
reviews

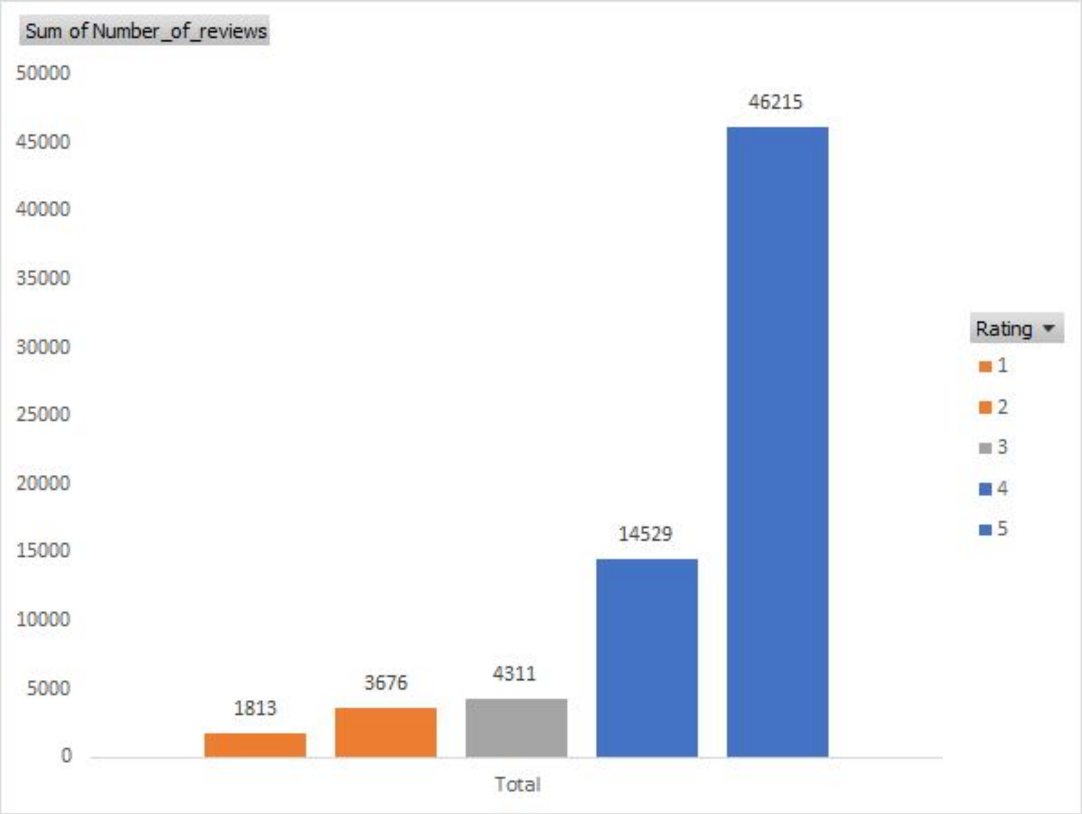
1,000
products



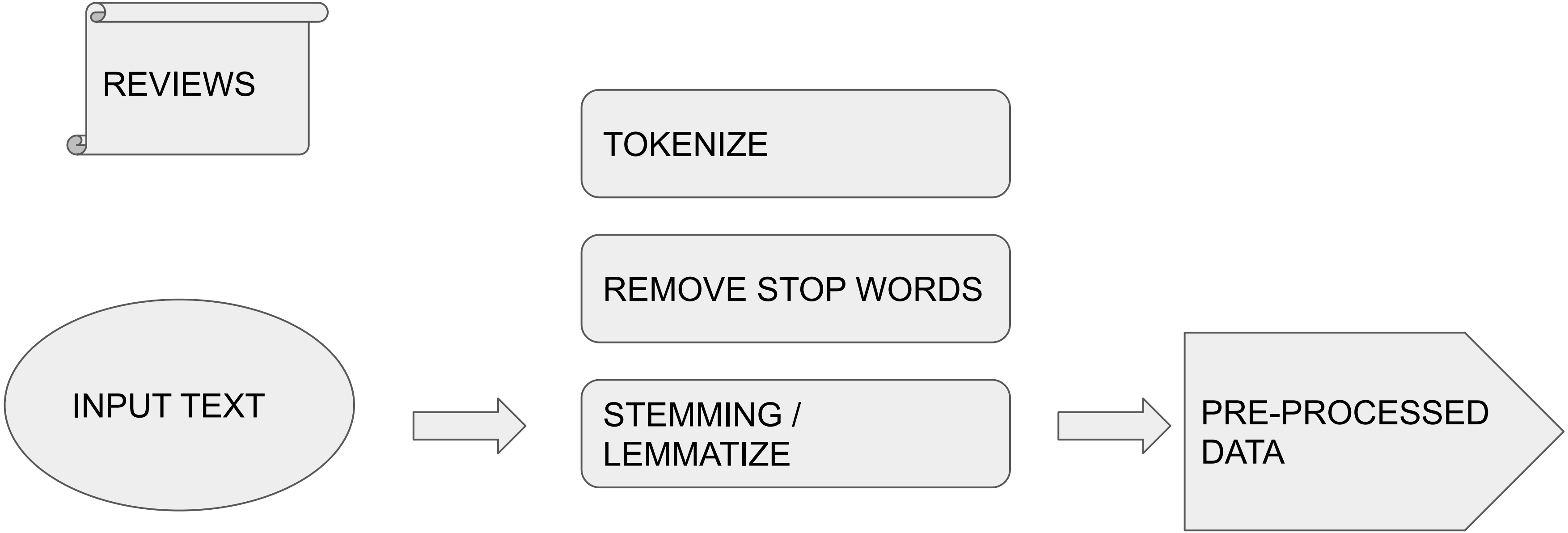
86% Positive reviews
8 % Negative reviews

555 reviews Clorox Disinfecting Wipes (most popular)

PRODUCT REVIEWS DISTRIBUTION



NATURAL LANGUAGE PROCESSING



TOPIC MODELING

TOPIC 1: Movie Review

MOVIE, ENJOYED, WATCH, LOVED, AWESOME, GODZILLA, CUTE, REALLY, WATCHING, ACTION, EXCELLENT, BEST, GREAT, WATCHED, RECOMMEND

TOPIC 2:Cleaning Product Review

WIPES, CLOROX, COLLECTED, PROMOTION, USE, CLEAN, EASY, DISINFECTING, CONVENIENT, KITCHEN, HOUSE, BATHROOM, QUICK

TOPIC 3: Skin Product Review (Moisturizer)

SKIN, PRODUCT, MOISTURIZER, OLAY, FACE, COLLECTED, PROMOTION, USE, FEEL, TOTAL, EFFECTS, LIKE, FEELS, SMOOTH, AGING

TOPIC 4: Laundry Detergent Review

TIDE, PODS, CLOTHES, CLEAN, LAUNDRY, SMELL, DETERGENT, USE, FRESH, EASY, SCENT, POD, PRODUCT

RECOMMENDATION SYSTEMS

CONTENT BASED FILTERING

Input -"soft skin cream"

'Olay Total Effects Daily Moisturizer
7-In-1 Anti-Aging, 0.5oz'
