# LISTERINE
# TEXT MINING

*BUSN 715 - Report date: Aug 26, 2020*

**Group 5**
Hao Deng
Daniel Le
Kim-Cuong Nguyen
Florence Wang

DREXEL UNIVERSITY
LeBow
College of Business
*Business Analytics*

# INDEX

# EXECUTIVE SUMMARY

With a motivation to understand how customers communicate about Listerine on vendor and social media platforms and where perception or performance could be improved, the team performed several text mining techniques on Listerine brand.

The project aims to answer two questions with two corresponding analyses. First, the team performed sentiment analysis across Listerine product categories and extracted customer insights related to overall Listerine brand perception and its product features that receive top attention from customers. Second, a comparison was made against competitor brands to visualize customer perceptions about brands across the oral care sector. As part of the analysis, the team used two datasets named BazarVoice and Brandwatch, including about 250,000 texts that were product reviews on different vendor platforms and product mentions on social media over the period of 2018-2020.

By performing the relevant analyses, the team could produce several conclusions. First, no obvious link between COVID-19 and Listerine product sentiment was found, even though positive sentiment dominated the reviews both pre-COVID and during the pandemic. Second, the causal effect between seasonal promotion or limited-edition packaging on positive reviews proved to be vague, while the effectiveness of giving free full-size samples was still debatable. Third, some product features are key to determining user satisfaction. Despite receiving numerous positive reviews, product flavor and taste were still controversial among customers' experience. White films left inside mouth after use was also an unexpected experience. Noticeably, the portable feature such as travel-size design proved to be a decisive factor that positively impacted the product sentiment.

As a result of the findings above, the team has some suggestions to improve sales and brand image under the customer's perspective. First, bundle sale strategies could be conducted; for example, combining classic flavors with other flavors that haven't satisfied customers to test customer responses while still maintaining the sales figures. Second, Listerine should perform productivity evaluation of corresponding business strategies; for example, campaigns like giving out free samples, more frequently to maximize cost utilization. The brand is also recommended to concentrate on improving key product features to keep satisfying their customers. Moreover, smart mouthwash products and alcohol-free products are of high demands in the current market and will be among the key drivers. It is recommended that Listerine invest more into its equivalent product portfolio to win customers in this segment.

# INTRODUCTION

As an Oral Care brand under Johnson & Johnson - a US multinational corporation in medical devices, pharmaceuticals and FMCG sector - Listerine is known as the first producer of oral health care products in the market and the world's top brand in its own sector thanks to first mover's competitive advantages. The brand has positioned itself as "kills bacteria that causes bad breath & helps prevent plaque build-up & gum disease while fighting off gingivitis" and has gained great reputation as one of the top oral care brands.

As shown in J&J Annual Report 2019, Listerine represented $1.5 billion, accounting for 2% of J&J total sales. Together with the annual growth of Listerine, the brand is undeniably in a growing, concentrated and highly competitive global space where different top brands are all expanding their own product portfolios with lots of innovation. The top competitors to Listerine include but not limited to Colgate, Sensodyne, Crest, Closeup, and Oral-B.

According the latest market research by Technavio, the global mouthwash market is expected to yield a compound annual growth rate of nearly 6% during the forecast period of 2019-2023. This huge market potential puts Listerine and other brands into a growingly fierce competition, but also leaves all the manufactures a valuable opportunity to enlarge their market share by innovating more product variants.

## OBJECTIVE

Market share growth is expected to be highly influenced by the company's well generating great insights into customer voice and initiating changes. In order to have a good understanding of customer voice, the team is motivated to perform some text mining techniques on Listerine brand. The study serves two purposes:

1/ To extract customer insights related to overall Listerine brand perception and its product features that receive top attention from customers.

2/ To visualize customer perceptions about brands across the oral care sector to understand how Listerine is positioned against its competitors.

These two objectives bring the team to the final goal of helping J&J understand how customers communicate about the brand and where perception or performance could be improved.

## DATA OVERVIEW

As part of the project, the team utilized two datasets named BazarVoice and Brandwatch. The table below summarizes several key information.

| | **BazarVoice** | **Brandwatch** |
|---|---|---|
| **Sources** | reviews on Listerine website (listerine.com) | - social media posts that mention the products (Twitter, Tumbler, etc.)<br>- reviews on other platforms (Amazon, Costco, etc.) |
| **Brands** | Listerine | Listerine & 20 competitors |
| **Time Period** | 01.2018 – 06.2020 | - 01.2018 – 06.2020 for Listerine<br>- 01.2018 – 04.2020 for others |
| **Data Size** | 8,591 | 241,339 |
| **Labelled - Category** | Yes | No |
| **Ratings provided** | Yes | No |

*Figure 1. BazarVoice and Brandwatch snapshot*

Basically, what the team would do is to combine the two datasets before drilling down into how the customers perceive current oral care products. However, there are several differences between the two datasets that are worth noting before they can be combined.

