Product Recommender System

Yunus Emre Işıkdemir

Yunus Emre Işıkdemir

Problem Definition

Positive Recommenced Index

Negative Recommended Index

Target Steps

01 Numerical analysis with continuous variables and categorical variables.

02 Polarity Scores of reviews via automated NLTK Sentiment Analyzer.

03 Sentiment analysis on reviews to predict recommendation index from scratch.

Yunus Emre Işıkdemir

Dataset Description

	Clothing ID	Age	Title	Review Text	Rating	Recommended IND	Positive Feedback Count	Division Name	Department Name	Class Name
0	767	33	NaN	Absolutely wonderful - silky and sexy and comf...	4	1	0	Initmates	Intimate	Intimates
1	1080	34	NaN	Love this dress! it's sooo pretty. i happene...	5	1	4	General	Dresses	Dresses
2	1077	60	Some major design flaws	I had such high hopes for this dress and reall...	3	0	0	General	Dresses	Dresses
3	1049	50	My favorite buy!	I love, love, love this jumpsuit. it's fun, fl...	5	1	0	General Petite	Bottoms	Pants
4	847	47	Flattering shirt	This shirt is very flattering to all due to th...	5	1	6	General	Tops	Blouses

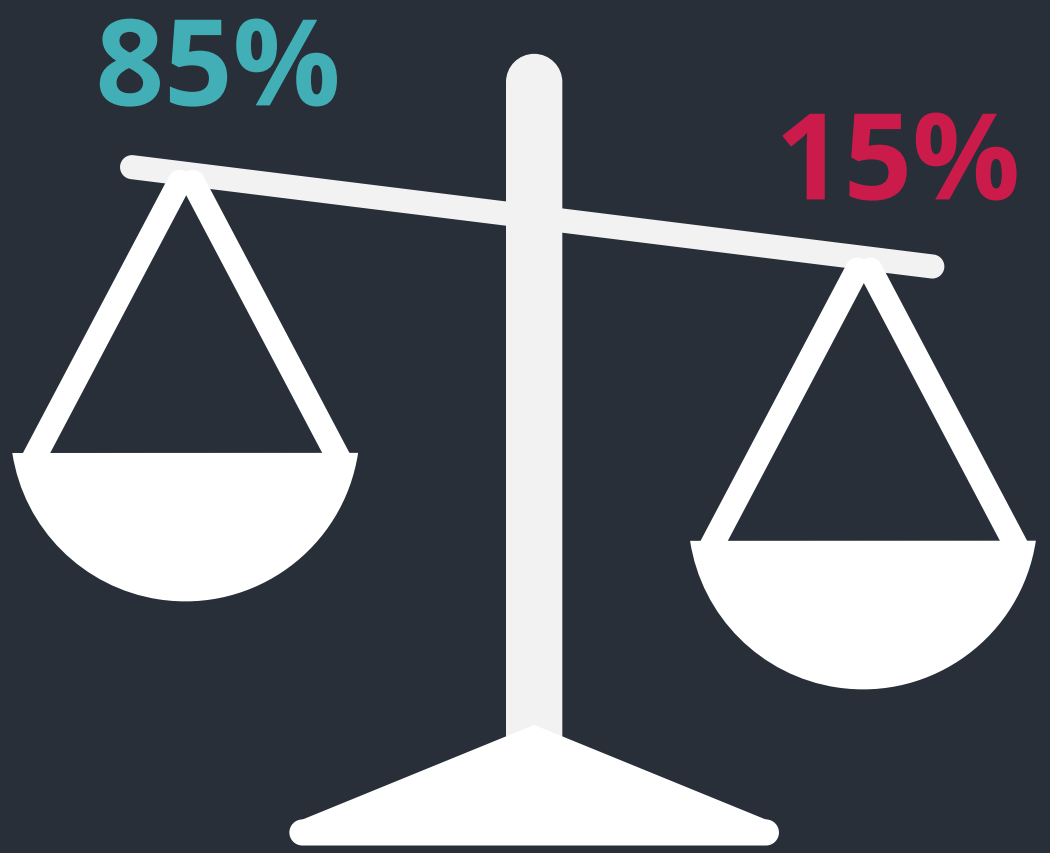
- Collection of customer reviews from an e-commerce website that has been anonymized privacy policy.
- 23486 customers and related information.
- 10 variables that describes the customer behaviours and products.

