

# **Review-based Popularity Analysis of the Online Fashion Industry**

**PRESENTED BY:**

**ANTONY PAULSON CHAZOOR | KOMAL JAISWAL | MUHTASIN RAHMAN | SHWETA PATIL**

# INTRODUCTION

- FAST FASHION INDUSTRY IS A PROFITABLE ARENA
- DRIVEN BY WORD OF MOUTH, PERCEIVED FASHION VALUE DRIVES SALES
- REVIEWS EXTREMELY IMPACTFUL, ESPECIALLY ONLINE WHERE THERE ARE NO TRIALS
- FUTURE-CENTRIC THINKING NEEDED TO FOLLOW CUSTOMER PREFERENCE AND FASHION TRENDS
- POPULARITY AMONG CONSUMERS PARAMOUNT FOR CONTINUED BUSINESS SUCCESS
- GOAL: PREDICT POPULARITY BASED ON KEY ADJECTIVES IN REVIEWS

What are the most frequent words used by reviewers?

Do popular products have characteristic keywords?

What are the most popular items?

## RESEARCH QUESTIONS

Possible to predict popularity based on review text?

## RESEARCH IMPLICATIONS

Popular products dictate future stock and promotions

Automated system to detect popular products

Unpopular products changed or advertised differently

Detect trends to adjust marketing strategy and research

Curb customer criticism and engender customer engagement

Decrease return rate and avoid restocking fees

# DATA DESCRIPTION

Kaggle.com, a data repository website

23486 rows, 10 feature variables for women's clothing

Main attributes:

-Clothing ID

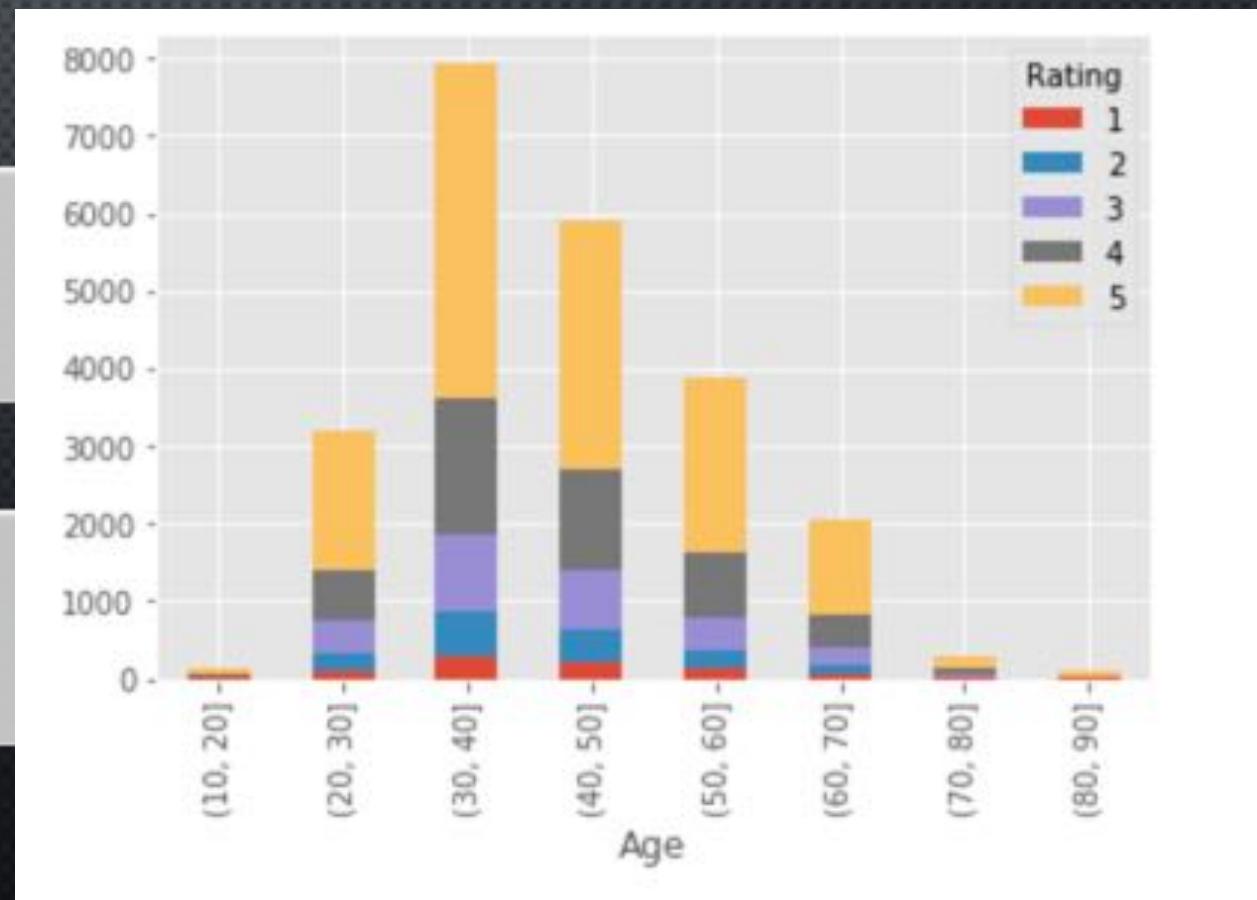
-Review Text

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

# EXPLORATORY DATA ANALYSIS

Age group-Ratings  
Breakdown