- The team noted that Brandwatch dataset does not have information on Listerine product category and ratings as the BazarVoice dataset. Therefore, the team would need to find a way to fill out these missing information so that the two datasets can be combined.

- While BazarVoice is comprised of only product reviews on Listerine website at listerine.com, Brandwatch dataset includes two types of information. Besides product reviews on different vendor websites like Amazon, Walmart, Target, etc., social media posts or comments that mention products under selected oral care brands are also included. This alerts the team that a number of texts is highly likely to not be true product reviews; instead, they can be promotional posts or simply posts that mention products for other purposes like entertaining, etc. Therefore, while assigning sentiment to Brandwatch texts, the team paid attention to assigning "Neutral" voice besides "Negative" and "Positive" to reduce the impact of such non-review texts.

- The time period for Listerine and other brands are different. While data for Listerine ranges from Jan 2018 to Jun 2020, data for competitor brands ends in Apr 2020. The team noted this for the comparison between Listerine and other brands.

By looking more into the dataset, the team also noted that Brandwatch dataset comes with a lot of noise that needs to be dealt with before analysis. The noise may arise from the fact that the data was pulled from social media and vendor platforms by some certain engines. Below are some sample reviews from the two datasets.

*(listerine.com)* Please bring it back! It's the only one I like. Finding a replacement will be difficult.

*(Amazon)* n n Terrible floss, falls apart too easily n n

*(Twitter)* RT @RisasDental Is your oral health a part of your New Year's resolutions? Here's a list of the best foods for a healthy smile to start off the year.

As can be from the samples, compared to BazarVoice texts which are customer reviews on Listerine websites with more "proper" language, Brandwatch texts have more noise and requires cleaning, and also, lots of slangs are used in social media posts. Plus, there are lots of duplicates in Brandwatch dataset that are "disguised" by retweet feature and special characters.

## METHODOLOGY

After investigating the data, the team performed necessary steps below to move the project forward.

### *Step 1:* Pre-process data

The first step is to clean the data. The team did necessary pre-processing tasks such as removing duplicates, removing promotional content, etc.

### *Step 2:* Build models to fill out missing values

To combine two datasets, it was a must to assign sentiment to all Brandwatch texts and assign product category to all texts of Listerine in that dataset. The team built two models accordingly for this purpose.

### *Step 3:* Generate insights

After having a complete and clean dataset, the team conducted three analyses on:

- sentiment across Listerine product categories;

- sentiment across brands in the oral care landscape;

- sentiment over top product features.

## DATA CLEANING

There are lots of texts in Brandwatch datasets to be duplicated in some ways. These can be divided into two main types.

The first type is texts that are technically the same but disguised by special characters and retweet features. Examples are as below.

Sometimes when I watch the Redskins play football, I like to roll up a listerine strip …..
RT @chaddukes Sometimes when I watch the Redskins play football, I like to roll up a listerine strip …..

zuode: she gargled Â my balls likeÂ  they were, Â listerine
zuode: she gargled Â my balls likeÂ  they were, Â listerine ????

*Figure 2. Duplicated Brandwatch texts*

To deal with those duplicates, the team performed some steps, including removing @ mentions, removing URLs, removing special characters, removing leading and trailing spaces, etc. At this step, the dataset was reduced from 241,000 to 142,000 texts.

The second type of noise is that lots of texts seem to be promotional contents, and thus are similar but not 100% the same.

….. The best way to ensure strong dental #health is to take care of your #smile and visit Dr. Barreto regularly: https://t.co/rBfUzvLHIO https://t.co/uKRsyAGc9r
….. The best way to ensure strong dental #health is to take care of your #smile and visit our #dentist regularly: https://t.co/1gmYaBwEeP https://t.co/5LVAMtmr1g

*Figure 3. Brandwatch texts of high similarity*

To remove those texts, the team calculated pairwise cosine similarity and applied a threshold of 0.9 to remove texts that bear high resemblance. The team removed 10,000 texts and got the final Brandwatch dataset of 132,000.

## MODELLING

### Brandwatch Sentiment Assignment

The team tried several different models to assign sentiment to Brandwatch texts. Several models gave good resuls on train and test data using BazarVoice dataset but predicted poorly on Brandwatch dataset when the team randomly selected 300 texts and manually checked sentiment accuracy. The team finalized the sentiment assignment using Vader with accuracy level of 86%. Three classes of sentiment, including Negative, Neutral, and Positive were assigned. A confusion matrix was generated using the result of 300 texts that were manually checked.