Yunus Emre Işıkdemir

Imbalanced Distribution

Recommended

Customers who are satisfied with the product.

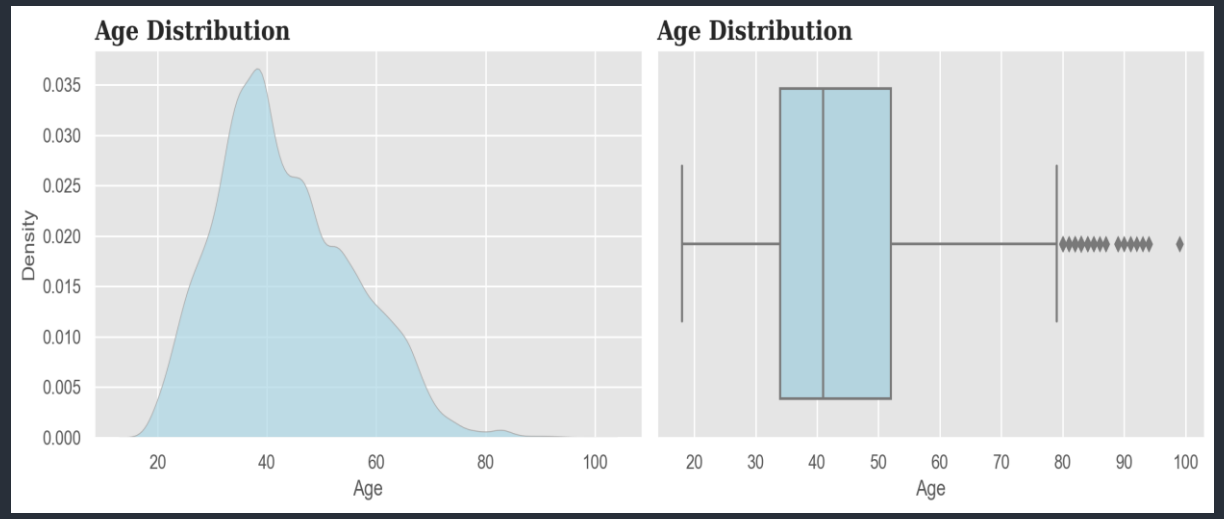


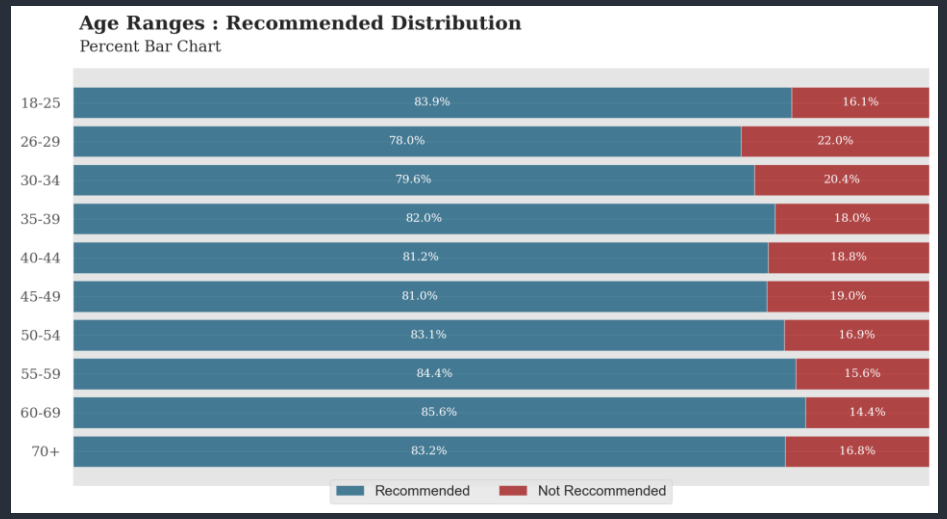
Nonrecommened

Customers who are not satisfied with the product.

