

SCOPE DATA

Customised Data Analysis Services

TEXT ANALYSIS CASE STUDY

TEXT ANALYSIS ON WRITTEN RE- VIEWS FROM AN ONLINE WOM- EN'S FASHION COMPANY USING R STUDIO

CONTENTS

*	3	ABSTRACT
*	4	1.TEXT ANALYSIS
*	4	2.SENTIMENT ANALYSIS
*	5	3.HOW WRITTEN REVIEWS CAN BE ANALYSED
*	6	3.1.OUTCOMES FROM A SENTIMENT ANALYSIS
*	7	3.2.TEXT ANALYSIS REVIEW EXAMPLE
*	7	3.3.CLEANING THE DATA
*	8	4.DEVELOPING A WORD CLOUD
*	9	5.SENTIMENT ANALYSIS
*	9	5.1.OVERALL SENTIMENT SCORE OF THE ONLINE WOMEN'S FASHION COMPANY
*	11	5.2.SENTIMENT REVIEW ANALYSIS OF THE ONLINE WOMEN'S FASHION COMPANY
*	12	5.3.APPLICATION OF FINDINGS, SOME STRATEGY EXAMPLE'S
*	13	6.BIGRAM - UNDERSTANDING THE CONTEXT RELATED TO SPECIFIC WORDS
*	14	7.OVERALL EMOTIONAL CLASSIFICATION OF REVIEW EXPERIENCE
*	14	7.1.EXAMPLES OF WORD ASSOCIATIONS
*	16	7.2.SENTIMENT SCORE BY PRODUCT CATEGORIES
*	17	APPENDIX 1 - DEEP DIVE - APPLYING THE THEORY
*	28	APPENDIX 2 – REFERENCE MATERIAL
*	28	APPENDIX 3 – GLOSSARY
*	29	APPENDIX 4 – R PACKAGES

ABSTRACT

AFTER AN ON-LINE SHOPPER HAS RECEIVED A DELIVERY OR A SERVICE HAS BEEN COMPLETED, THE RECIPIENTS ARE USUALLY PROMPTED TO PROVIDE SOME FEEDBACK ON WHAT THEY THINK OF THE PRODUCT OR SERVICE. A SIMPLE NUMERICAL SCORING SYSTEM CAN BE USED FOR RATING BUT THIS MISSES ON WHAT THE CUSTOMER IS FEELING I.E. THEIR "SENTIMENT". OBVIOUSLY, TO BE ABLE TO CAPTURE THESE EMOTIONAL TONE/ SENTIMENTS AND IF WAS THEN POSSIBLE TO ANALYSE THESE, ANY BUSINESS OR MARKETER WOULD BECOME AN "ENABLED VISIONARY". NOW THEY CAN BECOME ONE. TO DEMONSTRATE, IN THE "SIMPLIFIED" EXAMPLE BELOW, R STUDIO IS USED TO ANALYSE APPROXIMATELY 23,000 TEXT ONLY

1. TEXT ANALYSIS

After an on-line shopper has received a delivery or a service has been completed, the recipients are usually prompted to provide some feedback on what they think of the product or service. A simple numerical scoring system can be used for rating but this misses on what the customer is feeling ie their "SENTIMENT". Obviously, it would be highly desirable to be able to capture these feelings / sentiments and if was then possible to analyse these, any business or marketer would become an "enabled visionary". Now they can become one. To demonstrate, in the "simplified" example below, R Studio is used to analyse approximately 23,000 text only reviews.

A wide range of methods is covered under text analysis, it is the process of structuring un-structured text to provide insights. A few methods that will be explored here are Sentiment Analysis, Review Emotional Classification, Frequency Word Clouds (singular words) and Bigrams (two-word clouds – to understand the context related to specific words).

2. SENTIMENT ANALYSIS

SENTIMENT ANALYSIS a form of text analytics. involves analysing words in a written review (not scores) to understand the clients overall sentiment related to the topic (products or services in this case) – e.g., positive, negative, or neutral. This is significant because it is derived by comparing the words in a cleaned-up feedback review against "opinion biased terminology" – i.e., an analysis of the emotions contained within the review text that are associated with positive, negative or neutral feelings

3. HOW WRITTEN REVIEWS CAN BE ANALYSED

01

Manually

The collected reviews can be manually analysed however, analysing reviews manually requires some specific skill sets and business / industry knowledge. The analyst must also, be able to aggregate reviews and relevant data and to then apply a consistent categorization and rating methodology. Carrying out this process manually will have pitfalls and resource requirements as it typically requires a trained and experienced person and a robust process to provide accurate insights – filtering the noise.

02

Using R Studio

A simpler and more effective way of analysing written reviews is to have your Data Analyst use some of the available Packages in R Studio. This does require an advanced skill set in data analysis but as will be demonstrated, the results are worth it.

03

Rstudio

R Studio (R) is an open-source free software environment for statistical analysis, data analysis, modelling, and graphics. It is widely used amongst statisticians and data scientists. R has thousands of available packages for data analysis with a wide support community. It can connect with a wide range of widely adopted software and databases.

To demonstrate, in the "simplified" example below, R Studio is used to analyse approximately 23,000 text only reviews obtained for a Women's Ecommerce clothing company (data was sourced from Kaggle – a Machine Learning and Data Science Community website). The 23,000 reviews covered 20 product categories with multiple variations of each product – resulting in 1,206 discrete products.

3.1 OUTCOMES FROM A SENTIMENT ANALYSIS

THE OVERALL CUSTOMER EXPERIENCE OF YOUR USERS CAN BE REVEALED QUICKLY WITH SENTIMENT ANALYSIS

After the analysis has been run, the resultant bias of the words is provided as a positive, negative, or neutral sentiment. If the analysis results in identifying products which have a high negative sentiment, it may prove beneficial to address the reasons for negativity – to help reduce or eliminate negative sentiment.

The overall customer experience of your users can be revealed quickly with sentiment analysis, but it can get far more granular too.

A positive sentiment is beneficial because:

- High customer satisfaction related to the product/service resulting in the client leaving a positive review to assist others making the decision to buy/use, hence reducing hesitancy and increasing sales.
- If improvements /elimination in the product/service are the outcome required to improve positive sentiment, this can result in fewer returns, credits or refunds as customer's will be happy with the product/service.
- Helps identify potential opportunities to adjust price points for strong (high positive sentiment) products/services.

Sentiment Trends:

- Monitor customer/reviewer's sentiment towards product/brand due to strategic changes that have been made.
- Identifying improvements in strengths of product/s/brand (compared to other products or performance compared to previous periods).

3.2. TEXT ANALYSIS REVIEW EXAMPLE:

A Sentiment Review Analysis is the process where data is structured and cleaned in a way such that it can be analysed. The analysis is performed using a readily available pretested package of code. The table below is an example of the raw unprocessed text (in .csv format) from the 23,000 reviews which were analysed.