**PREDICTED CLASS**

|  | Negative | Neutral | Positive |
|---|---|---|---|
| **Negative** | 88 | 8 | 4 |
| **Neutral** | 10 | 84 | 6 |
| **Positive** | 3 | 11 | 86 |

(TRUE CLASS on vertical axis)

*Figure 4. Sentiment Confusion Matrix*

The confusion matrix showed that Vader tended to predict more Neutral than Negative and Positive, which the team believed to be fair concerning the nature of Brandwatch dataset. Regarding BazarVoice, the team assigned reviews with 1-2 stars as Negative, 3 as Neutral and 4-5 as Positive. By combining BazarVoice and Brandwatch, the team got a sentiment summary as below:

| Sentiment | Data |
|---|---|
| Negative | 1,806 |
| Neutral | 127,488 |
| Positive | 10,866 |
| **Total** | **140,160** |

*Figure 5. Sentiment data summary*

## Brandwatch Category Assignment

To assign product category to Listerine texts in Brandwatch dataset, the team build a predictive model using word embeddings, training and testing on BazarVoice dataset by 70-30. The team performed modelling on 9 out of 13 categories and eliminated categories with too few samples.

```
LISTERINEÂ® On-The-Go Oral Care Products                        3727
LISTERINEÂ® SENSITIVITY                                         1069
LISTERINEÂ® NIGHTLY RESET                                        897
LISTERINEÂ® Floss Products                                       749
LISTERINEÂ® ZERO alcohol-free Mouthwash Collection               712
LISTERINEÂ® Antiseptic Mouthwash Collection                      626
LISTERINEÂ® ULTRACLEANÂ® Tartar Control Mouthwash Collection     343
LISTERINEÂ® SMART RINSEÂ® Kids Mouthwash Collection              246
LISTERINEÂ® TOTAL CARE Mouthwash Collection                      120
LISTERINEÂ® Fluoride Toothpaste Collection                        40
LISTERINEÂ® NATURALS Mouthwash Collection                        34
LISTERINEÂ® HEALTHY WHITEâ?¢ Teeth Whitening Mouthwash Collection  26
LISTERINEÂ® FLUORIDE DEFENSEâ?¢                                     2
```

*Figure 6. Frequency of Listerine product categories in BazarVoice dataset*

Figure 7 & 8 shows confusion matrix of the model on the train and test data. The model predicts fairly on both train and dataset with accuracy level of 88% and 81% respectively.

By comparing the two confusion matrices, it can be inferred that:

- The model predicts most correctly on *Floss Products*, *On-The-Go Oral Care Products*, and *SENSITIVITY* – the categories that have the highest number of samples.

- The model tends to misclassify mostly among *Mouthwash Products*. This can be understood that customers do not clearly specify features that differentiate the mouthwash collections, leading to the model finding it hard to assign the correct mouthwash collection.

| TRUE CLASS \ PREDICTED CLASS | Antiseptic Mouthwash | Floss Products | NIGHTLY RESET | On-The-Go Oral Care | SENSIT-IVITY | SMART RINSEÂ® Kids Mouthwash | TOTAL CARE Mouthwash | ULTRACLEANÂ® Tartar Control Mouthwash | ZERO alcohol-free Mouthwash |
|---|---|---|---|---|---|---|---|---|---|
| Antiseptic Mouthwash | 325 | 9 | 38 | 17 | 15 | 1 | 0 | 23 | 3 |
| Floss Products | 2 | 516 | 0 | 4 | 0 | 0 | 0 | 0 | 0 |
| NIGHTLY RESET | 13 | 0 | 516 | 14 | 20 | 2 | 2 | 37 | 15 |
| On-The-Go Oral Care | 6 | 4 | 18 | 2578 | 5 | 6 | 0 | 3 | 2 |
| SENSITIVITY | 13 | 2 | 62 | 22 | 609 | 7 | 1 | 30 | 4 |
| SMART RINSEÂ® Kids Mouthwash | 0 | 1 | 7 | 4 | 4 | 135 | 0 | 7 | 6 |
| TOTAL CARE Mouthwash | 28 | 8 | 4 | 4 | 6 | 1 | 28 | 3 | 10 |
| ULTRACLEANÂ® Tartar Control Mouthwash | 16 | 4 | 55 | 8 | 11 | 8 | 0 | 129 | 8 |
| ZERO alcohol-free Mouthwash | 10 | 3 | 40 | 12 | 11 | 8 | 9 | 43 | 367 |