'Olay Regenerist Deep Hydration Regenerating Cream'
"L'oreal Paris Revitalift Triple Power Deep-Acting Moisturizer"
'CeraVe SA Renewing Cream'
'Shea Moisture African Black Soap Problem Skin Facial Mask'
'Tree Hut Shea Body Butters, Coconut Lime, 7 oz'
'Nivea Extended Moisture Body Lotion

Input -"laundry"

'Clorox Disinfecting Wipes Value Pack Scented 150 Ct Total'
'Tide Original Liquid Laundry Detergent - 100 Oz'
'Lysol Disinfectant Spray
‘Early Morning Breeze, 19.0 Ounces'
'Oxiclean Laundry Stain Remover Spray'
'Tide Pods Detergent, Botanical Rain, 61ct'
'Storkcraft Tuscany Glider and Ottoman
‘Beige Cushions, Espresso Finish'

FUTURE WORK

- Gather reviews from users who have actually used the product
- Identify Fake Reviews
- Try various Clustering algorithms

APPENDIX - Topic Interpretations

Topic Interpretations

TOPIC_0: Thriller Movie Review

MOVIE, ENJOYED, WATCH, LOVED, AWESOME, GODZILLA, CUTE, REALLY, WATCHING, ACTION, EXCELLENT, BEST, GREAT, WATCHED, RECOMMEND