Review Text
Absolutely wonderful - silky and sexy and comfortable
Love this dress! It's sooo pretty. I happened to find it in a store, and I'm glad I did bc I never would have or I had such high hopes for this dress and really wanted it to work for me. I initially ordered the petite small (
I love, love, love this jumpsuit. It's fun, flirty, and fabulous! every time I wear it, I get nothing but great com
This shirt is very flattering to all due to the adjustable front tie. It is the perfect length to wear with leggings
I love Tracy Reese dresses, but this one is not for the very petite. I am just under 5 feet tall and usually wear
I added this in my basket at the last minute to see what it would look like in person. (store pick up). I went w
I ordered this in carbon for store pick up, and had a ton of stuff (as always) to try on and used this top to p
I love this dress. I usually get an XS but it runs a little snug in bust so I ordered up a size. very flattering and t
I'm 5'5" and 125 lbs. I ordered the s petite to make sure the length wasn't too long. I typically wear an XS re
Dress runs small esp where the zipper area runs. I ordered the sp which typically fits me and it was very tig
This dress is perfection! so pretty and flattering.
beautifully made, lined and reminiscent of the old retailer quality. it is lined in the solid periwinkle-colored

3.3. CLEANING THE DATA

- The Women's Clothing Data Set – in a .csv file format, contained approximately 23,000 reviews covering 1,206 products.
- Before commencing the analysis, a review of the data set identifies a total number of reviews (23,000) across 5 departments (Bottoms, Dresses, Intimates, Jackets, Tops) with a total of 20 Product Categories within the departments and 1,206 Product sub-categories. Each of the 1,206 Products has their individual Stock Keeping Unit identifications or SKUs which provides specific information on the silhouette (style of the product) and/or colour.
- Note that the quantity of customer reviews has a direct correlation to the quantity of sales – i.e., you purchase the item then review it.

Feedback text can contain other words, symbols, numbers, punctuation, stemming (e.g., argue vs argues vs arguing) co joining words (e.g. for, and, nor, but, or, yet, and so) and stop words (ie words that occur frequently but don't provide a lot of insights (such as the, I, she, either etc)). Other specific words which are removed to avoid the data being skewed, for example a name of a product or business., these all have to be removed to provide greater accuracy to the insight. These can be removed using a R package (tm package).

4. DEVELOPING A WORD CLOUD

The first step in the analysis is looking at a high-level view of approximately 23,000 reviews by creating a FREQUENCY WORD CLOUD. This is done by condensing each of the reviews into a string of text and counting the times a single word appears. A FREQUENCY WORD CLOUD illustrates the most used words in the review/feedback dataset – this is very simply displayed graphically with higher the word usage, the larger the word appears in the cloud.

As illustrated in the FREQUENCY WORD CLOUD, the top most frequent words, “fit” appeared 11,416 times, “love”; 11,324, “look”; 9,304 and “like”; 7,849 all being of a positive sentiment suggesting there is an overall positive sentiment towards the product range

FREQUENCY WORD CLOUD



5. SENTIMENT ANALYSIS

TEXT REVIEWS CAN ALSO BE INTERROGATED TO DELIVER AN INDICATION OF EITHER AN OVERALL POSITIVE, NEGATIVE, OR NEUTRAL SENTIMENT, THAT IS THE COMPARISON IF THE TOTAL

5.1. OVERALL SENTIMENT SCORE OF THE ONLINE WOMEN’S FASHION COMPANY

Using the “⁽²⁾Syuzhet” Package in R (the ⁽²⁾Syuzhet package is tested code with relevant libraries of words) is used to establish whether there is a positive, neutral, or negative sentiment related to an analysed word.

Applying the ⁽²⁾Syuzhet analysis over all the 23,000 reviews provides a positive and negative score for each review. Taking the net of these scores (positive – negative) provides an overall Net SENTIMENT score for the review, greater than 0 being net positive, less than zero being net negative and net zero being neutral.

The count of positive and negative sentiment words in all reviews for all product categories is shown below, indicating an overall positive sentiment related to the company’s products, a score of +33,000, a third more positive words are used than negative.

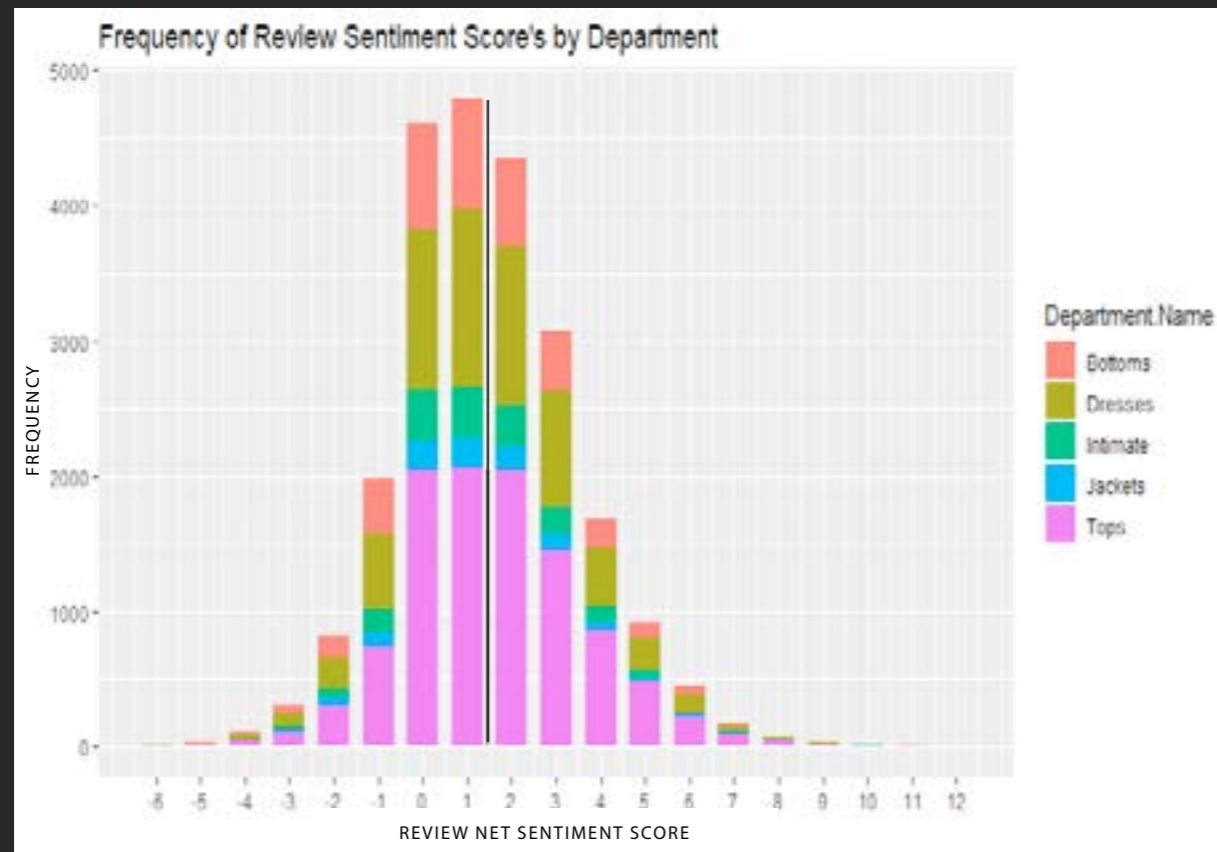
Negative Score	Positive Score	Net Sentiment Score (Positive – Negative)	Average Review SENTIMENT Score (33,865/23,486)
34,077	67,942	33,865	1.44

DISTRIBUTION OF REVIEWS NET SENTIMENT SCORES

Below shows the distribution of the 23,000 review's net SENTIMENT scores, ranging from a score of -6 to 12 across the departments. The average is 1.44.