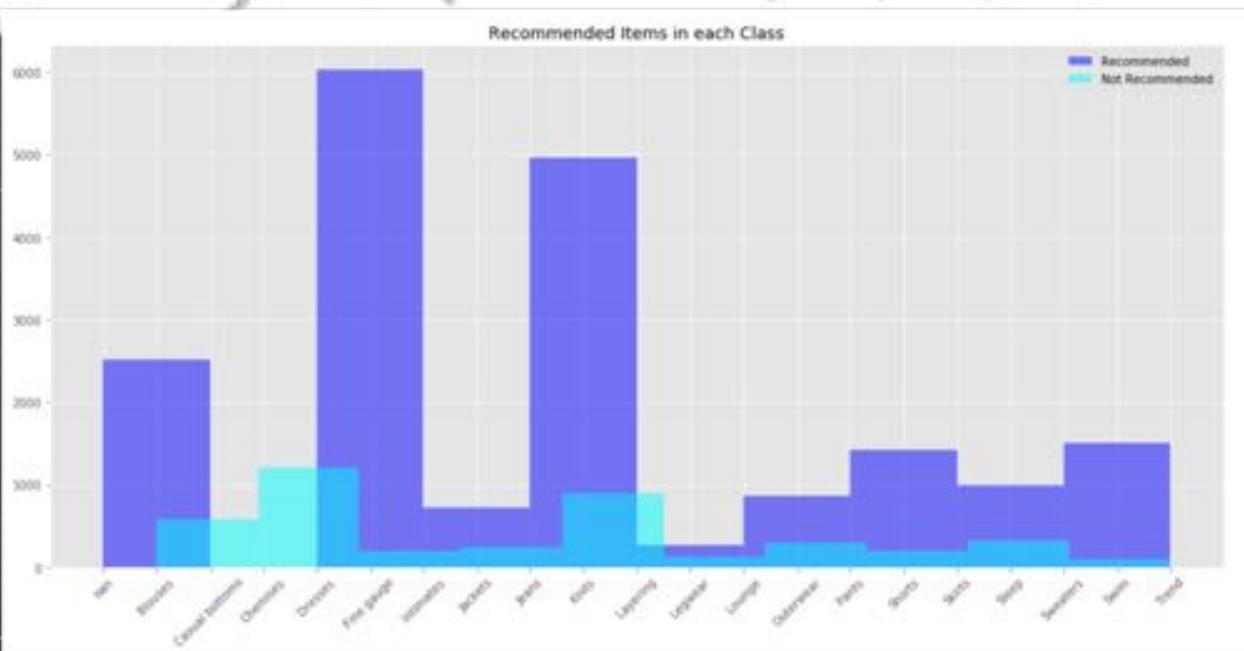
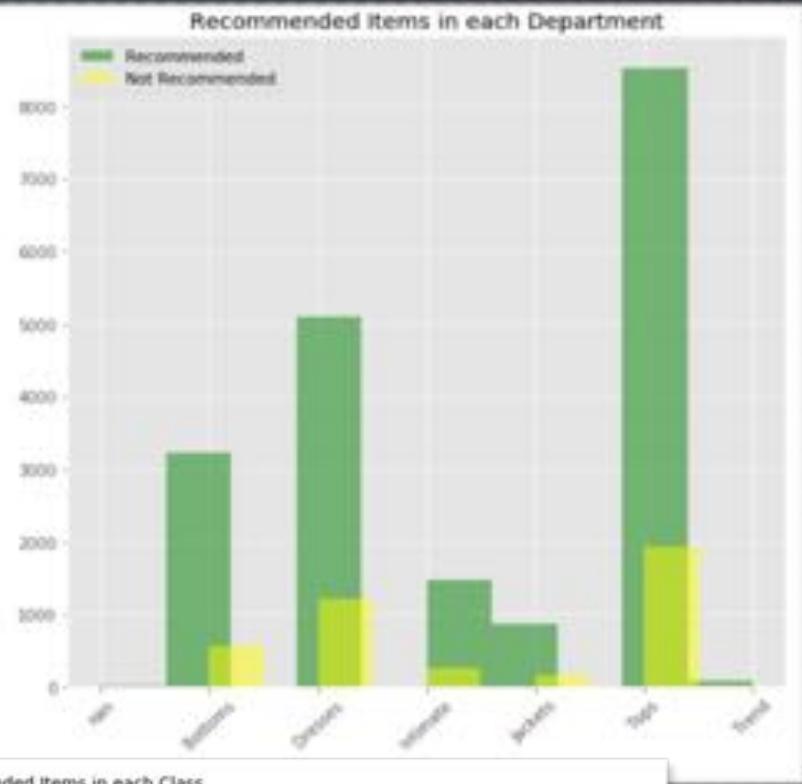
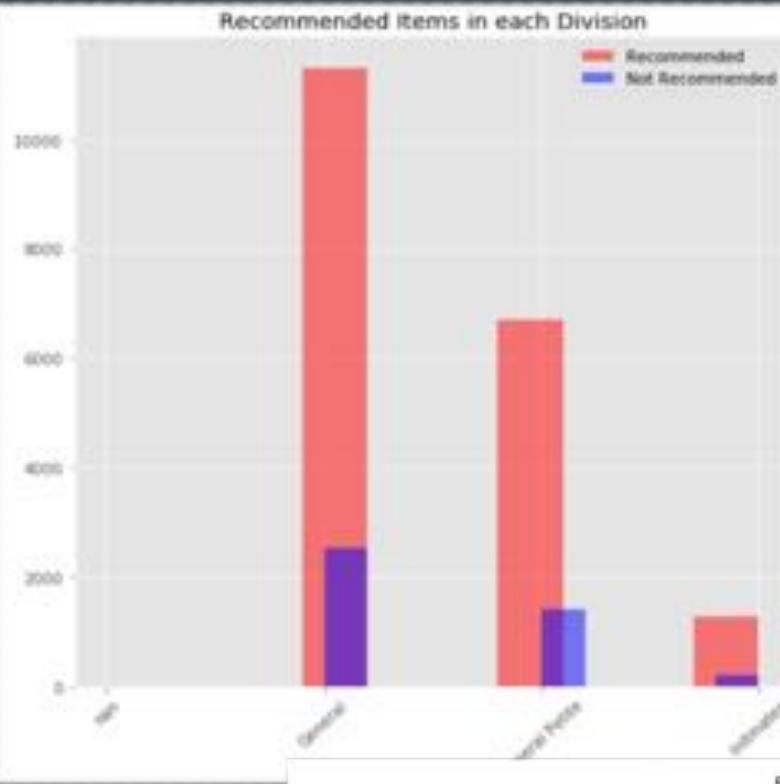
Age group 30-40  
gave best ratings  
and review the most