EDA – Age Distribution

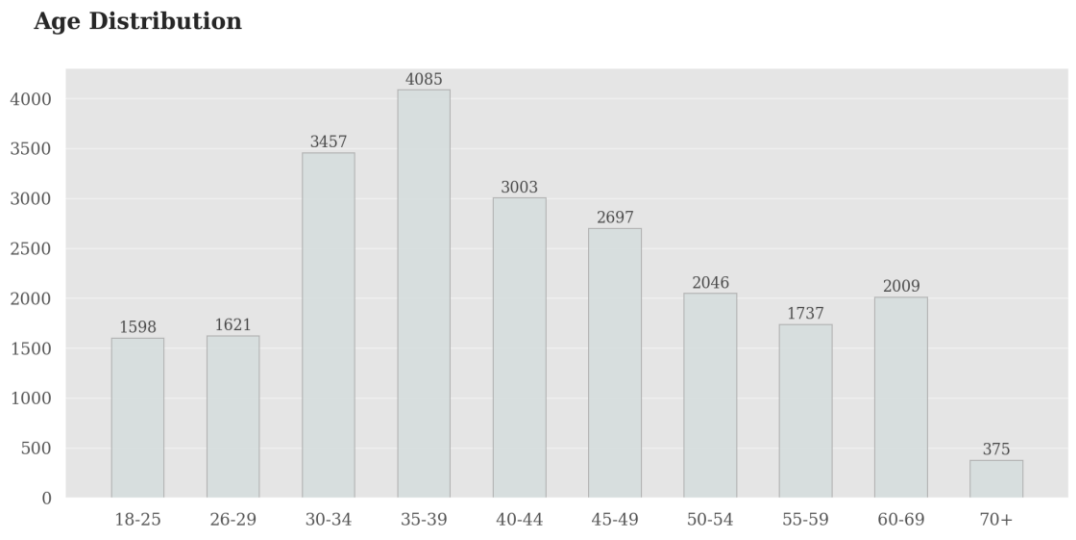
- Most customers are in the 37-43 age range. Young customers are in the minority.
- There are some rare observations which are 99 years old. This does not fit into general distribution.
- With interquantile range method, it is possible to say that +78 years old are rare.