*Figure 7. Confusion matrix on train data*

| TRUE CLASS \ PREDICTED CLASS | Antiseptic Mouthwash | Floss Products | NIGHTLY RESET | On-The-Go Oral Care | SENSIT-IVITY | SMART RINSEÂ® Kids Mouthwash | TOTAL CARE Mouthwash | ULTRACLEANÂ® Tartar Control Mouthwash | ZERO alcohol-free Mouthwash |
|---|---|---|---|---|---|---|---|---|---|
| Antiseptic Mouthwash | 116 | 14 | 21 | 21 | 4 | 5 | 0 | 12 | 2 |
| Floss Products | 4 | 210 | 1 | 7 | 3 | 1 | 0 | 0 | 1 |
| NIGHTLY RESET | 14 | 4 | 194 | 22 | 17 | 3 | 1 | 17 | 6 |
| On-The-Go Oral Care | 7 | 9 | 23 | 1049 | 7 | 4 | 1 | 2 | 3 |
| SENSITIVITY | 11 | 2 | 37 | 9 | 234 | 6 | 1 | 10 | 9 |
| SMART RINSEÂ® Kids Mouthwash | 1 | 1 | 3 | 4 | 0 | 68 | 0 | 3 | 2 |
| TOTAL CARE Mouthwash | 6 | 2 | 3 | 2 | 4 | 3 | 2 | 1 | 5 |
| ULTRACLEANÂ® Tartar Control Mouthwash | 10 | 2 | 17 | 13 | 3 | 2 | 0 | 52 | 5 |
| ZERO alcohol-free Mouthwash | 12 | 6 | 17 | 9 | 3 | 5 | 5 | 24 | 128 |

*Figure 8. Confusion matrix on test data*

# ANALYSIS RESULTS

## Listerine Product Performance

*How do customers think about Listerine products?*

### Better 2020

Upon the result of Vader sentiment model, Listerine has the largest number of comments across its competitors. Moreover, Listerine also has the largest percentage of positive review of all 3 years. In total, Listerine has around 85% of neutral sentiments and around 13% of positive sentiments whereas 95% of competitors' sentiments are classified as neutral. It is noted that in year 2020, the available months on BrandWatch dataset for Listerine is from January till June while the available data for rest of the brands is till April. Listerine's positive sentiment is still on the top of the leader board and almost reaches the positive sentiment numbers from previous years thanks to Listerine's promotions on Antiseptic Mouthwash and other products at the beginning of the year.
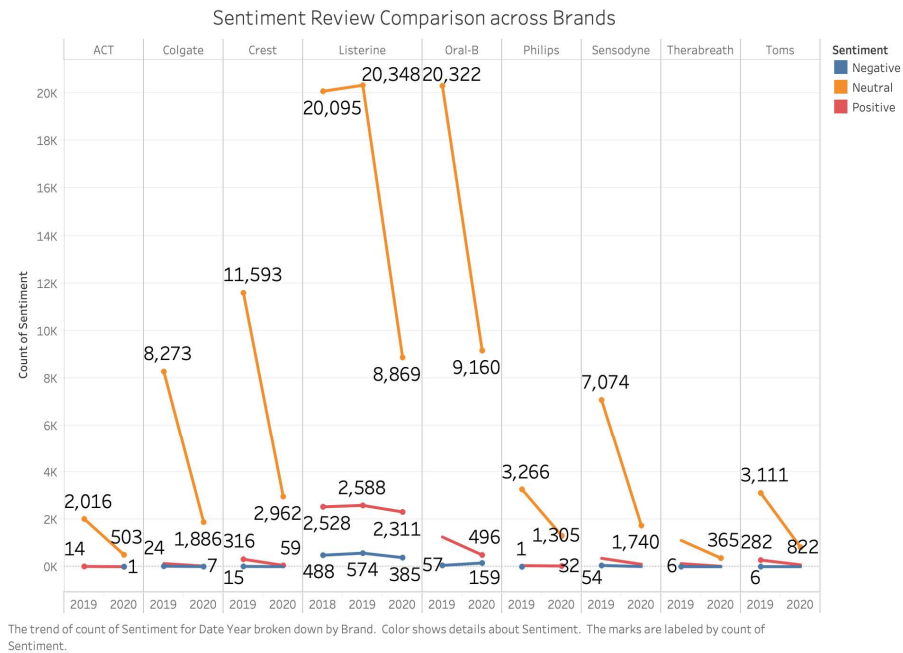


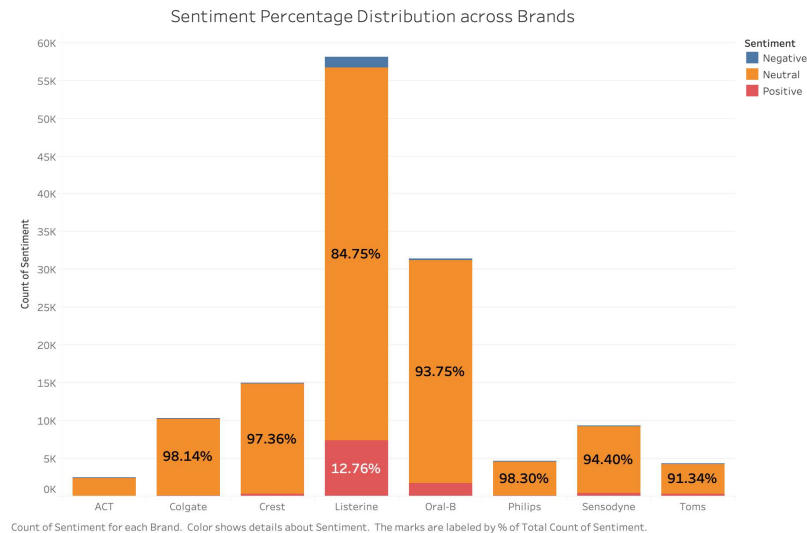*Figure 9. Sentiment review quantity across brands*



*Figure 10. Sentiment review distribution across brands*

## Seasonal trend?

The team decided to keep top 7 products instead of all 13 products since the review numbers for the last 6 products are not sufficient for analysis. On Figure 11, there are several review peaks showing among all 7 products.