The most recommended items are from General Division

The most recommended items are from the Tops, Dresses and Bottoms department

The most recommended items are from Dresses, Fine Gauge and Intimates Class



Reviews have approx.  
200 or more words in  
length

Reviews of ratings 3 and 4 have greater length

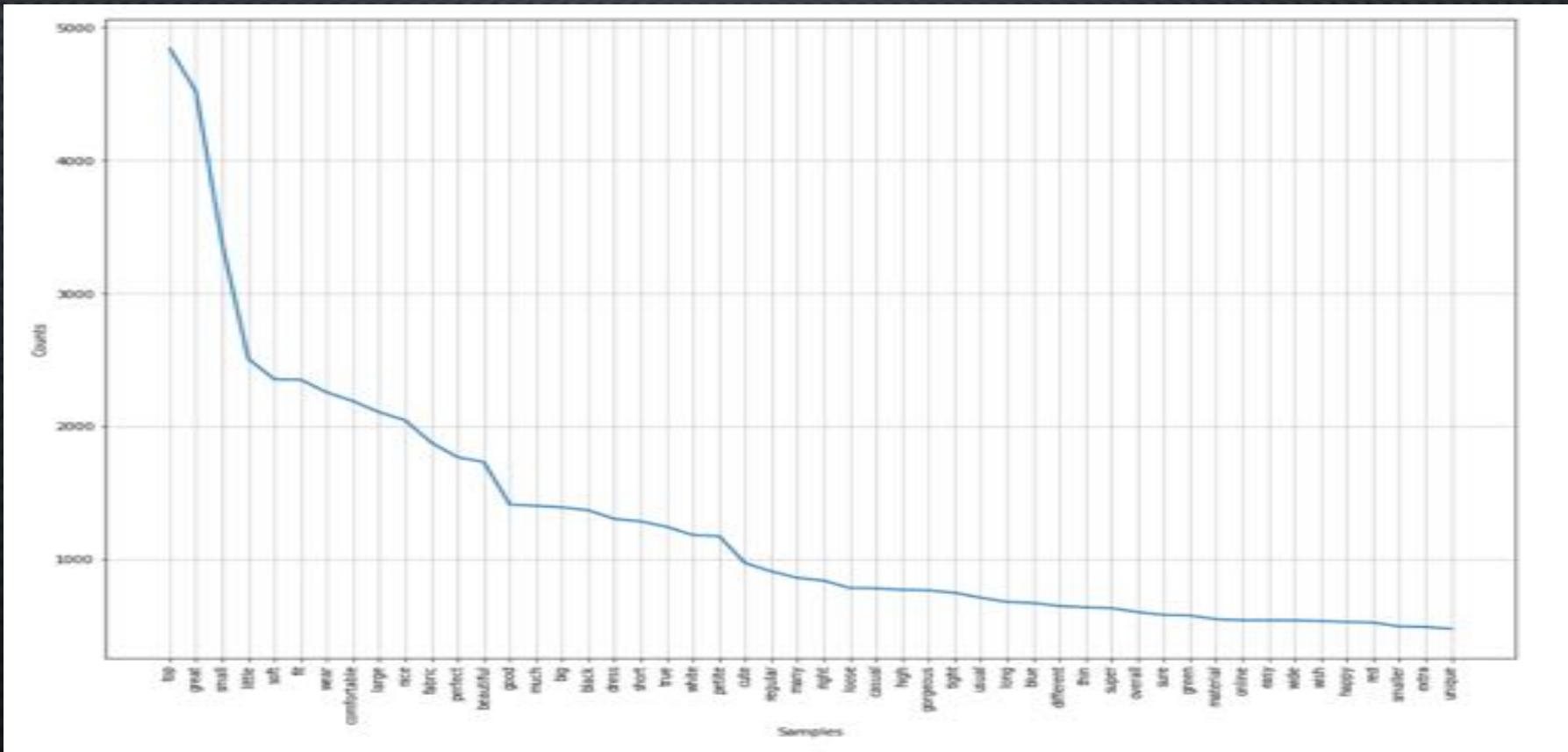
482,899 words in  
reviews corpus



Top 50 adjectives  
picked with at least  
500 repetitions

'JJ', 'JJR' and  
'JJS' part of  
speech  
tagging

Top  
adjectives -  
Top, Great,  
Small, Little



Select popular apparel:  
Highest number of reviews

Clothing ID	Count
1078	1024
862	806
1094	756
1081	582
872	545
829	527
1110	480
868	430
895	404
936	358

# METHODOLOGY

Part of Speech tagging, removing stop-words, punctuations to extract adjectives

('small', 3377), ('little', 2508), ('soft', 2354)

Dataset partitioned to validation and training to approx. (70:30)

Collate all reviews of these popular items into one

50 most frequent adjectives used for prediction

Clothing ID	Review Text	dress	size	top	love	fit	like	wear	great
867	This sweater is the bomb. It's soft, flatterin...	14.0	99.0	182.0	100.0	86.0	78.0	80.0	79.0
176	This one pulled across my shoulders and neck ...	0.0	0.0	0.0	1.0	1.0	2.0	1.0	0.0
693	I love these pants. like, obsessively love, li...	0.0	0.0	0.0	3.0	1.0	3.0	0.0	0.0
1095	....that's what i look like wearing this dress	523.0	185.0	57.0	99.0	147.0	108.0	101.0	65.0

# DATA MANIPULATION

50 new columns with individual adjective frequency

"Popular" column based on number of reviews

Columns: clothing ID, reviews, adjectives, popular

# LOGISTIC REGRESSION

Training on 824 records,  
validating on 384

Only 4  
unpopular  
items predicted  
as popular

96.3%  
accuracy

	Predicted Unpopular	Predicted Popular
Unpopular	316	4
popular	9	26

Only 9 popular  
items predicted  
as unpopular

Only 13 items  
misclassified

# DECISION TREE

Easy to interpret

Only 4  
unpopular  
items predicted  
as popular

98% accuracy

	Predicted Unpopular	Predicted Popular
Unpopular	321	4
popular	3	27

Only 3 popular  
items predicted  
as unpopular

Only 7 items  
misclassified

# CONCLUSION

- 30-40 YEAR OLDS LEAVE MOST REVIEWS, FOLLOWED BY 40-50 AND 50-60 AGE GROUPS
- TOPS ARE MOST RECOMMENDED
- AMONG THOSE DRESSES, FINE GAUGE AND INTIMATES MOST RECOMMENDED
- REPRESENTATIVE WORDS PRESENT IN REVIEWS OF POPULAR PRODUCTS
- MODELS UTILIZING ADJECTIVES SUCCESSFUL IN PREDICTING CLOTHING POPULARITY
- DECISION TREE BEST MODEL WITH 98%

# FURTHER IMPROVEMENTS

The Model learned to predict unpopular items better

Balance by collecting more data on popular items and training the model to predict popular items

Features in the original dataset can be used to make a holistic prediction model

THANK YOU!

Questions?