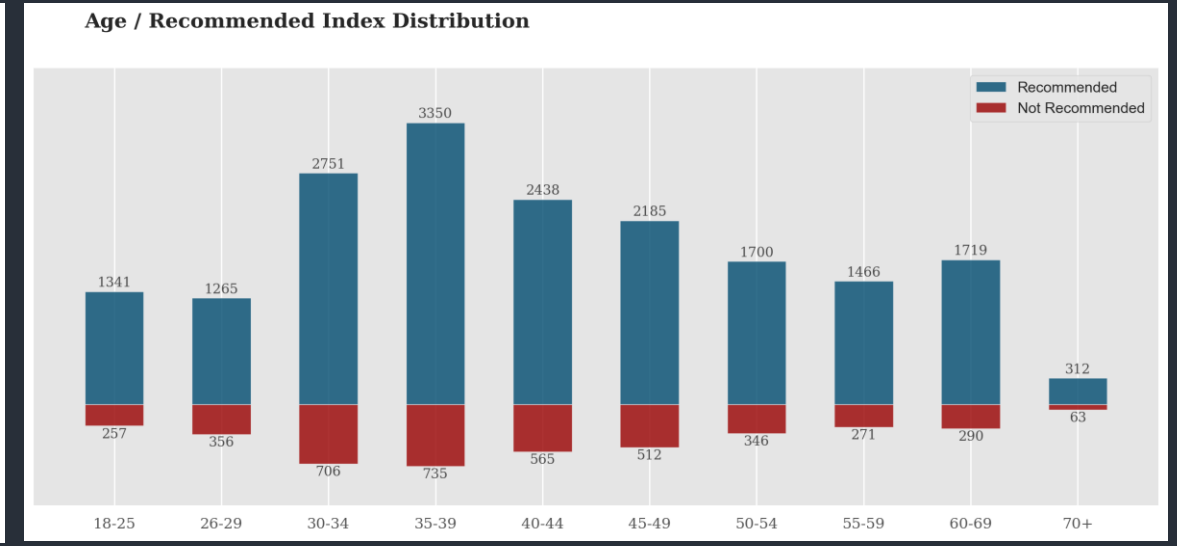




Categorize the Age Variable

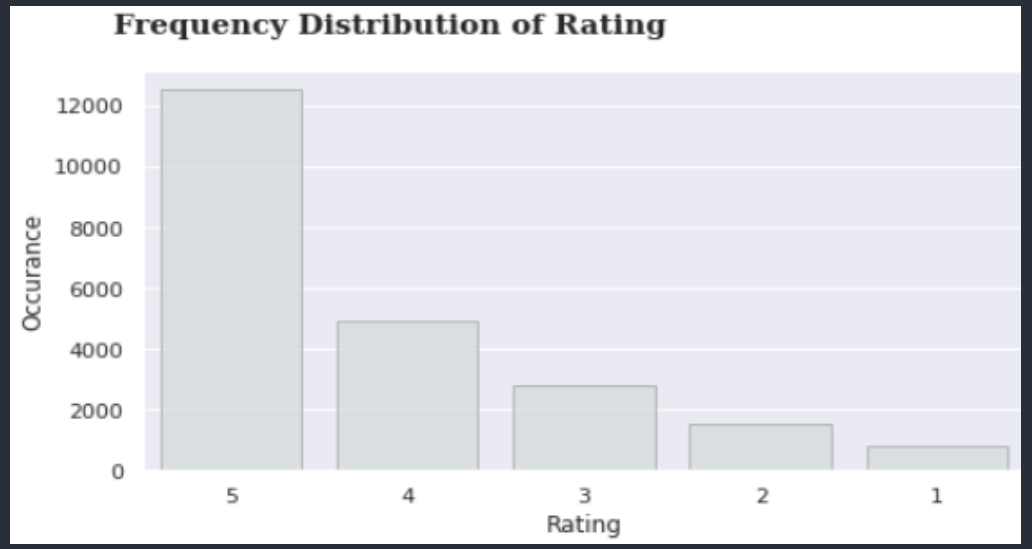
- The age variable was categorized into a specific range and crossed with the recommendation target variable. The rates of recommendation and non-recommendation according to age are clearly seen.