As shown, there are a few products within the departments that have Net SENTI-

MENT score below 0. investigating this tail of the distribution (<0) to determine which product categories and respective products are driving these reviews. less than 0) of the distribution and which products are pulling down the average net SENTIMENT Score.



*Vertical line indicates the average sentiment score 1.44

On completion of the net SENTIMENT SCORE, the next step will be to investigate the clients' emotions behind the positive and negative scores.

5.2. SENTIMENT REVIEW ANALYSIS OF THE ONLINE WOMEN'S FASHION COMPANY

- EMOTIONAL CLASSIFICATION: The words used in each review are scored based on comparison with a list in the (1) NRC Emotional Association Lexicon Paper which is a list of English words that are associated with eight emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and provides an emotional context to the positive and negative reviews. Overall, the reviews are weighted towards TRUST (25%) followed by JOY (26%), ANTICIPATION (22%) – suggesting not only a loyal and returning customer base but that they are also enthusiastic about the products appearances.
- Overall, the sentiment for the products offered by the Online Women's Fashion company is positive – as indicated by the (1)NRC Emotional Association Lexicon score, the SENTIMENT SCORE is net positive by 33% (that is there are one third more positive words used then negative)

Over all products the average net SENTIMENT SCORE for a review is 1.44 where above zero is a net positive review and below 0 is a net negative review. Therefore, the average review is positive. (Summation of all the individual comment's positive minus the negative word sentiments divided by the total number of comments to give the average net SENTIMENT SCORE)

(1) Refer to Appendix for references

5.3. APPLICATION OF FINDINGS, SOME STRATEGY EXAMPLE'S:

- Through implementing strategic product/services changes, resulting in improved client satisfaction, the average overall sentiment will increase resulting in an improvement in the positive perception and ideally higher sales. It is also anticipated that the number of negative reviews will also reduce.
- If we perform further analysis on the dataset and investigate some options as to how to increase the net SENTIMENT SCORE -- e.g., what could shift the average SENTIMENT SCORE say to 1.50 from 1.44 a 4.1% increase?

POSSIBLE STRATEGIES TO IMPROVE THE SENTIMENT SCORE:

01

Strategy 1

Discontinue product categories/products that have negative/low SENTIMENT SCORES.

- This will refocus the company on higher SENTIMENT SCORE products
- Minimize production and stranded cost (e.g., marketing, samples).

02

Strategy 2

Resolve issues with high volume products with low SENTIMENT SCORE (by identifying possible issues with products and assessing if they can be resolved)

- Result in improved SENTIMENT SCORE and likely increased sales volume of the product as purchaser satisfaction will improve
- Reduce returns/credits/refunds

03

Strategy 3

Focus on marketing/sales strategy/methods to increase sales volume of high positive SENTIMENT products

(NOTE: THERE ARE LIMITATION TO THIS TYPE OF METHODOLOGY SUCH AS NEGATIVE WORDS CAN BE USED IN A POSITIVE WAY WITHIN A SENTENCE, THIS ANALYSIS DOESN'T TAKE THIS INTO ACCOUNT THEREFORE SOME REVIEWS MIGHT BE SKEWED MORE NEGATIVE THAN THEY SHOULD BE. THE ANALYSIS IS ALSO LIMITED TO THE QUALITY OF THE DATA (I.E., SPELLING))

6. BIGRAM - UNDERSTANDING THE CONTEXT RELATED TO SPECIFIC WORDS:

A FREQUENCY WORD CLOUD with the 2 most common words is called a Bigram. A BIGRAM is used to provide more context on the words that are being used consecutively or in conjunction with the words that appear the most. This is done by *TOKENISING consecutive sequences of words using a package called "TIDYTEXT." After cleaning the text of 'STOP-WORDS'. Interestingly the resulting CLOUD shown below indicates the word fit is mostly associated with PERFECT and reviewers 'HIGHLY RECOMMEND' the product.



As can be seen, the BIGRAM suggests a very positive experience with associated words such as 'SUPER CUTE' and 'LOVE LOVE' and 'HIGHLY RECOMMEND' appearing. It is notable that the word FIT is used frequently, and with the positive spin of the majority being FITS PERFECTLY. With SIZE8 and SIZE2 being mentioned indicates some smaller sizes are more popular and/or some sizing may run small and will need to be investigated.

**Refer to Glossary for more information*

7. OVERALL EMOTIONAL CLASSIFICATION OF

A deeper analysis allows the Analyst to attach an EMOTIONAL perspective to the SENTIMENT. This is achieved through scoring words used in a sentence based on their comparison to a list of words in the NRC Sentiment Lexicons within the SYUZHET PACKAGE which are associated with eight emotions (i.e., anger, fear, anticipation, trust, surprise, sadness, joy, and disgust). It provides a score for each of the eight emotions based on the presence of a

word corresponding to that emotion (The ⁽¹⁾ NRC WORD-EMOTION ASSOCIATION LEXICON Score), for example scores from this data set for a single sentence can range from 0 -10 for each of the 8 emotions.

7.1. EXAMPLES OF WORD ASSOCIATIONS

ANTICIPATION

“I SAW THIS TOP IN STORE AND IMMEDIATELY PICKED UP THE RED ONE”

TRUST

“THIS IS ONE OF MY FAVOURITE NEW RETAILER BUYS”

