DEVELOPING A WORD CLOUD

A FREQUENCY WORD CLOUD ILLUSTRATES THE MOST USED WORDS IN A TEXT DATASET

- 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.
- 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.
- The first step in the analysis is looking at a high-level view of the 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.

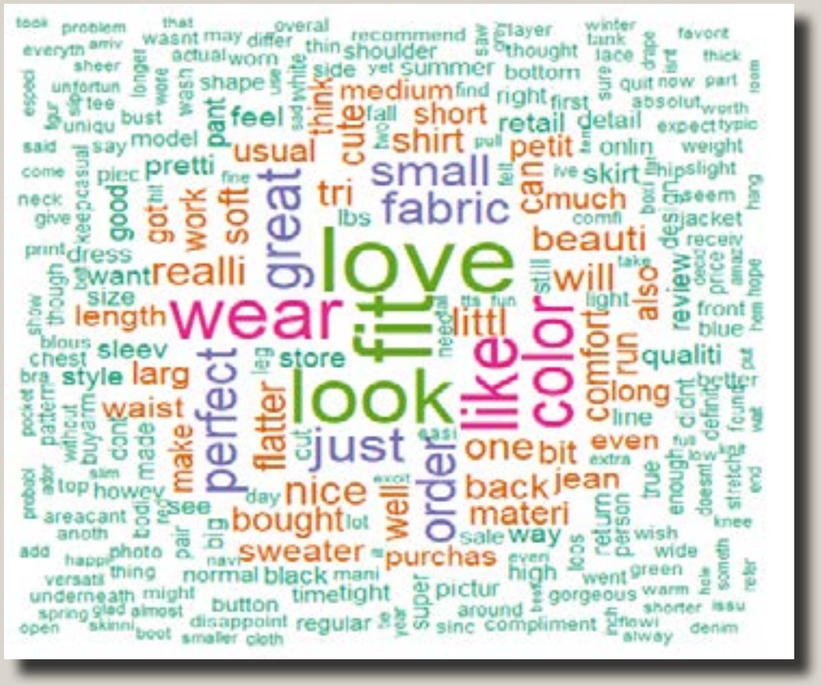
CLEANING THE DATA

- 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 6 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.

DEVELOPING A WORD 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 rang

FREQUENCY WORD CLOUD



THE HIGH FREQUENCY WORDS SUGGEST AN OVERALL POSITIVE SENTIMENT

DEVELOPING A WORD CLOUD

OUTCOMES OF A WORD CLOUD

- Condenses 23,000 written reviews into a high-level overview of how customers respond to the brand/company.
- Can be used for induvial products, departments.
- Identify categories that may need to be identified within the data set. In this case categorize reviews into categories such as
  - Fabric Reviews – reviews related to fabric
  - Fit Reviews – reviews related to fit
  - Colour Reviews – reviews related to colour
  - Order Reviews – reviews related to the order
  - Where further analysis can be performed on these categories.
- Combined with other data to determine aspects related to high sales volume products

APPENDIX 1: FURTHUR READING

- Part 1 - DEVELOPING A FREQUENCY WORD CLOUD
- Part 2 – DEVELOPING A BIGRAM
- Part 3 – Text Analytics – Sentiment Analysis
- Part 4 - Text Analytics – Emotional Classification
- Part 5 –Full case Study – Text Analysis on written reviews from a Women's Online fashion company using RStudio

APPENDIX 2 - 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")

VISIT

www.scopedata.net

INFO@SCOPEDATA.NET