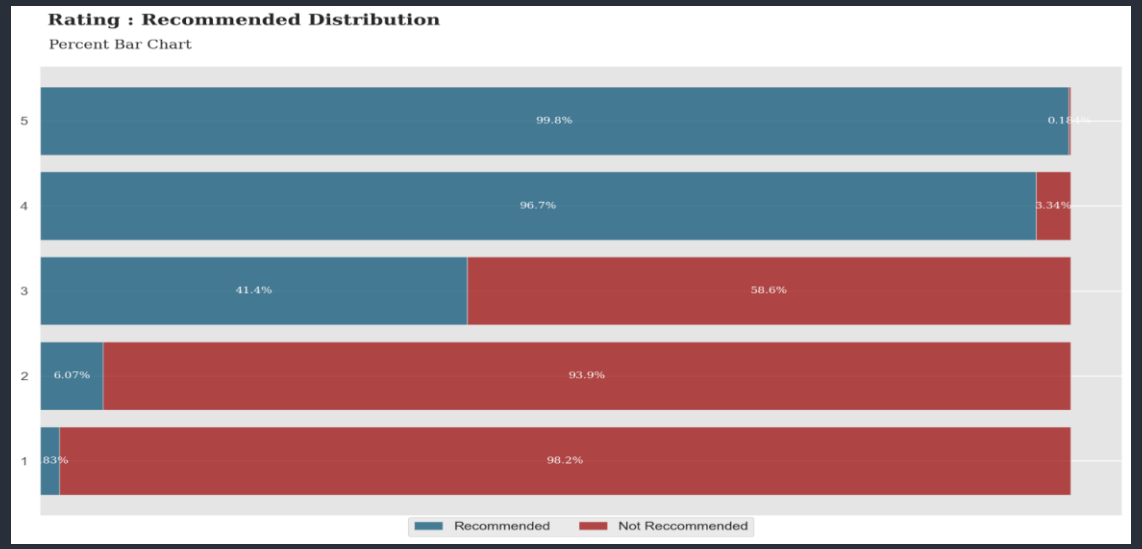




Rating vs Recommendation

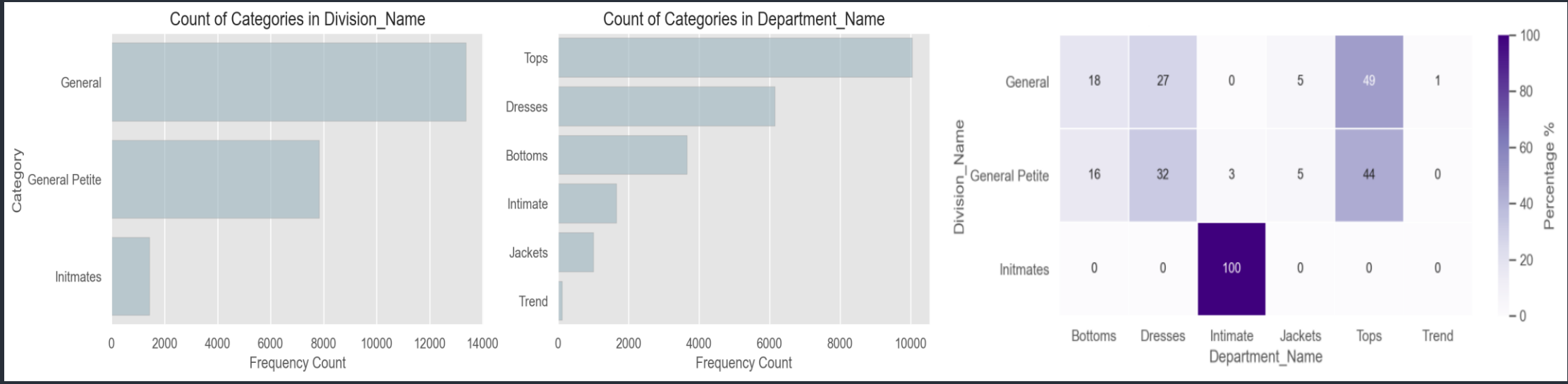
- When the rates of recommendation are examined according to the given, those who voted higher tend to recommend the product.





Division Name and Department Name

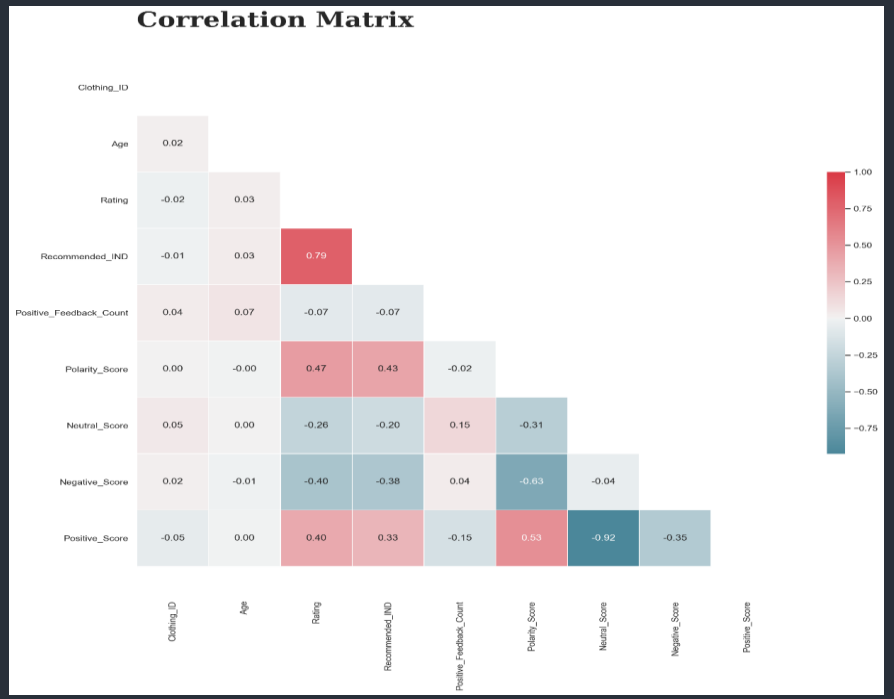
- The top sellest products are shown. In this case, tops and dresses are more demanded products.
- According to the general division, most of the products belong to the top department in terms of percentage.