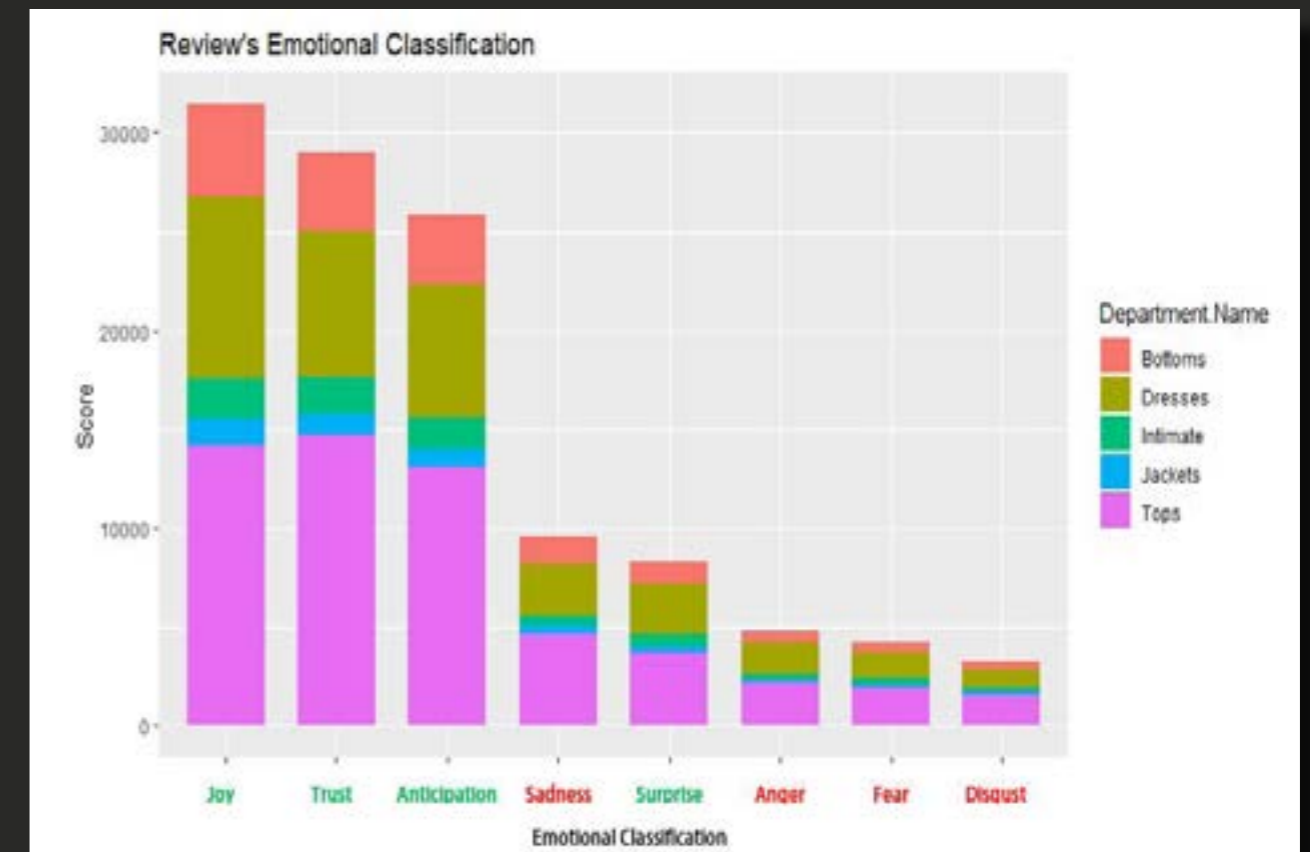
ANGER

“DO NOT WASTE YOUR MONEY ON THIS”

THE GRAPH BELOW SHOWS THE EMOTIONAL CLASSIFICATION OF ALL PRODUCT REVIEWS.

This can be used to track the impact of any operational/process change implemented (e.g., changes to product range/customer service responses/delivery times etc) by tracking results over time will indicate if the positive trend is increasing/stable/decreasing and if the negative emotions are decreasing.

BELOW, POSITIVE EMOTIONS ARE HIGHLIGHTED IN GREEN AND NEGATIVE EMOTIONS HIGHLIGHTED IN RED ON THE X-Axis:

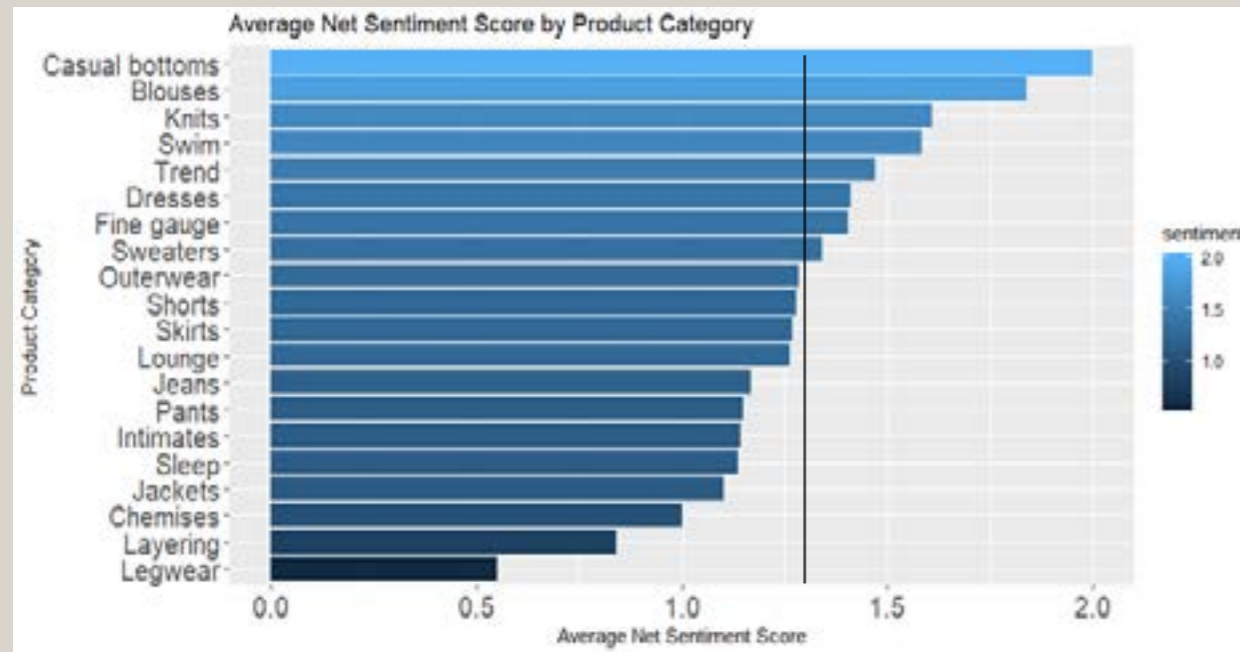


- As can be seen, in the reviews, the respondents expressed JOY, TRUST and ANTICIPATION emotions in their reviews most frequently of the 8 Emotional Classifications – suggesting a loyal and returning customer base.
- Overall, the reviews had mainly positive emotions. There could be further refinements made in the reviews associated with negative emotional classification such as DISGUST, FEAR and ANGER.
- More analysis will need to be done to assess which product category is driving a particular emotional classification or positive/negative result.

7.2. SENTIMENT SCORE BY PRODUCT CATEGORIES:

AS THE OVERALL DEPARTMENTS HAVE A POSITIVE SENTIMENT, WE WILL NOW LOOK AT THE PRODUCTS CATEGORIES DRIVING THIS SENTIMENT.

By analysing the net SENTIMENT Scores (positive and negative) using the “⁽²⁾SYUZHET” package across each product category review and then taking the average of the net of the scores (the net score being the positive value minus the negative value) identifies the SENTIMENT SCORE for each product category.



The vertical dotted line indicates the average sentiment score 1.44.