TOPIC_1:Cleaning Product Review

WIPES, CLOROX, COLLECTED, PROMOTION, USE, CLEAN, EASY, CLEANING, DISINFECTING, CONVENIENT, KITCHEN, HOUSE, BATHROOM, QUICK, PRODUCT

TOPIC_2: Skin Product Review (Moisturizer)

SKIN, PRODUCT, MOISTURIZER, OLAY, FACE, COLLECTED, PROMOTION, USING, FEEL, TOTAL, EFFECTS, LIKE, FEELS, SMOOTH, AGING

TOPIC_3: Movie Review

GOOD, ORIGINAL, PRICE, PRETTY, SEQUEL, ACTION, LIKE, STORY, QUALITY, REALLY, BETTER, LIKED, NICE, BUY, JUST

TOPIC_4: Household Cleaning Supplies Review

GREAT, PRODUCT, PRICE, WORKS, BUY, SMELLS, RECOMMEND, MOVIES, QUALITY, USE, VALUE, CLEANING, JOB, STORY, DEAL

TOPIC_5: Laundry Detergent Review

TIDE, PODS, CLOTHES, CLEAN, LAUNDRY, SMELL, DETERGENT, USE, FRESH, EASY, SCENT, POD, JUST, PRODUCT, USED

TOPIC_6: Hair Product Review

HAIR, CONDITIONER, SHAMPOO, SOFT, RECEIVED, PRODUCT, FREE, COLLECTED, PROMOTION, INFLUENSTER, TESTING, OILY, PURPOSES, OPINIONS, MASK