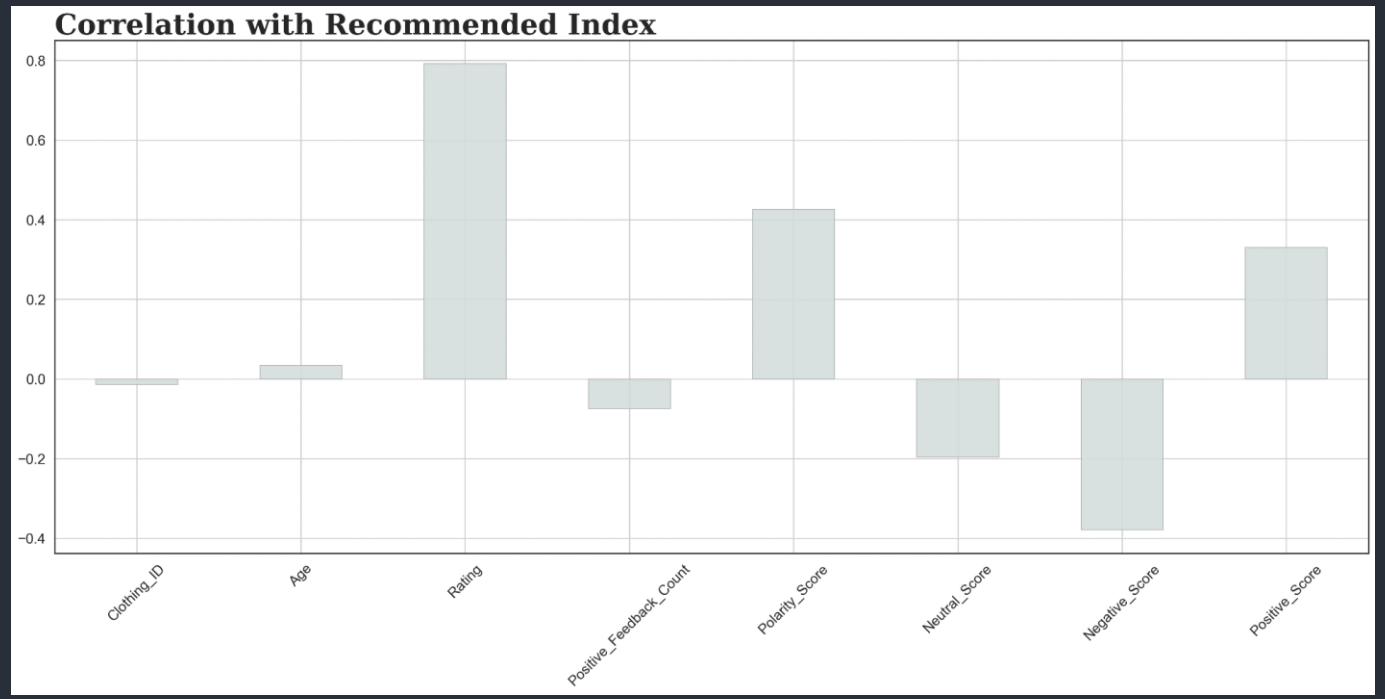


Yunus Emre Işıkdemir

Correlation Analysis

- Rating and recommendation index are highly correlated in positive direction.





