Team 7

APAN 5902: Data Science Consulting

Project:At-Home Hair Coloring Demand and Sentiment Post-Covid Lockdown

August, 12 2020

Section 1. Project Overview

1.1 Description and Justification

The beauty Industry was impacted due to COVID19 with the closures of hair salons. This change created demand for DIY beauty products. For instance, in the DIY hair coloring category, sales increased 115% when compared to the months before the start of the quarantine. Before COVID, the global market for hair coloring was expected to increase to 40 billion U.S. dollars by 2023. Now the areas of the U.S are starting to re-open; our goal is to determine if the time in quarantine and the risk for COVID have changed the purchasing patterns of consumers.
The goal of this assessment is to determine if there is demand for DIY hair coloring products post-lockdown and to deliver actionable data insights that will guide our client's investment strategy in the hair coloring segment. Additionally, our client wants guidance in regards to the Marketing strategy for launching a new hair coloring product.

1.2 Goals and Objectives as Documented in Initial Pitch

1. Analyze the relationship between Covid-19 confirmed cases and online sales of hair coloring products, if there is a correlation.
2. Compare Google Trends for hair coloring related keywords in 2019 and 2020, examine and analyze the difference.
3. Analyze reviews of hair dyeing products by sentiment analysis and keywords extraction, to identify the preferred features from customers' perspective.

1.3 Scope and Sources of Data

1. New York Times github: Covid Cases in the U.S. from March 2020 until August 2020.
2. Amazon reviews from the Top 4 Selling DIY hair coloring products. Reviews are from U.S. customers between October 2019 until August 2020.
3. Google Trends: Searching frequency for 'hair dyeing' from October 2019 until August 2020.

Section 2. Assumptions & Methodology

2. 1 Regression & Null Hypothesis

The initial part of our project entails an exploratory approach to determine if there is any correlation between Covid-19 cases and the Online sales of at home hair coloring products. To deliver those insights, we defined the following Null Hypothesis and Alternative Hypothesis.

Null Hypothesis: There is no correlation between online sales of at home hair coloring products and cases of Covid-19.

Alternative Hypothesis: There is correlation between online sales of at home hair coloring and cases of Covid-19.

2.2 Assumptions

1. The number of reviews from Amazon are a proxy for Sales.
2. Keyword and Sentiment Analysis were used to provide early warnings of sudden changes in consumer preference which can signal future changes in demand. Additionally, Keyword and Sentiment Analysis are providing insights to meet our client's request for Marketing insights related to product development.

2.3 Approach & Deliverables

The following analytical methods were used in this project. Each approach is reviewed in detail in section 4.

1. Regression
2. Google Trends analysis
3. Keyword Analysis
4. Sentiment Analysis

The information is presented in a dashboard which is user-friendly and allows for user selections.

Section 3. Benefits and Outcomes

3.1 Reproducible generation of results

The tools and processes presented in this project can be easily replicated and continued. The results delivered by our analysis process are reproducible, from web scraping the most recent data, to updating the dashboard. Subsequently, these actions will continue to deliver the latest insights.

3.2 Problems solved

1. Delivered improvements on current products and suggestions towards the upcoming new products launches, for example, we found customers are concerned about the relationship between hair dyeing and cancer. Thus, we propose to add more natural ingredients to let customers feel safer to use the hair dyeing products.
2. There is correlation between online sales of at home hair coloring and cases of Covid-19.
3. Compared Google Trends for hair coloring related keywords in 2019 and 2020, examine and find the difference, people are concerned about hair coloring products and health related issues, such as breast cancer.
4. Analyzed reviews of hair dyeing products by sentiment analysis and keywords extraction, to identify the preferred features from customers' perspective.

Section 4. Deliverables

4.1 Recommendations on how to use the dashboard

The Dashboard is divided into several components that provide key insights.