Yet, none of them are seasonal behaviors but caused by special event after drilling into the output data and filter the top words around the period.
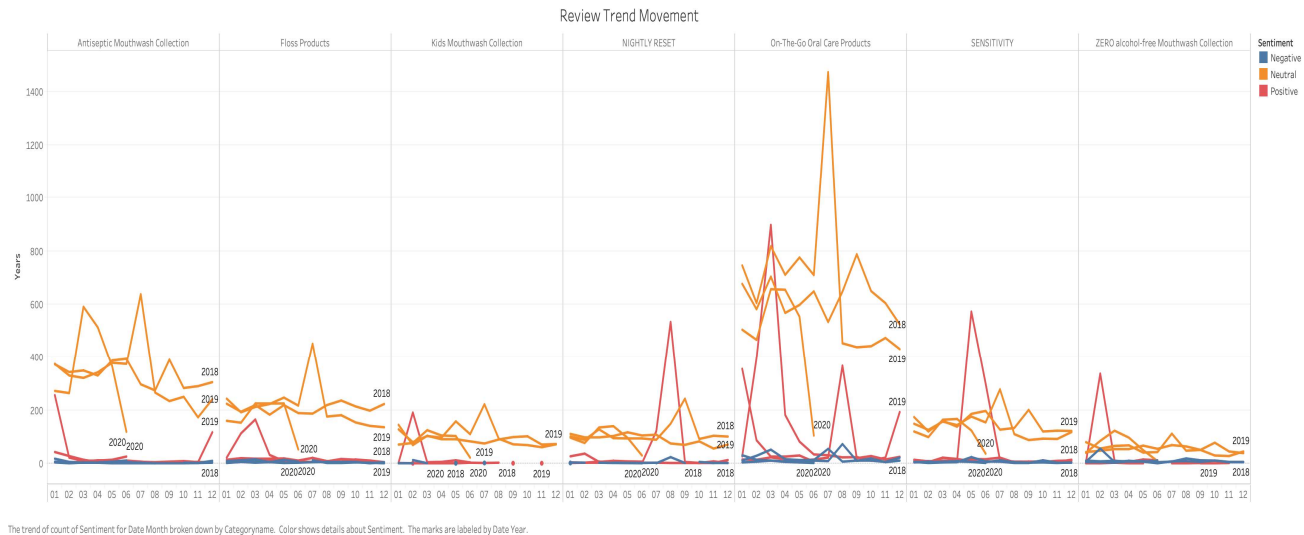


*Figure 11. Trend movement across top 7 categories*

## Antiseptic Mouthwash Collection

On the first product showing on the left of the Figure 12, Antiseptic Mouthwash Collection has two neutral sentiment peaks on July 2018 and March 2020. On July 2018, Listerine found a campaign to celebrate pride weekend targeting LGBTQ community.

The design of rainbow packaging raised the heat of discussion over internet. Another peak on March 2020, it was the time when COVID-19 had an exponential growth in United States. People were discussing whether Listerine mouthwash could help them prevent COVID-19 by killing the virus using mouthwash.
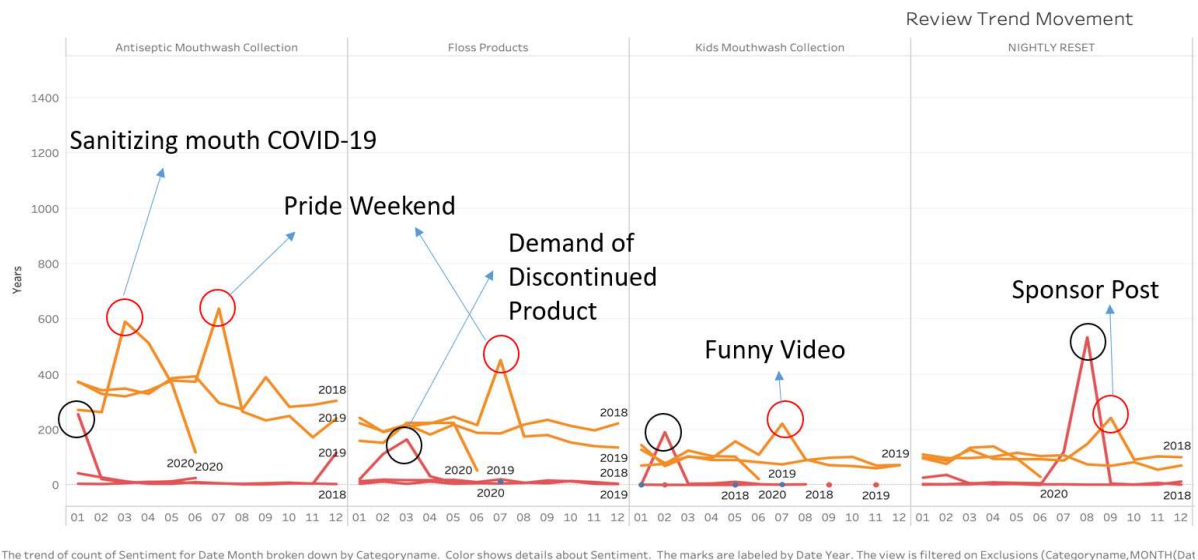