Literature Reviews

TFIDF Vectorization

Beyond TFIDF Weighting for Text Categorization in the Vector Space Model. (P. Soucy and G.W. Mineau)

VADER Sentiment Analysis

Vader: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. (C.J. Hutto and Eric Gilbert)

Analysis of Reviews

- Automated extracting polarity scores from reviews via NLTK Sentiment Analyzer.

Review_Text	Polarity_Score	Neutral_Score	Negative_Score	Positive_Score	SentimentVader
I love this shirt. i have it in red. i found t...	0.8932	0.272	0.000	0.728	Positive
After reading the reviews, i had high hopes fo...	0.9729	0.664	0.000	0.336	Positive
This has a heftier fabric than seems to be in ...	0.9427	0.792	0.027	0.181	Positive
Let me start with: i love this dress --- but i...	0.5727	0.340	0.226	0.434	Positive
I am usually a regular xs with retailer tops. ...	0.9291	0.700	0.000	0.300	Positive
Loved these in the store but the 00 was sold o...					
I know it's a bit of a moot point since the ch...					

Focus on Non – Recommended Reviews

- Focusing on the persons who have not recommended the product may be provided added value to the company.

Common Words

- Common words of recommended and nonrecommended products' review.

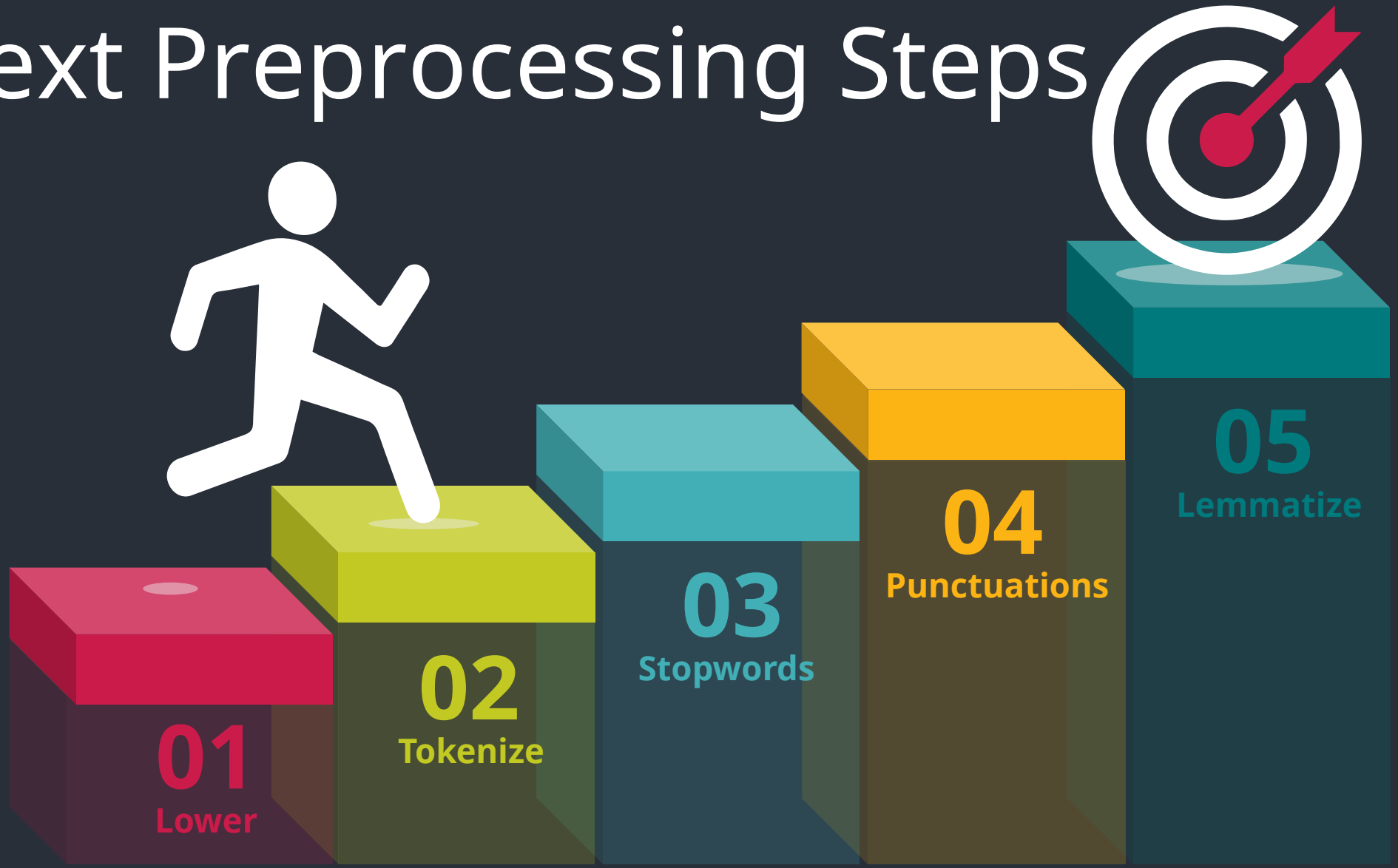
Non Recommended Items N-Gram

- Important to note that ‘love’ is used with combination of ‘really wanted love’.

	1-Gram	Occurrence	2-Gram	Occurrence	3-Gram	Occurrence	4-Gram	Occurrence	5-Gram	Occurrence
0	dress	1976	wanted love	243	really wanted love	70	really wanted love dress	15	reference 5 7 125 lb	3
1	like	1780	going back	215	wanted love dress	65	looked like maternity top	10	going back wanted love dress	3
2	top	1572	looked like	187	really wanted like	40	really wanted like dress	9	reference measurements 38 30 40	3