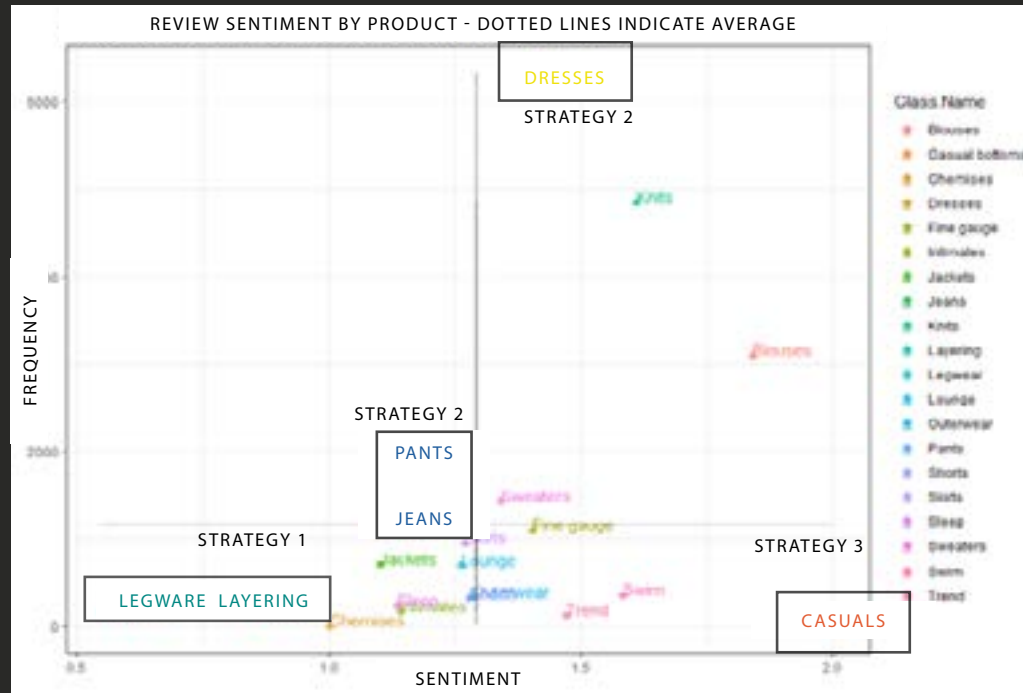
As shown:

- CASUAL BOTTOMS have the highest average SENTIMENT SCORE.
- LEGWEAR, LAYERING has the lowest SENTIMENT SCORE.
- 12 Product categories are below average SENTIMENT SCORE.

APPENDIX 1 - DEEP DIVE - APPLYING THE THEORY

INTERPRETING SENTIMENT SCORE

If we expand the analysis to look at the interplay between the number of reviews (sales) and the SENTIMENT SCORE for each product – where reviews correspond to sales – as below, we

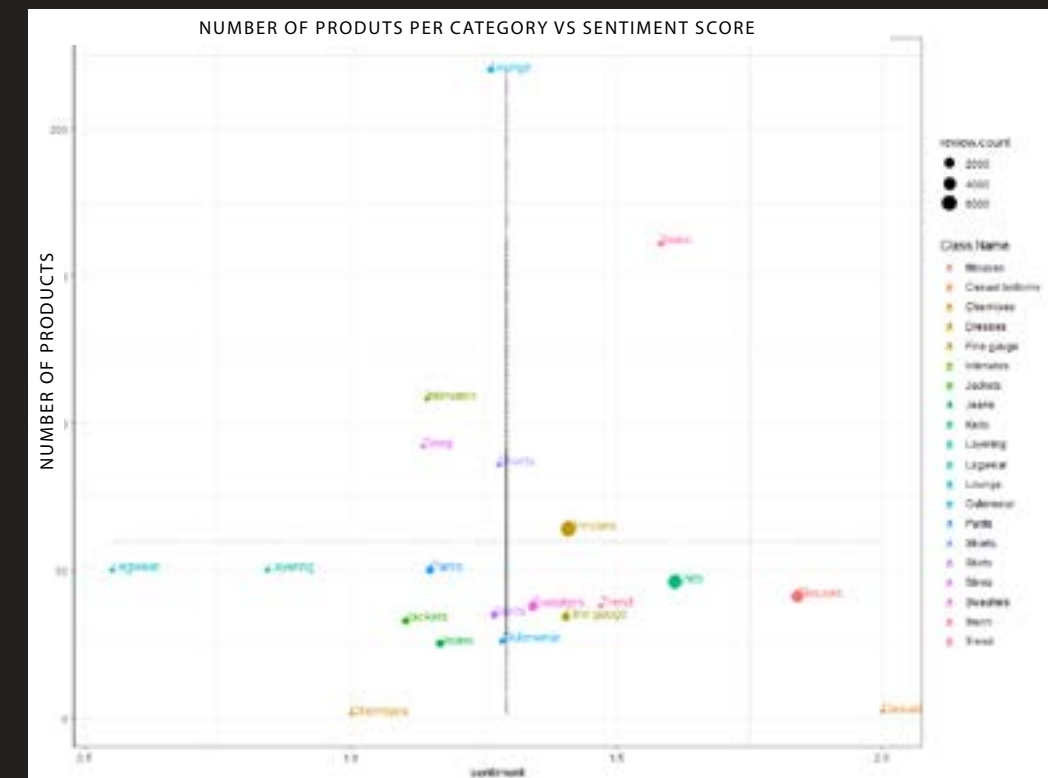


-The vertical dotted line indicates the average sentiment score 1.44.
 -The horizontal dotted line indicates the average number of reviews per product category 1,174.

- DRESSES have the largest number of reviews but not the highest SENTIMENT SCORE. Suggesting a popular reasonably well thought of product. The high review number indicating a high sales volume. The above average SENTIMENT SCORE indicating a “consistent selling product”.
- PANTS & JEANS have an above average number of reviews (sales volume) but below average SENTIMENT SCORE. Suggesting not quite as popular as dresses – could possibly be not selling as well because the SENTIMENT SCORE is below average – worth looking into to determine possible reasons for the lower SENTIMENT SCORE and explore options to improve it.
- CASUAL BOTTOMS have the highest SENTIMENT SCORE but one of the lowest numbers of reviews (sales volume). Suggesting a popular and high-quality product among a few purchasers/reviewers
- LEGWEAR has the lowest SENTIMENT SCORE and one of the lowest number of reviews (sales), suggesting the legwear products need to be reviewed to determine what needs attention – should they be discontinued – what are the carrying cost etc.

LOGICALLY, WE CAN GO ON TO A DEEPER ANALYSIS OF THESE OBSERVATIONS TO PRODUCT CATEGORIES TO IDENTIFY CAUSES OF LOWER SCORES, WHICH IF ACTED UPON COULD HELP IMPROVE THE OVERALL POSITIVE SENTIMENT SCORE:

Below we look at the SENTIMENT SCORE relative to the number of products per product category: As shown there is disparity in the range of products per product categories, indicating there may be potential to streamline product categories. Sleep, intimates, and lounge product categories have the highest number of products but low SENTIMENT and low sales volume (review counts).



Below we look at some strategies we could use to the product lines to drive a positive improvement in the SENTIMENT SCORE – resulting in improved sales (more reviews). Of course, over the complete product range, eliminating low performing products would also have the desired effect of improving the net SENTIMENT score.