*Figure 12. A closer look into trend movement*

**Floss Products**

Same as Antiseptic Mouthwash Collection on July 2019. The rainbow packaging raised attention over the internet. Moreover, in February 2018, positive sentiment went to peak because Listerine was claiming to discontinue woven floss. The announcement caught attention from frequent users, then they protested and left positive sentiment to express the love of woven floss.

**On-The-Go Oral Care Products**

As the most popular products produced by Listerine, On-The-Go has the most sentiment peaks over 3 years. Same as other popular products, On-The-Go gains attention from the rainbow packaging. Other than that, it also gains attention from prank video from twitter. The content of video on September 2018 was about a young adult uses On-The-Go product for unusual behavior, which led the discussion about negative impression of the product.

**Did positive sentiment peak bring back positive sales?**

From 2018-2020, all positive sentiment peaks over all products except Floss products have the same reason that triggers the positive sentiment peak. Listerine has been implementing free full-size sample give out strategy across all products for years. Some potential customers/users who received free sample left their positive reviews on product page. Portion of the positive reviewers promised he/she would buy the same product again because of the good impression/usage experience of the free sample. Interestingly, all of the positive sentiment reviews from free sample give out have the similar official tone compare to the neutral and negative sentiment reviews. To summarize, Listerine is able to create sentiment

momentum using free sample give out strategy. However, the scale of the cost of the strategy and the return of increasing sales is still a question mark based on the result of sentiment analysis. Further combined analysis of sentiment data timeline and sales records is necessary. Due to the limitation of exploring sales records, we suggest J&J analysis team take sentiment analysis result and cross reference with sales record in order to discover better insights from datasets.

## Top Discussed Product Features

*The team performed further analysis to find out the features in Listerine and other brands that are discussed most frequently in reviews for better insights, based on which the team can give strategic recommendations. LDA model was applied in this analysis.*

### Most discussed Listerine positive topics

The most common topics include clean & fresh feeling, product tastes, mint flavor, chewability & swallowability, store & portability, gum care, and easy-to-use characteristics. The common topics imply that customers do care about feelings after use and convenience of products.

By referring to Figure 13 on the next page, it is easily noted that On-The-Go Oral Care products gain the highest number of reviews, especially on their chewability and swallowability, clean & fresh feeling, and mint-flavored features. Since it is of great convenience to product users to clean their mouth wherever their needs arise, portable size or travel size products become a new trend among customers. Additionally, floss products are essential for daily oral use. The convenience to store and bring, and the great gum care that the products provide to customers are also highly appreciated.
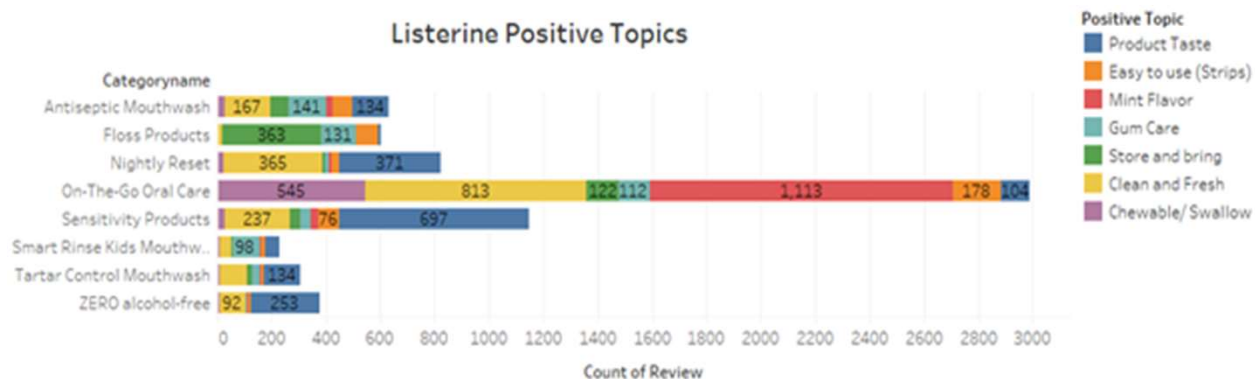
## Listerine Positive Topics



*Figure 13. Listerine positive topics*

Regarding almost all mouthwash products, the number of reviews relating to product taste occupies the prominent position in positive reviews. This means that the tastes of Listerine mouthwash products are widely accepted by customers. Therefore, it is recommended that Listerine should keep their classic tastes and come up with some bundle sale strategies to boost the sale of the products that still receives lots of conflicting reviews.