3	would	1348	looks like	153	made look like	29	really wanted like top	9	photos reference measurements 38 30	3
4	fit	1327	really wanted	151	wanted love top	28	5 4 120 lbs	8	medium photos reference measurements 38	3

Yunus Emre Işıkdemir

Text Preprocessing Steps



Count Vectorization Results

- Logistic regression with considering the class distribution, it is obtained that more balanced results.

Methods	Sensitivity (%)	Specifity (%)
Logistic Regression	95	57
LR with adjusted weights	88	77
Naive Bayes	93	60
Support Vector Machines	96	54

TFIDF Vectorization Results

- Logistic regression with considering the class distribution, it is obtained that more balanced results.

Methods	Sensitivity (%)	Specifity (%)
Logistic Regression	97	48
Logistic Regression with adjusted weights	86	83
Naive Bayes	100	02
Support Vector Machines	95	57

Deeper Analysis on Words

- Words which are filledwith greens are mostly occured in recommended comments.
- Reds are mostly occured in non-recommended comments.

+5.655	love
+5.176	perfect
+4.627	great
+4.527	little
+4.166	comfortable
+3.789	with
+3.490	soft
+3.319	compliments
+3.259	fits
+2.855	bit
+2.826	unique
+2.822	perfectly
+2.567	comfy
+2.563	happy
+2.522	jeans
+2.438	size
+2.391	glad
+2.365	slightly
+2.331	feminine
+2.311	nicely
+2.212	amazing
+2.135	easy
+2.133	fun

-2.480	would
-2.500	return
-2.530	odd
-2.632	even
-2.674	maternity
-2.690	way
-2.775	excited
-2.987	poor
-3.125	bad
-3.196	back
-3.251	looked
-3.343	unflattering
-3.536	huge
-3.575	not
-4.048	returned
-4.059	was
-4.235	cheap
-4.311	returning
-4.559	wanted
-5.106	disappointed

i am not really love this because shapeless

if you have any cleavage this dress will look awful. the ties drop pretty low & look trashy over cleavage. recommend otherwise - good fabric cool look