TOPIC_7: Food / Personal Care Product Review

LOVE, SMELL, PRODUCT, AWESOME, ABSOLUTELY, HOUSE, AMAZING, KIDS, JUST, MAKES, PRODUCTS, SMELLS, LIP, COLOR, MOP

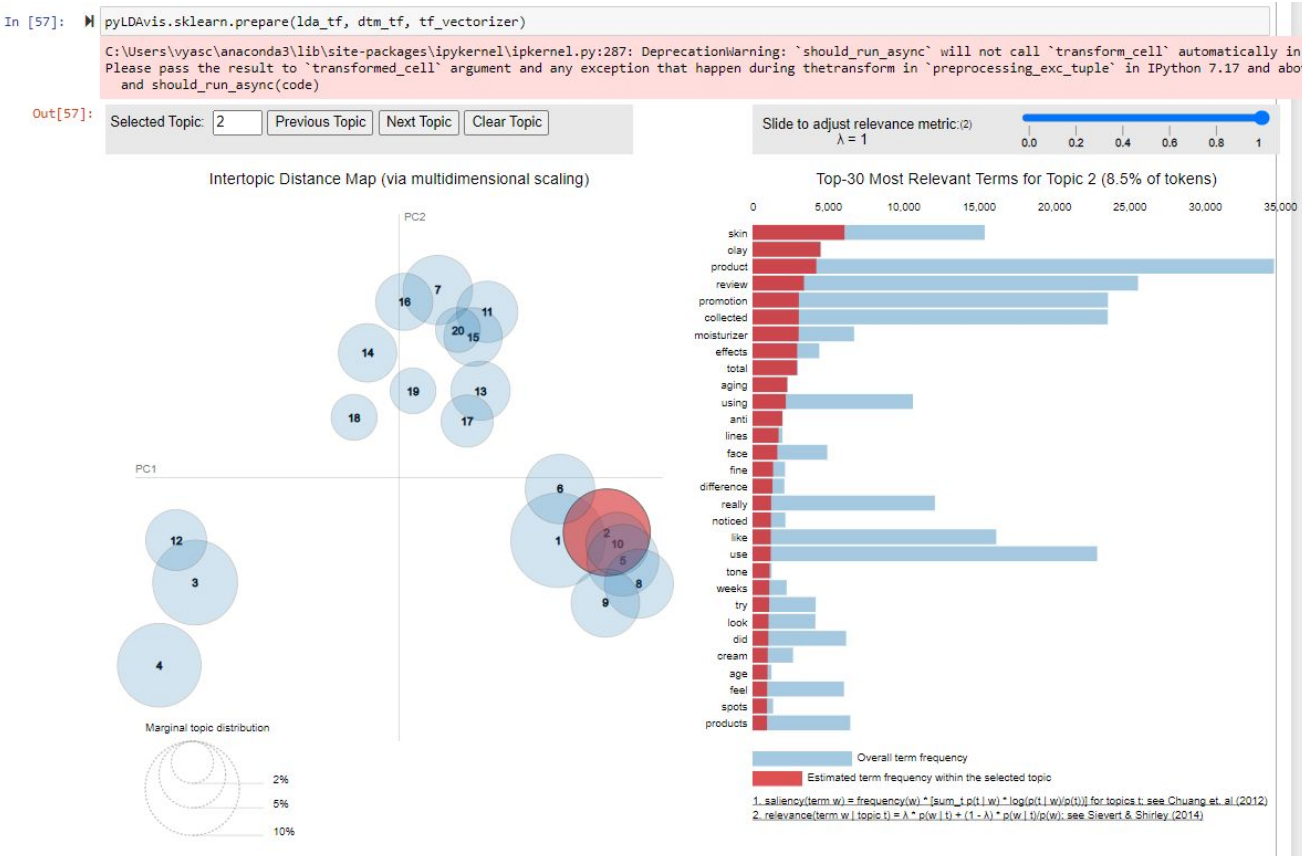
TOPIC_8: Comedy Movie Review

KIDS, FUNNY, ADULTS, LOVED, CUTE, MOVIE, WATCH, ENTERTAINING, HILARIOUS, ADULT, FUN, PETS, HUMOR, REALLY, ENJOY

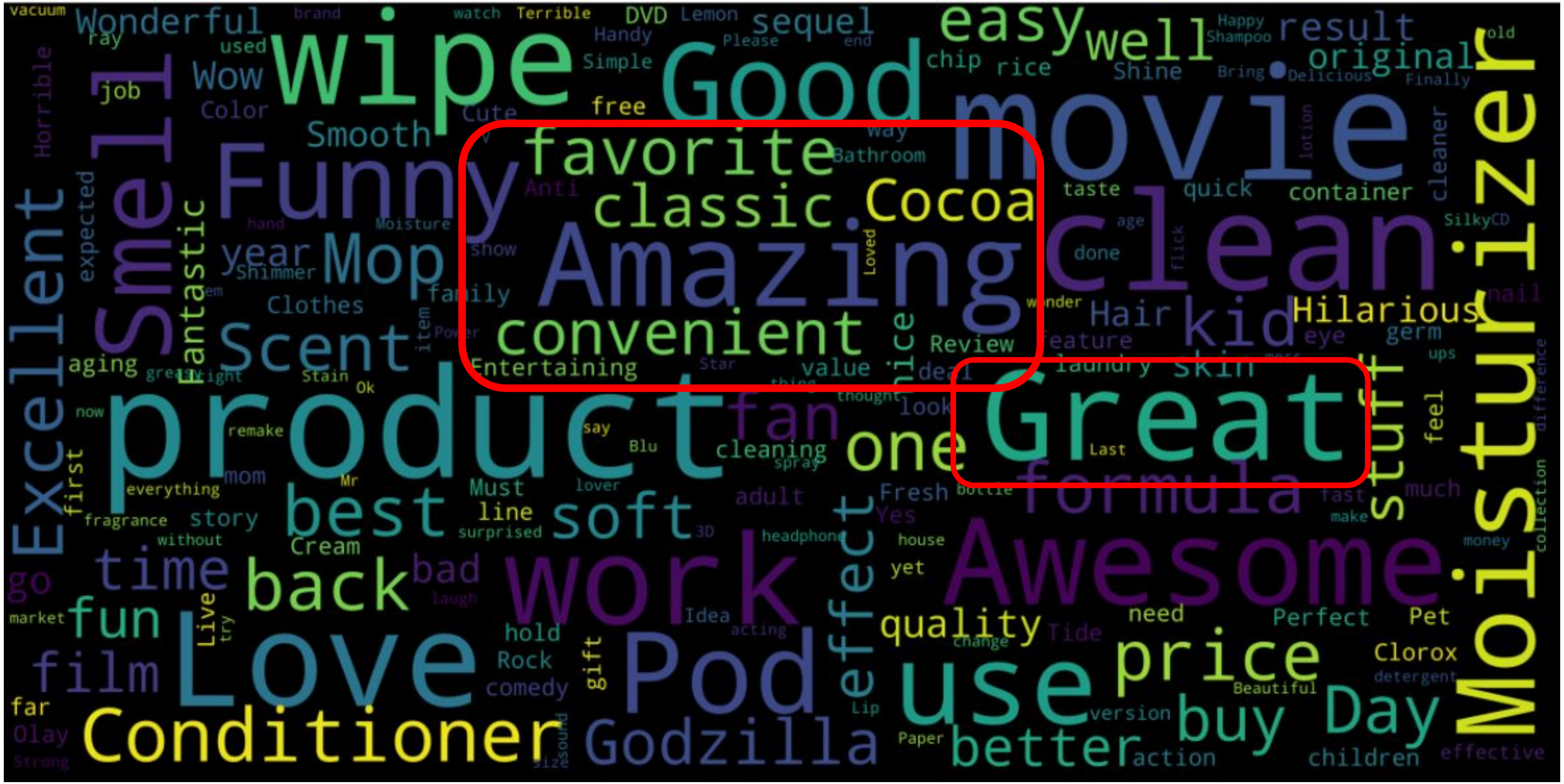
TOPIC_9: Kids/Family Movie Review

FAMILY, FUN, WATCH, ENJOYED, NIGHT, ENTIRE, MOVIE, ENJOY, FILM, FRIENDS, CUTE, AGES, ENTERTAINING, PETS, FRIENDLY

APPENDIX - visualization using pyLDAVis



APPENDIX - Word Cloud



APPENDIX -

Document - Term Matrix

Document / Topic Matrix

In [73]:

```
doc_topic_matrix_df  = pd.DataFrame(doc_topic_matrix).add_prefix('topic_')
doc_topic_matrix_df[['reviewsConcat','reviews_keywords']] = product_df[['reviewsConcat','reviews_keywords']]
doc_topic_matrix_df.head(10)
```

Out[73]:

	topic_0	topic_1	topic_2	topic_3	topic_4	topic_5	topic_6	topic_7	topic_8	topic_9	reviewsConcat	reviews_keywords
0	0.001586	0.000513	0.002394	0.021652	0.000316	0.006999	0.000000	0.021560	0.000000	0.000000	Just Awesome i love this album. it's very good...	['just awesome i love this album it s very go...
1	0.000000	0.025483	0.006359	0.114957	0.000000	0.000000	0.008534	0.000000	0.000000	0.000000	Good Good flavor. This was collected as part ...	['good good flavor this was collected as par...
