**BRAND OVERVIEW**

Stats

-# products

-# reviews, avg # reviews/product, avg rating, most reviewed product (+#)

Tables/figures

-Distribution of: (i) ratings across all products, (ii) average ratings by product, (iii) rating counts by products

-# reviews per month & avg rating per month

-[Automated] text form summary of metrics, i.e. “90% of reviews are rated above 4 stars”

-Pictures of top products & brand logo

**PRODUCT LEADERBOARD**

Tables/figures

-Product all-time leaderboard

># ratings, avg rating, stdev rating, first & last rating date, avg # ratings/month

-Product trends

>compare last K months with previous K months – record n1,n2,increase in avg rating

**NLP STATS & COUNTS**

Tables/figures

-Simple statistics

>Length: number of characters, words, sentences.

>Length: Average word length (in chars), average sentence length (in words)

>Diversity (<https://en.wikipedia.org/wiki/Diversity_index>)

>>number of unique words, density of unique words, Shannon Entropy, Simpson Index

-Preprocessing:

>number of punctuation, whitespace, capitals/lowercase, unicode/emojis

>stopwords: number of and most common stopwords

>length after preprocessing (number of words)

-Rank-frequency plot for words + either show word on hover or put barchart/datatable on the side

>fit zipf distribution

-Word cloud and comparison cloud for positive vs negative reviews

-Most frequent words (by count or tf-idf) + bigrams + trigrams

**SENTIMENT ANALYSIS**

Tables/figures

-Review leaderboard (i.e. rank reviews by most positive to least)

>metadata (product name, text, review date), sentiment score, helpful votes

>could also include average sentiment in the product leaderboard

-Review sentiment over time

>frequency of negative/positive words over time

-Sentiment distribution (for all products or just 1)

-Emotion distribution in all reviews & classify reviews by emotion

-Most common positive & negative words/bigrams

>bigrams which start with a negation word (not, no, without)

>bigram network

>can also do previous 2 with pairwise correlation instead of bigrams

>Word cloud and comparison cloud for positive vs negative sentiment

-# positive & negative words for each product + ratio

-Vector space plot colored by sentiment, emotion, helpful votes, etc.

-Allow user to input a word or set of words & then get counts (word must be in vocabulary) and other statistics

>might be useful to do things with synonyms (find all words similar to “taste” and get counts)

**TOPIC MODELING**

Tables/figures

-Vector space plot colored by clusters

-Cluster sizes, number of unique words in each cluster, top words in each cluster

**PRODUCT COMPARISON**

-

**REFERENCES**

**Product-aspect mining**

=[Feature-based Customer Review Mining](https://nlp.stanford.edu/courses/cs224n/2007/fp/johnnyw-hengren.pdf)

=[A Review of Natural Language Processing Techniques for Opinion Mining Systems](https://www.researchgate.net/profile/Shiliang_Sun2/publication/309691845_A_Review_of_Natural_Language_Processing_Techniques_for_Opinion_Mining_Systems/links/5a628c3caca272a158199ea4/A-Review-of-Natural-Language-Processing-Techniques-for-Opinion-Mining-Systems.pdf)

=[Opinion Mining of Customer Feedback Data on the Web](https://dl.acm.org/doi/pdf/10.1145/1352793.1352842?casa_token=eMWF8uL_zFoAAAAA:1dtIUtRbnanDw3M39oeSTcc-GcAFvCWsK85os9EtCUgkc6dZSh-FE8L-Vs_tCKGcKR2_fbw053bSEQ)

=[An overview of keyword extraction techniques](https://www.r-bloggers.com/2018/04/an-overview-of-keyword-extraction-techniques/)

=[Domain-Specific Aspect-Sentiment Pair Extraction Using Rules and Compound Noun Lexicon for Customer Reviews](https://www.mdpi.com/2227-9709/5/4/45)

=[Text Mining Customer Reviews For Aspect-based Restaurant Rating](https://arxiv.org/abs/1901.01642)

**Noun-adjective pairs**

=<https://stackoverflow.com/questions/51308482/wish-to-extract-compound-noun-adjective-pairs-from-a-sentence-so-basically-i-w>

=<https://stackoverflow.com/questions/17251156/finding-out-adjectives-describing-a-noun-using-stanford-nlp>

=<https://stackoverflow.com/questions/53493052/any-elegant-solution-for-finding-compound-noun-adjective-pairs-from-sentence-by>

**Word relevance**

=<https://towardsdatascience.com/6-tips-to-optimize-an-nlp-topic-model-for-interpretability-20742f3047e2>

=[LDAvis: A method for visualizing and interpreting topics](https://www.aclweb.org/anthology/W14-3110.pdf)

Keywords:

-Text mining, opinion mining, customer review, customer feedback, NLP, keyword extraction, finding noun-adjective pairs,

**R PACKAGES**

https://quanteda.io/articles/pkgdown/comparison.html

https://analyticsindiamag.com/top-10-r-packages-for-natural-language-processing-nlp/

https://cran.r-project.org/web/views/NaturalLanguageProcessing.html

tidytext

tm

NLP

quanteda

koRpus

corpus

openNLP

spacyR

udpipe

RcmdrPlugin.temis

lsa

RWeka

stringr

text2vec

wordcloud

word vectors

SnowballC - stemmer

H2O - word2vec & parallelized ML algos

Monkeylearn

NLP

Preprocessing

-Convert to lowercase

-Remove extra whitespace

-Remove tokens containing numbers OR punctuation & unicode

>careful of contractions (-'), end-sentence chars (.?!), and phrase breaks (;,--)

-Remove tokens shorter than 3 chars

-Remove pronouns (I him her)

Options

-Stopwords: No, Standard, Standard+Task-specific (& configurable pct threshold)

-Lemmatization: No, Yes

-Filter extremes: No, Yes (& configurable no\_lower,no\_upper,max\_tokens)

-Ngram length: 1, 2, 3

-Keep valence shifters: No, Yes (only applicable if stopwords/filtering)

>should be used with lemmatization (i.e. aint --> {tok: ai, n't} {lem: be, not})

>note that a token "nt" will be removed when short tokens are filtered

>15% of valence shifters contain an apostrophe

Filters

-Products

-Star rating: 1, 2, 3, 4, 5

<https://jkunst.com/highcharts-themes-collection/> # highcharter themes

<https://github.com/nik01010/dashboardthemes> # shinydashboard themes

<https://stackoverflow.com/questions/58432278/set-the-font-in-highcharter-r-plot> # highcharter plot customization

<https://stackoverflow.com/questions/53519783/set-minimum-maximum-width-for-box-with-shinydashboard-r> # bootstrap boxes

<https://stackoverflow.com/questions/19611254/r-shiny-disable-able-shinyui-elements> # disable (gray out) inputs