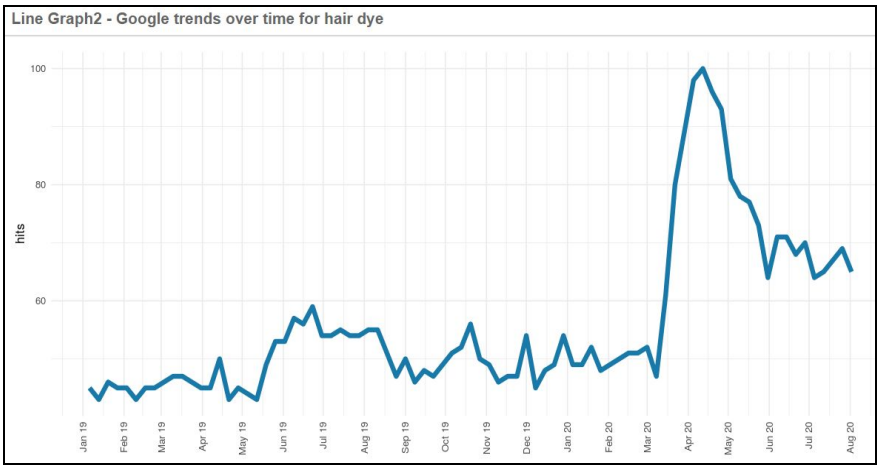
1. **Regression & Google Trends Section**

The Google Trends component is used to conduct analysis and comparative research.

The user can monitor:

- Rising Queries for hair dye
- Recognize regions of interest
- Choose the best performing keywords

- Monitor interest

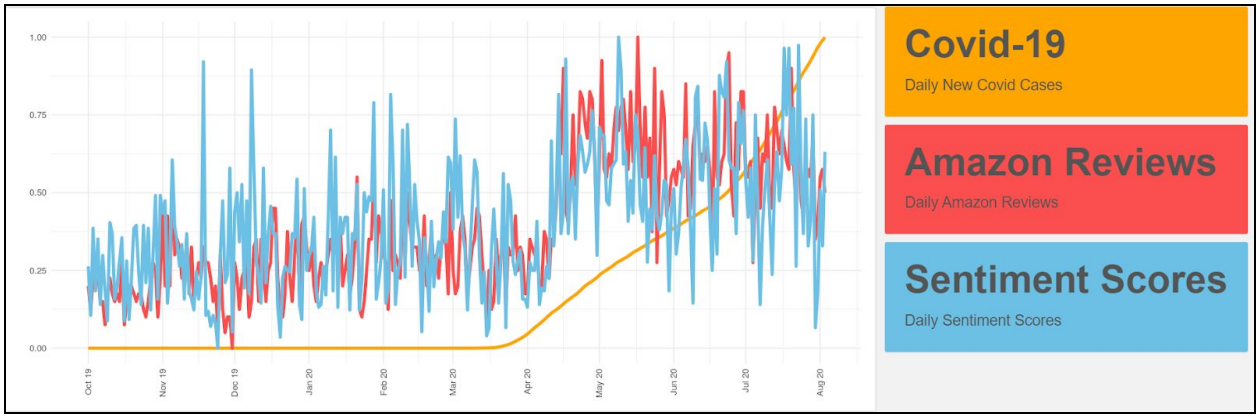


Volume trend graph of the "hair dye" search from January 2019 until August 2020, shows the sharp increase in search volume since March, and the interest stays high compared to the pre covid period.

In addition, we can see interesting insights from the "top" and "rising" queries, which we can later compare with sentiment and keyword analysis. From the top queries we can see that the most popular query is "how to dye hair", followed by "brown hair dye", "black hair dye", "blonde hair dye", "red hair dye", we can conclude that, brown hair is the most popular hair color, followed by black, etc.

From the rising queries we can see that people are concerned about the chemical in hair dye, and associated cancer risk. Over the period of little more than a year, searching query for "hair dye and breast cancer risk" increased by 950%, which shows a high level of concern.

2. Covid-19 & Trend



Features are rescaled - range in [0, 1]

The graph illustrates the daily count of Amazon reviews (Amazon reviews are used as a proxy of sale), daily sentiment scores (The higher the blue line the more positive is the sentiment score) and new Covid cases per day, between October 2019 and August 2020. If we look at the trend over time we can see the Amazon Reviews and positive sentiment increased dramatically since March, it shows a strong positive trend and positive correlations.