2	0.000000	0.000000	0.000000	0.124778	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	Good Good flavor.	['good good flavor ']
3	0.000710	0.005852	0.006265	0.000550	0.000000	0.015386	0.000228	0.000000	0.000000	0.000000	Disappointed I read through the s on here befo...	['disappointed i read through the s on here be...
4	0.000995	0.000000	0.021566	0.000604	0.000000	0.002488	0.001356	0.000000	0.000150	0.000000	Irritation My husband bought this gel for us. ...	['irritation my husband bought this gel for us...
5	0.002911	0.000000	0.009563	0.001856	0.003526	0.007866	0.003688	0.023571	0.000071	0.000000	Not worth it My boyfriend and I bought this to...	['not worth it my boyfriend and i bought this ...
6	0.001914	0.000000	0.013072	0.002739	0.001625	0.013974	0.005396	0.001658	0.000638	0.000000	Disappointing Bought this earlier today and wa...	['disappointing bought this earlier today and ...
7	0.001705	0.000000	0.015346	0.000000	0.001866	0.011936	0.011050	0.001258	0.003975	0.000000	Not happy at all I bought this product for my ...	['not happy at all i bought this product for m...
8	0.000342	0.000000	0.011775	0.002326	0.006826	0.010428	0.004096	0.001242	0.001502	0.014411	Very disappointing My husband and I bought thi...	['very disappointing my husband and i bought t...
9	0.001820	0.000000	0.005896	0.003862	0.001375	0.015794	0.001282	0.002238	0.000531	0.000000	Don't buy Got as a surprise for my husband the...	['don t buy got as a surprise for my husband t...