TEXT ANALYTICS - 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 words used are either positive or negative in sentiment

OVERALL SENTIMENT SCORE OF THE ONLINE WOMEN'S FASHION COMPANY

To demonstrate, in the "simplified" example, 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.

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.

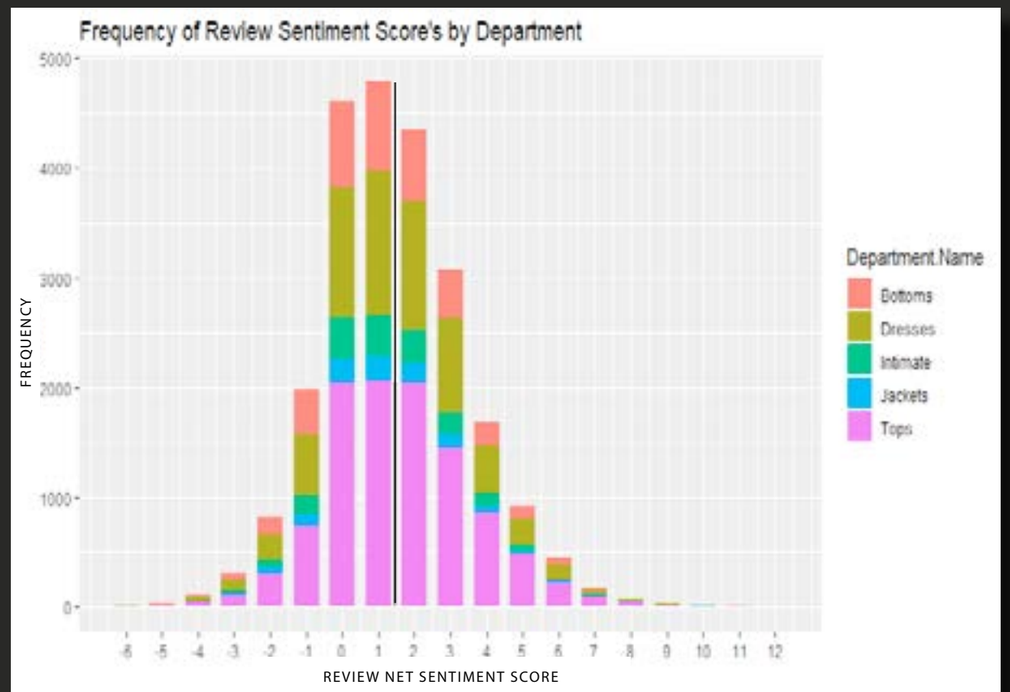
The count of positive and negative sentiment words in all reviews for all product categories is shown below, indicating an overall positive sentiment related to the company's products, a score of +33,000, a third more positive words are used than negative.

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 SENTIMENT 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.



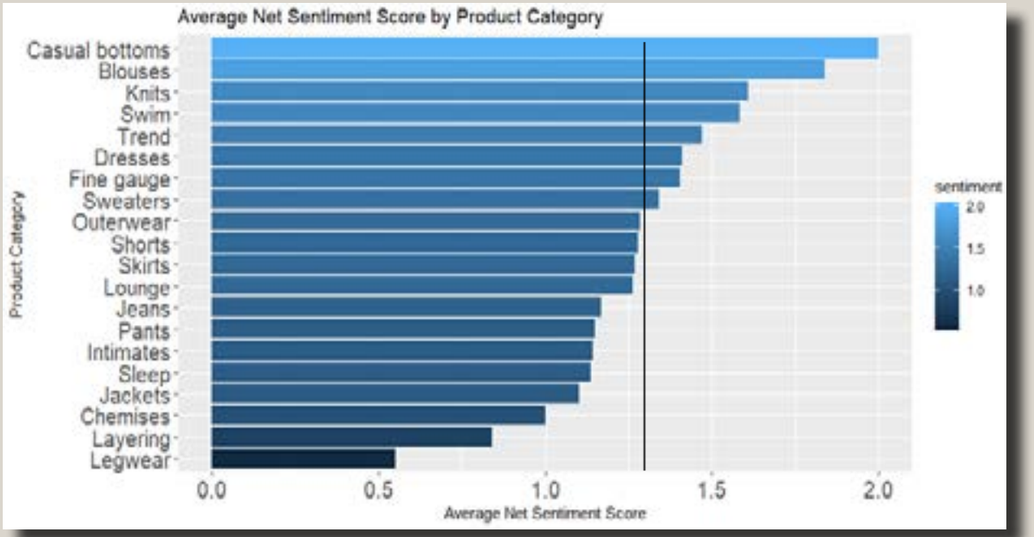
**Vertical line indicates the average sentiment score 1.44*

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

TEXT ANALYTICS - SENTIMENT ANALYSIS

AS THE OVERALL DEPARTMENTS HAVE A POSITIVE SENTIMENT, WE WILL NOW LOOK AT SOME OF 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.

TEXT ANALYTICS - SENTIMENT ANALYSIS

OUTCOMES FROM A NET SENTIMENT SCORE

Segmenting products by their net sentiment score and analysing the drivers of the sentiment can assist with potential product strategies.

Combining sentiment analysis with other text analytic methods, products and categories can be further investigated, for example:

- Address product categories/products that have negative/low SENTIMENT SCORES to improve sales or reduce stranded costs by discontinuing or minimising product range
- Resolve issues with high volume products with low SENTIMENT SCORE (by identifying possible issues with products and assessing if they can be resolved)
- 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 1 – REFERENCE MATERIAL

REFERENCES:

[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 2 – 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.

APPENDIX 3 – FURTHUR READING

- Full case Study – Text Analysis on written reviews from a Women's Online fashion company using RStudio

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(wordcloud)
library(reticulate)
library(crfsuite)

VISIT

www.scopedata.net

INFO@SCOPEDATA.NET