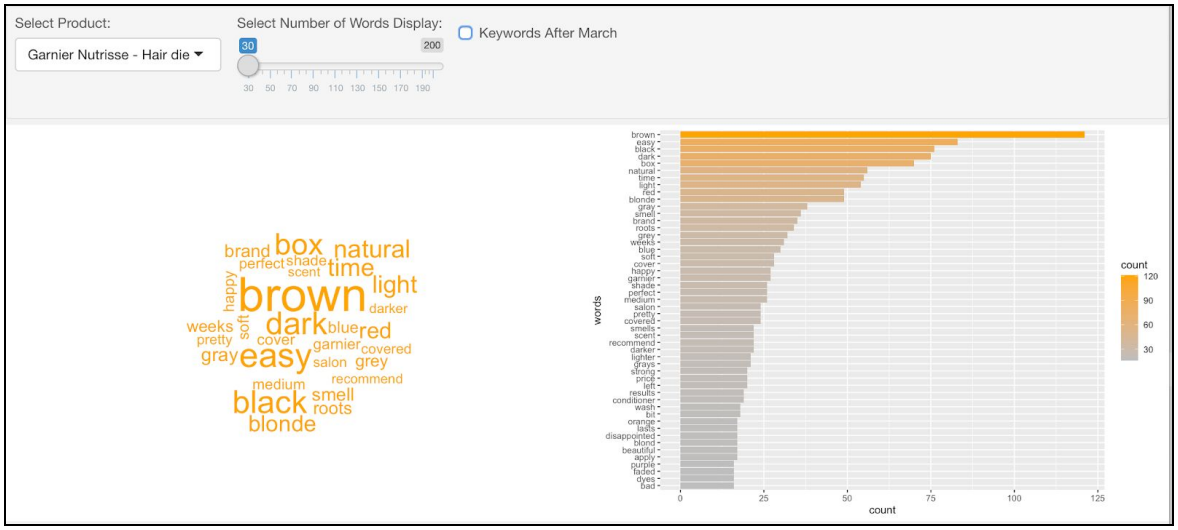
Variable	Estimate	Std. Error	t value	Pr(>|t|)	Coef.Lower.95	Coef.Upper.95
(Intercept)	7.8	0.673	11.583	0	6.48	9.12
cases	0	0	10.944	0	0	0
sentiment	0.966	0.085	11.38	0	0.8	1.133

The regression output shows that predictor variables - Covid-19 daily cases and Sentiment scores are statistically significant, because their p values are close to zero. It also shows positive correlations between dependent and independent variables and confirms our findings.