STRATEGY 1: PRODUCT MIX

DISCONTINUE PRODUCTS CATEGORIES/PRODUCTS THAT HAVE NEGATIVE/LOW SENTIMENT SCORES (E.G., PRODUCTS WITH POOR SILHOUETTE FIT, POOR STYLE, MATERIAL ETC)

- Legwear – what legwear products are driving the sentiment down?
- Explore the impact discontinuing 5 Product categories with below average SENTIMENT SCORE with low review volume (<250 reviews, < 4 reviews per product where average product category is 1,174 and 25 respectively)– these are undergarment type products LEGWEAR, LAYERING, CHEMISES, SLEEP, INTIMATES (apart from lounge and swim this is majority if the intimate department product offering) – These Product Categories also have

Investigate 4 Products (SKUS) (140,264,281,169)

- Negative SENTIMENT Score Products: Products 140,264,281 have lowest SENTIMENT SCORE and only 1 review each – sales volume) therefore suggested to be discontinued to minimize production costs.
- Highest Review Products with Low Sentiment Score: Product 169 (32 reviews) has the highest volume for the product category, its low score due to some reviewers finding issues with the fleece lined leggings. If we look at the WORD CLOUD, we can see that being itchy or the fit isn't overly tight explains the reason:

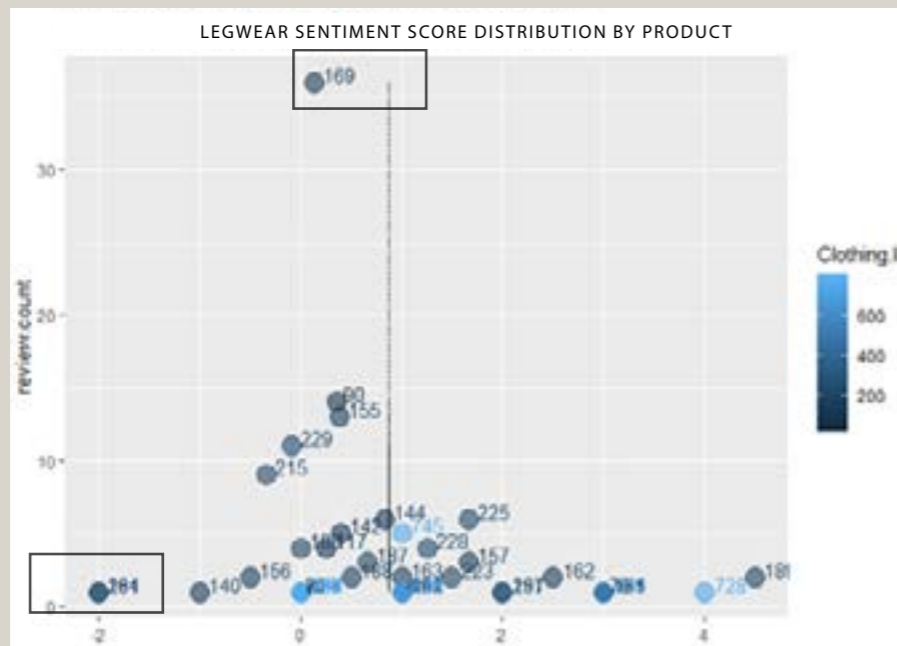


LEGWEAR REVIEW SENTIMENT DISTRIBUTION BY PRODUCT: A LOW NUMBER OF REVIEWS (165) AND OFFERS 50 PRODUCTS.

Legwear Review Sentiment Distribution by Product: a low number of reviews (165) and offers 50 products.

STRATEGY 1- LEGWEAR : SUMMARY

- As the average product reviews number is 1,174 and legwear has a total of (165) reviews, with an average review per product of 3– this suggests that discontinuing this product category and refocus on higher performing product varieties would be appropriate.



-The vertical dotted line indicates the average sentiment score 0.55.
 -The horizontal dotted line indicates the average number of reviews per product category 3

Impact of discontinuing 5 Product categories with below average SENTIMENT SCORE and low review volume (sales) on overall average SENTIMENT SCORE?

If we remove LEGWEAR, LAYERING, CHEMISES, SLEEP, INTIMATES which combined have a total of 694 (less than 3 reviews per product) reviews from the total of 23,000 reviews or 3% of total reviews: Discontinuing these results in an Overall SENTIMENT SCORE moving from 1.44 to 1.46 a 1.4% increase.

However, these product categories contribute to 300 SKUs of the 1,206 discrete products offered (25%), resulting in a savings to production costs.

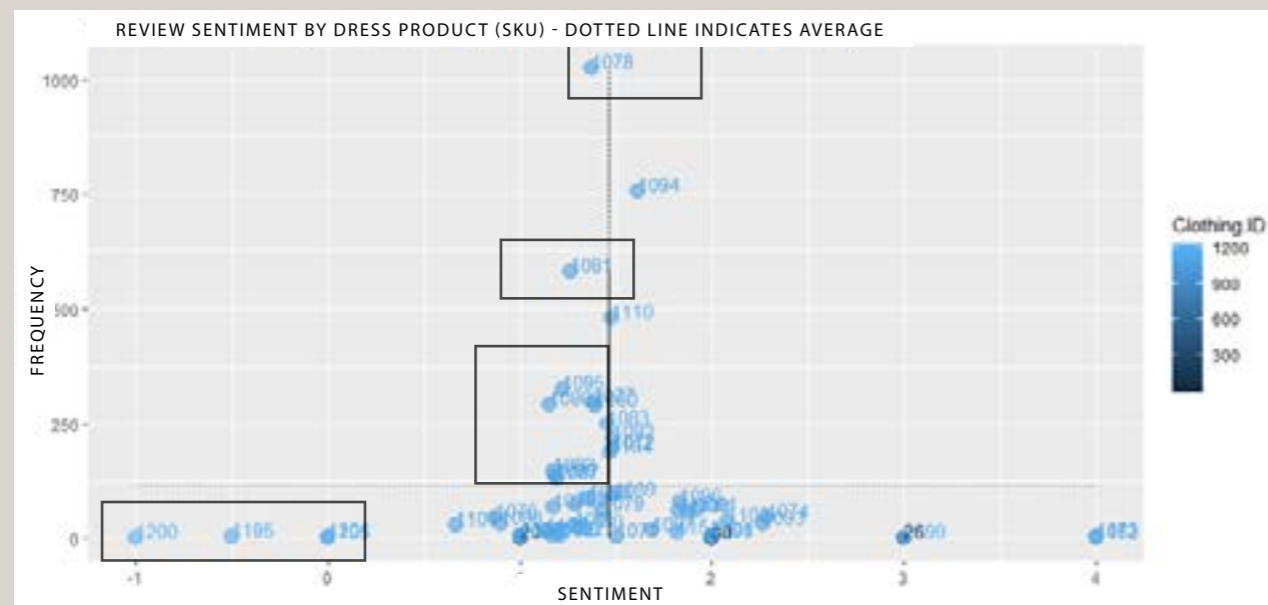
THEREFORE, THE IMPACTS OF DISCONTINUING HIGHER VOLUME PRODUCT NEEDS TO BE INVESTIGATED ON ITS IMPACT TO OVERALL PERFORMANCE.