**Most discussed Listerine negative topics**

By referring to Figure 14, it is noted that the most common negative topics are white films left, mint flavor/replacement, strong flavor, bad taste and bad taste after use.

On-The-Go Oral care products also have the highest negative review quantity due to strong flavor and bad taste in and after use. Those features are not complained that much in

traditional mouthwash. As mentioned above, portable size or travel size products are trendy. Thus, Listerine may consider enhance production capacity, and adjust marketing strategies on travel-size mouthwash products to make up for the group of customers who complain about unsatisfied feelings from On-The-Go Oral Care products.

Besides, some of the mouthwash products are complained to leave white films in mouth after use; for example, Sensitivity Products and ZERO alcohol-free. In fact, Sensitivity Products are alcohol-free mouthwash products. The white films are inevitable – the result when ingredients, like sodium lauryl sulfate, react in mouth with saliva Listerine may consider inviting some dental experts to officially explain this after-use impact, eliminating customers' confusion with plaque or any other dental health disorder. Also, Listerine should put more emphasis on alcohol-free feature in products' advertisements.
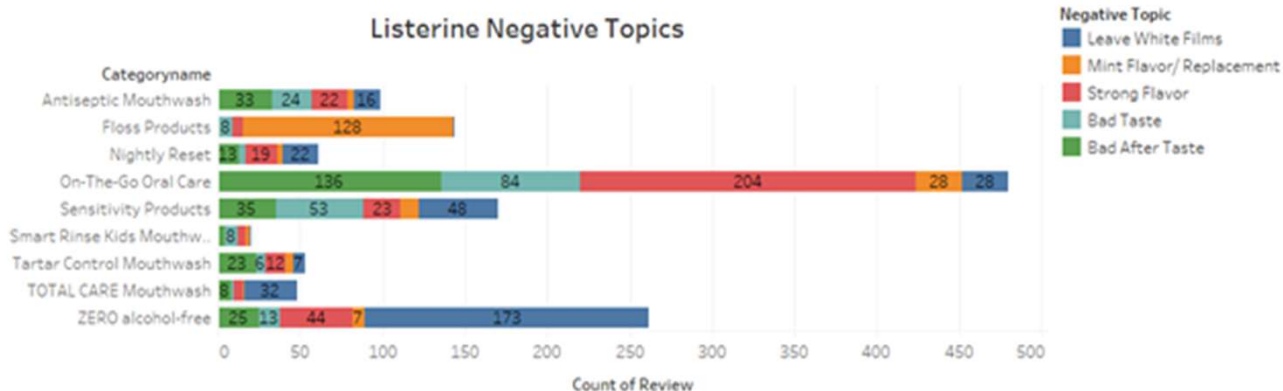
## Listerine Negative Topics
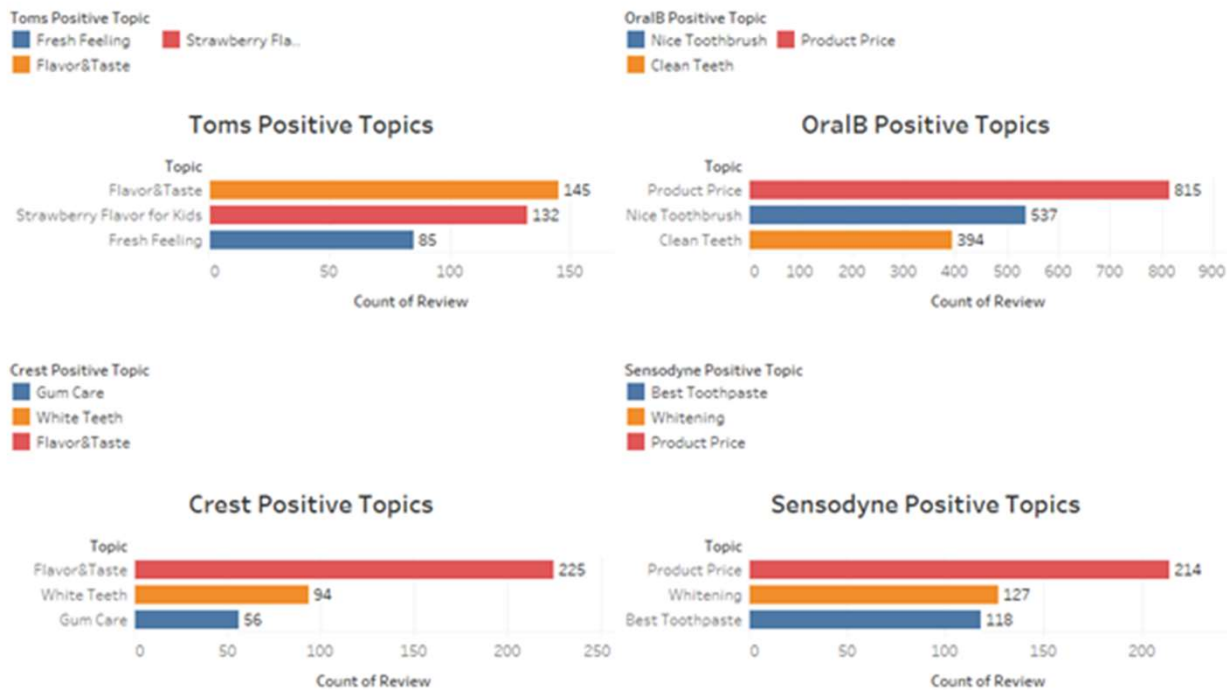