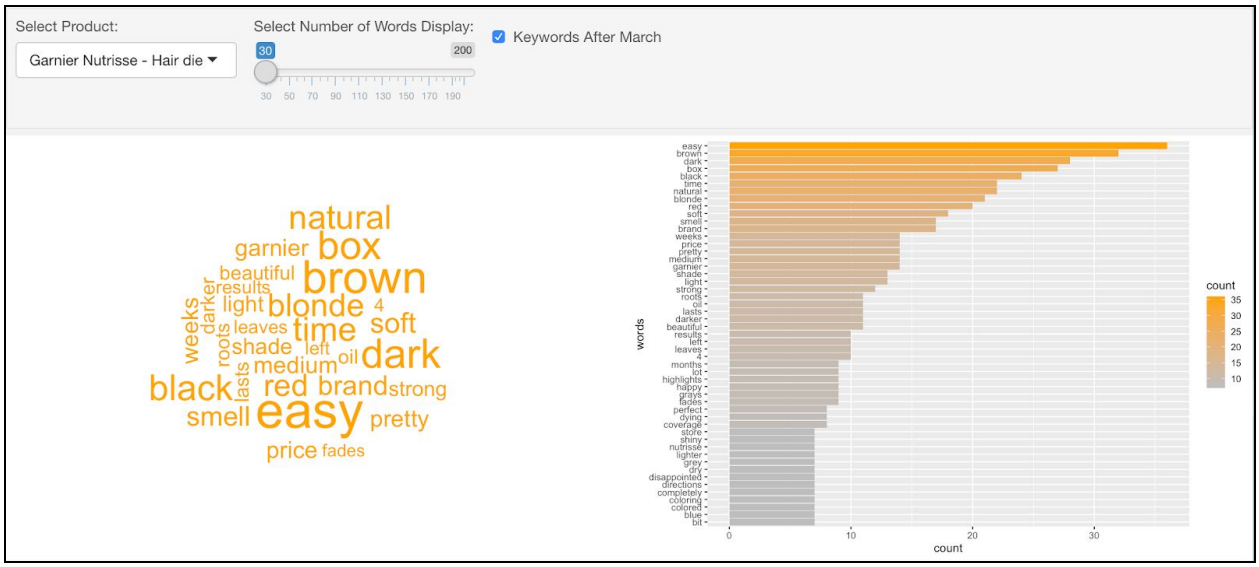
3. Keyword Analysis

Keyword analysis by analyzing word frequency could help to quantify what customers reviews are about and to measure how important a word plays. Here it also helps companies to better understand customers' feedback and demand about each brand's products.

Before starting the real analysis, we first customized stop words. In English, words like "the", "is", "of", etc. are the most frequently used. Also, since our project is related to hair coloring products, then we also removed words such as "hair", "dye", "product", etc. By doing these, we could achieve more concise results. In the shiny dashboard, we can first choose the brand and select the number of words that appear most frequently. Also, we built bar charts that provide differences of results before March (from 2019-10-01 to 2020-02-29) and after March (from 2020-03-01 to 2020-08-03) when COVID-19 became serious and people started to quarantine in the United States.



For example, we chose Garnier Nutrisse – Hair dye, and then extracted top commonly used words. We could set the number to 30 so that it shows the top 30 most frequently used words in the review data on the brand before March. Based on the above graph we noticed customers mention most about color words such as “brown”, “dark”, “black”, etc. Also, ease of use, “natural”, and “time” that a product takes have been mentioned very frequently. These information represent that customers care about the effect after doing hair dye by themselves, wanting to know if the dyed hair color is the same as the color shown on the package; concern about how much time they need to take to use products; and concern about the quality of hair dying products.



On the other hand, after checking the keywords results after March, we noticed that the word that comes up most frequently is “easy”. This change shows that what customers are most concerned about now is the convenience of using products. In addition, besides color related words, words such as “time”, “box”, “smell” appear more

frequently than before March. These differences may be due to the fact that people are more sensitive and pay more attention to these details without the help from hair stylists. We plotted bar charts showing on the right of the shiny dashboard. These charts show differences in word frequency based on customer reviews before and after March.

In fact, based on calculated results, the company could notice information such as customers' concerns and preferences - which color is most popular with customers and which kinds of color/product's quality needs to be improved. Based on the keyword analysis, companies may gain some directions and ideas on building the future products, helping to explore new markets, increase conversions, and make decisions.

Marketing is inherently analytics. Keyword analysis helps to 1) optimize spend – distribute more budget to successful keywords and reduce wasteful spending on those that have default or are not popular 2) notice market trends – knowledge of keyword search frequency provides insight into market behavior which companies could notice and apply to businesses 3) maximize usage time – after analyzing keyword performance, companies could optimized spend their time 4) find new markets – use keyword analysis to expand and discover more specific queries.

Keyword analysis is the starting point and cornerstone on search marketing and also could help parts in sentiment analysis.

4. Sentiment analysis

The purpose of sentiment analysis is to gain understanding about the current market competition, by discovering what customers are talking about for one product and its main competitors', which is beneficial for guiding future product investment and product development decisions.

In the dashboard, from user input, we can choose which two products to compare. If we choose Garnier Nutrisse and Keracolor. Regarding the positive feedback, for Garnier Nutrisse, as we can see from the word cloud below, customers are talking more about 'easy', 'strong', while for Keracolor, customers are talking more about 'soft', 'vibrant', and 'clean'. For negative feedback, both products' top negative keywords are 'dark' and 'smell', and customers of Keracolor are more likely to talk about 'smell'.