STRATEGY 2: HIGH VOLUME PRODUCTS

RESOLVE ISSUES WITH HIGH VOLUME PRODUCTS WITH LOW SENTIMENT SCORES I.E., TO THE LEFT OF THE VERTICAL DOTTED LINE

- Low sentiment products with high-volume product categories such as DRESSES. As Dresses have the highest Sales Volume (6000 Reviews) but only the 7th highest SENTIMENT Score suggesting possible issues within some high-volume dress products. Resolving this could impact the overall sentiment score.
- Low sentiment products with above average volume (horizontal dotted line average volume of 1,174) product categories PANTS and JEAN

DRESSES –HAVE THE HIGHEST NUMBER OF REVIEWS – 6000 (CORRESPONDING TO SALES VOLUME) HOWEVER NOT THE HIGHEST SENTIMENT SCORE.



-The vertical dotted line indicates the average sentiment score 1.40
 -The horizontal dotted line indicates the average number of reviews per product category 99

INVESTIGATE 9 PRODUCTS (SKUS) (1200,1195,1205,1078,1081,1095,1077, 1086,1080)

- Negative SENTIMENT Score Products: 1200, 1195, 1205 products have low review volume therefore discontinue
- Highest Review Product & Below Average Sentiment Score: 1081, 1078:

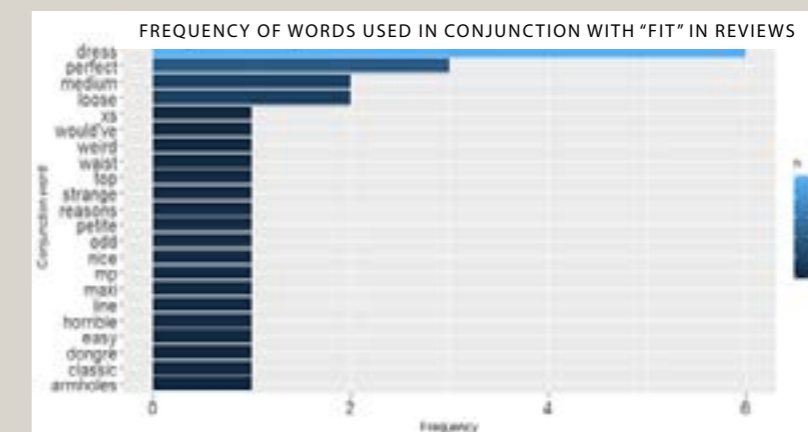
SKU 1078

- 1078 (1024 reviews) as shown in the below Bigram fit is mentioned a lot, with the fit being tight up top with mention of 'petite size', 'rib cage', 'bust line', 'bit tricky'. Filtering fit on the right, shows some think its 'odd', 'petite' and 'strange'.

SKU 1078 BIGRAM CLOUD



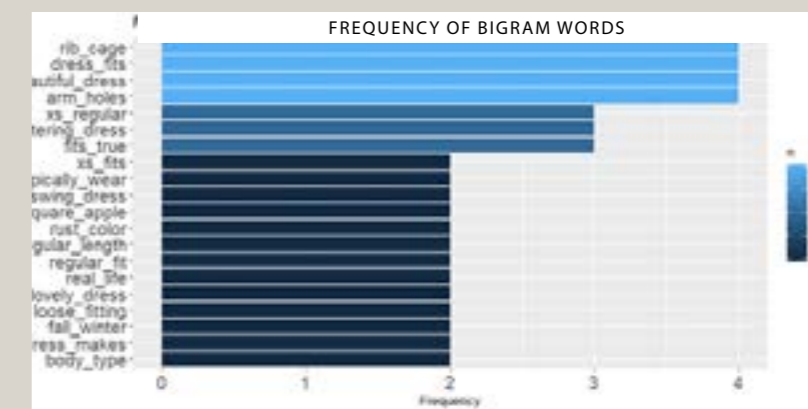
SKU 1078 BIGRAM GRAPH



SKU 1081

1081 (582 reviews) again this dress has issues with sizing and fit, the graph to the right identifies the frequency of consecutive words 'rib cage' a 'square apple' and 'body type'.

SKU 1081 BIGRAM GRAPH



BELOW AVERAGE SENTIMENT AND ABOVE AVERAGE VOLUME DRESS PRODUCTS:

- Dresses: 1095 (327 review), 1077 (297 reviews), 1086(291 reviews), 1080 (249 reviews)
- As indicated by the Bigram frequency graphs below: it appears the size curve and fit needs to be reviewed for these products for lower and higher size ranges as there are mention of lower sizes 4,6,2 and larger sizes 10,12 being pointed out with comments such as comments 'rib cage', 'petite size', suggesting people are commenting on fit because they may have read previous reviews about sizing and want to provide information around their experience (however customers that purchases in the middle of the size curve appear to have experienced better

STRATEGY 3: IMPROVE SALES VOLUME

- Focus on methods to increase volume of high positive SENTIMENT SCORE products
- If we look at CASUAL BOTTOMS and BLOUSES, both showing well above average (vertical line) SENTIMENT SCORES but having low sales volume more investigation is required to investigate driver of low volume.

SUMMARY OF REVIEWED STRATEGIES

TO MINIMIZE DISCONTINUATION OF PRODUCT GROUPS AND FOCUS ON PRODUCTS TO ACHIEVE AN AVERAGE SENTIMENT OF 1.5 WOULD BE TO :

Discontinue Product Categories:

- LEGWEAR, LAYERING, CHEMISES, SLEEP, INTIMATES product categories as they have least amount of reviews (volume) and highest product range, resulting in cost saving of production and other associate stranded costs (e.g., marketing, samples).

Discontinue products in product categories:

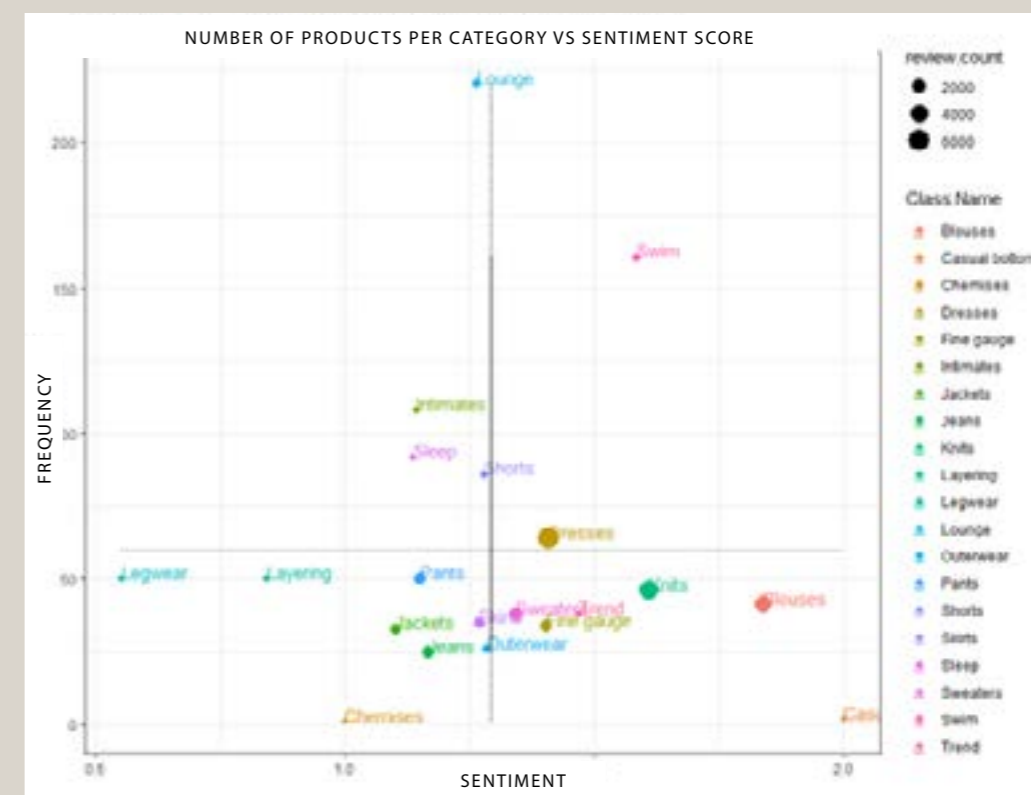
DRESSES (1200, 1195, 1205, 1095, 1077, 1086, 1080) to drive volume in other DRESS products

Resolve issues in:

- DRESSES (1078,1081) or for next season to increase sales and reduce returns and credits
- PANTS (32 of 50 products) AND JEANS (17 of 25 products) to increase sales and reduce returns and credits

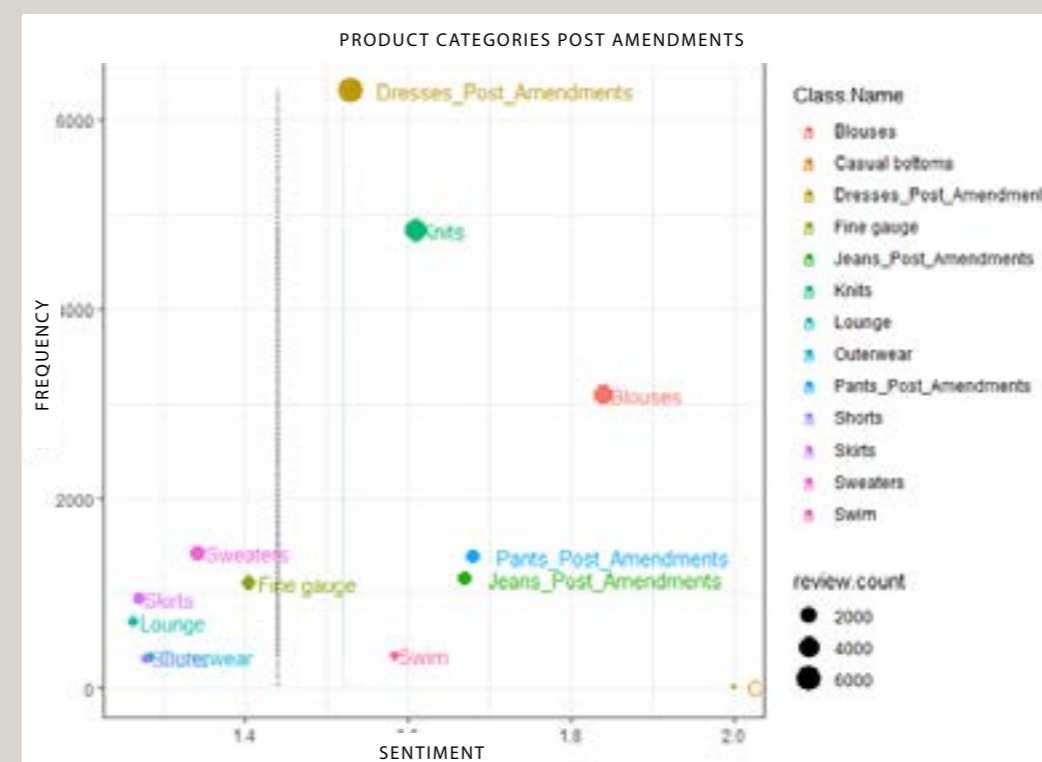
Resulting in overall SENTIMENT SCORE from 1.44 to 1.53 (6% increase), increase in sales of other products, reduction in returns and credits, and savings in cost of production and stranded costs related to discontinued products.

PRE AMENDMENTS - PRODUCT CATEGORIES SENTIMENT SCORE



POST AMENDMENTS - PRODUCT CATEGORIES SENTIMENT SCORE

RESULT: AVERAGE SENTIMENT SCORE MOVED FROM 1.44 TO 1.53



CONCLUSION:

- The overall Sentiment of products offered by the company is positive, where the customers not only have TRUST and JOY related to the products, but they are also EXCITED about the products appearances.
- Reviewing the tail distribution of products (low volume products) to identify which should be discontinued as this would increase the overall positive sentiment and save production costs.
- Refining/addressing (if possible) the issues related to specific products within product categories will further reduce negative scores and emotions in reviews and would help sales.
- Monitoring the SENTIMENT SCORES by product category/ SKU over time will identify the benefits of any refining/adjustments that have been made on the identified item's performance.
- This with combined with sales data and other data (marketing per product category) will provide more insights to performance of potential opportunities related to cost and margin of products.

APPENDIX 2 – REFERENCE MATERIAL

REFERENCES:

(1) NRC Lexicon Emotions Papers:

- Mohammad, Saif M. and Turney, Peter D., NRC EMOTION LEXICON, National Research Council Canada (NRC), viewed 12/8/2021, <<http://saifmohammad.com/WebPages/lexicons.html> >
- More information:
- <http://www.saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>

(2) Syuzhet R Package Information

- Matthew L. Jockers , Nebraska Literary Lab, viewed 12/8/2021, <<https://www.rdocumentation.org/packages/syuzhet/versions/1.0.6>>

APPENDIX 3 – GLOSSARY

GLOSSARY:

- Tokenisation - is breaking up text into smaller units called tokens which can be individual words, phrases or sentences. In this case words or consecutive words are used.
- NRC Lexicon Emotions scored words based on the comparison of eight emotions (i.e., anger, fear, anticipation, trust, surprise, sadness, joy, and disgust).

APPENDIX 4 – R PACKAGES

While performing the analysis using R, the analyst can call on specialised packages to perform detailed analysis of the data.

Packages used to perform detailed analysis.

```
library(tidyverse)
library("tm")
library("SnowballC")
library("wordcloud")
library("RColorBrewer")
library("syuzhet")
library("ggplot2")
library("tidytext")
library("glue")
library(DT)
library(tidytext)
library(dplyr)
library(stringr)
library(readr)
library(reticulate)
library(crfsuite)
```




VISIT

www.scopedata.net

INFO@SCOPEDATA.NET