*Figure 14. Listerine negative topics*

*Figure 15. Positive topics across competitor brands*

Plus, regarding floss products, customers are in great favor of the convenience to store and bring the products. However, there is a bunch of users to complain the mint flavor and the irreplaceability feature. As a result, Listerine may consider the possibility to produce replaceable floss and unflavored floss.

**Most discussed topics on other brands**

In contrast, the most popular topics among other popular brands are price, taste and teeth whitening. Noticeably, natural ingredients and flavor are decisive features for parents when choosing oral care products for kids.

Listerine, as a mainstream mouthwash brand, has all aspects products such as multiple flavor options, Smart Rinse Kids mouthwash collection, Health White whitening mouthwash, etc., making the brand extremely competitive in the mouthwash market.

Besides, Oral-B mostly focuses on 'smart' toothbrush, which is the field that Listerine is yet to enter.

## CONCLUSION

Listerine is a well-known brand in the oral care industry with multiple product lines ranging from mouthwash, breath freshener to floss. Due to its popularity, Listerine receives a huge amount of reviews from multiple social media and online marketplace channels. The analysis objective is to derive the change in customer satisfaction over time, especially before and after the coronavirus pandemic (COVID-19), and across product categories. Besides, the analysis also aims for identifying which product features are welcomed by the customers or need to be improved.

By performing sentiment analysis on over 140,000 reviews, the team would highlight the points below:

- Although the positive sentiment is the overall theme of all reviews, there is no obvious evidence to indicate the impact of COVID-19 on the Listerine product sentiment.

- Mapping the analysis with some aligned business strategies, we could also see that seasonal

promotion or limited-edition packaging did not help much in increasing the positive reviews, while the effectiveness of giving free full-size samples is still debatable.

- The team's analysis using LDA model pointed out some highlighted product features that have vital roles in user satisfaction. The product flavor and taste are controversial among customers' experience, while some products leaving white films inside the mouth after use is also an unexpected experience. Nevertheless, the portable feature such as travel-size design is determined as a decisive factor that positively impacts the product sentiment.

## RECOMMENDATION

Based on our findings, we have some suggestions that may help Listerine improve their product sales and brand image under the customer's perspective.

- For short term strategy, Listerine should consider bundle sale strategies, such as combining classic flavors with other flavors that haven't satisfied customers to test customer responses while still maintaining the sales figures.

- In the long term, Listerine should perform productivity evaluation of corresponding business strategies more frequently to maximize cost utilization. This will help the company reduce the operation cost by not wasting resources without timely monitoring in unproductive campaigns like giving out free samples. Concentrating on improving product features stated above as leading to negative sentiment is also a compulsory next step for Listerine to keep satisfying their customers.

- According to research by Technavio, smart mouthwash products and alcohol-free products are

of high demands in the current market and will be among the key drivers of the market. It is recommended that Listerine invest more into its equivalent product portfolio to win customers in this segment.

## LIMITATION

Despite great efforts in performing this analysis, the team still encountered with several limitations.

- The amount of data used in analysis into top product features is limited, since most texts in Brandwatch dataset are assigned as neutral (nearly 90%). As part of this analysis, only 12,000 positive and negative reviews in 3 selected consecutive years were used, which means that there is a risk that the team could not cover several important features that the data could not capture.

- The assignment of product categories to Listerine texts in Brandwatch beas some limitation since the team could not check the accuracy of category prediction. Due to limited knowledge of all the covered product categories, the team found it difficult to correctly assign categories manually by ourselves; therefore, the team had to completely relied on the model for this assignment.