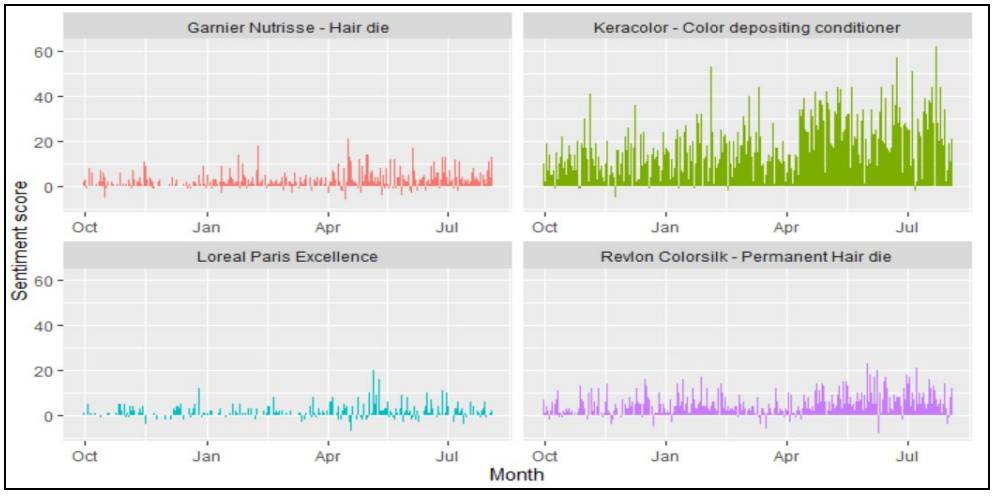
(1) Word cloud for Garnier Nutrisse

(2) Word cloud for

In addition, by comparing sentiment scores' change over time, we can find Keracolor has more customers' positive reviews and the number of total reviews, while L'oreal Paris has more negative customers' reviews. Meanwhile, the number of positive reviews/total number of reviews increased significantly after April 2020, which is an indication of the increasing purchase trend for hair coloring products due to the covid19 outbreak.

To summarize, by sentiment analysis, we gained insights about the competition landscape of hair-coloring products, the strength and weakness of each product, and how customers' attitude towards different products changes over time.

(3)Sentiment score change by brand over time



Section 5. Conclusion

In today's world, demand forecasting is impacted by constant changes in consumers' behaviors due to the proliferation of new distribution channels and new products, which is the zeitgeist of the late 2010's. Additionally, a "new normal" has been defined by Covid19. Consequently, our project aims to help investors by providing marketing recommendations in the online DIY hair coloring industry and help launch a hair coloring product that primarily sells online.

In Regression & Google Trends, we calculated p-value and drew a regression line to prove that there is a relationship between Covid-19 confirmed cases and online sales of hair coloring products. We also analyzed google trends for hair coloring related keywords in 2019 and 2020 to show that customers are concerned about hair coloring products and health related issues, such as breast cancer. By extracting customer review keywords before and after March, companies could then not only notice customers' shopping preferences but also gain some ideas on designing more successful products. Keyword analysis here helps companies to optimize spend, be aware of market trends, make better use of time, and explore new markets. In addition, sentiment analysis helps to identify key emotional triggers that drive customers' decisions; track overall customer satisfaction by viewing the positive and negative words; and upsell opportunities because with sentiment analysis, companies could easily identify customers' emotion about products and make future improvements.

Section 6. References

Nandini Suuda Siva (2019, Aug 1). *What is keyword Analysis? and How to do Keywords Analysis? – DMT*. Digital Marketing Trend. Web. Retrieved from http://www.digitalmarketingtrend.info/what-is-keyword-analysis-and-how-to-do-keywords-analysis-dmt/

Silge Julia & Robinson David (2020, Mar 7). *Text Mining with R A Tidy Approach*. Web. Retrieved from https://www.tidytextmining.com