
# Social Media Analysis Report

***On Rhode Skin by Hailey Bieber***

By Gladys Chan

# 1 Introduction

In today's digital era, social media has become one of the most essential and crucial tools that allows businesses to thrive in the competitive market. These businesses invest their time and money on popular social media platforms like Instagram, and TikTok to reach their customers, gain insights from social media data and increase brand awareness [1]. For instance, one of the most impacted industries that are transformed by the power of social media marketing is the beauty industry. According to Schmidt [2], the lockdown from the global pandemic in 2020 has shifted the society's focus towards good personal mental health and self-care. From this, many users have turned to social media platforms to buy or promote viral beauty products that would improve personal self-care. Hence, with the significant growth among consumers in social media platforms through viral skincare and makeup trends, the increase of sales and brand recognition has sky-rocketed for the beauty industry. From TikTok alone, there are at least 46 million posts with the hashtag: #beauty and 22 million posts for #skincare. Through this increasing interest in skincare and makeup, many businesses in the beauty industry have integrated social media platforms as a part of their marketing strategy to increase sales and brand recognition [3]. Additionally, not only are there rising interest in skincare and beauty among consumers, but among influencers and celebrities as well. Malis [4] mentions that in recent years, there have been trends where celebrities are curating their own skincare and makeup brands. Through the use of social media platforms and their identity as celebrities, these companies have seen significant growth in brand awareness all over the world. For example, Rare Beauty by Selena Gomez (achieving 4 million followers on TikTok as of 2024) and Kylie Cosmetics by Kylie Jenner (3.5 million followers in TikTok) have seen substantial increase in brand awareness since their first product launch because of their identity and impact in social media [5]. With the amount of strong competitors in social media, it could be daunting for small and medium beauty companies to establish themselves in a highly competitive landscape with bigger entities that includes Rare Beauty and Kylie Cosmetics as their founders have a strong impact on social media. However, with the proper use of social media analytics, not only is it possible to gain insights about the business, increase customer reach and brand awareness but also having the potential to establish a stronger social media presence in these popular platforms.

With the great impact of social media marketing, it can be concluded that almost all businesses in the beauty industry selling skincare and makeup products have an online presence [3]. However, social media marketing can only go so far if social media analytics along with machine learning methods on social media data are not utilized effectively. Although beauty brands may have the platform to showcase their products with visually appealing and attractive content on their profile to increase sales, it is not enough for them to fully understand their customers' behavior and engage with them on a personal level [6]. To combat this problem, social media analytics come into play. With data collected from the beauty brand's popular social media platforms like Instagram and TikTok, they are able to perform data analytics techniques like sentiment analysis to understand how their customer's feel about their product and brand, topic modeling to find out what product is currently trending, and engagement analysis to understand how their customers interact with their content [7]. By applying social media analytics techniques like these and many more, these beauty brands have the opportunity to create a much more personal and engaging content that resonates with their target customers and potentially establish themselves in the highly competitive market of viral beauty brands.

In this report, data analysis will be conducted on a small celebrity-owned skin-care/makeup brand's (Rhode Skin by Hailey Bieber) social media platforms: Instagram and TikTok to help them ***understand their business nature, how their target audience interacts with their brand, how the company influences their customers online, creating marketing awareness and increasing brand loyalty.*** To conduct the research in exploring how Rhode Skin can utilize social media analytics to understand their business and customers, a few social media analytics techniques will be applied such as data exploration and descriptive analysis, sentiment analysis, topic modeling and engagement analysis. The data needed for these techniques will be scraped from the three most popular social media platforms: Instagram and TikTok. This paper will first start off with exploring a literature review of topics that are related to the social media analytics techniques used and the background of the beauty brand chosen for this paper. With these compiled literatures, there can be a comprehensive understanding of how each of the social media analytics methods (chosen for this paper) functions and how it contributes to the overall research objective. Next, the methodology will show what was utilized to perform the social media analytics techniques. The methodology presents the process of data extraction, data cleaning, exploratory data analysis and the social media analytics techniques: ***sentiment analysis***, ***engagement analysis*** and ***topic modeling***. In the results, there will be a presentation of findings from the data analysis conducted in methodology and the interpretation of the results.

# 2 Methodology

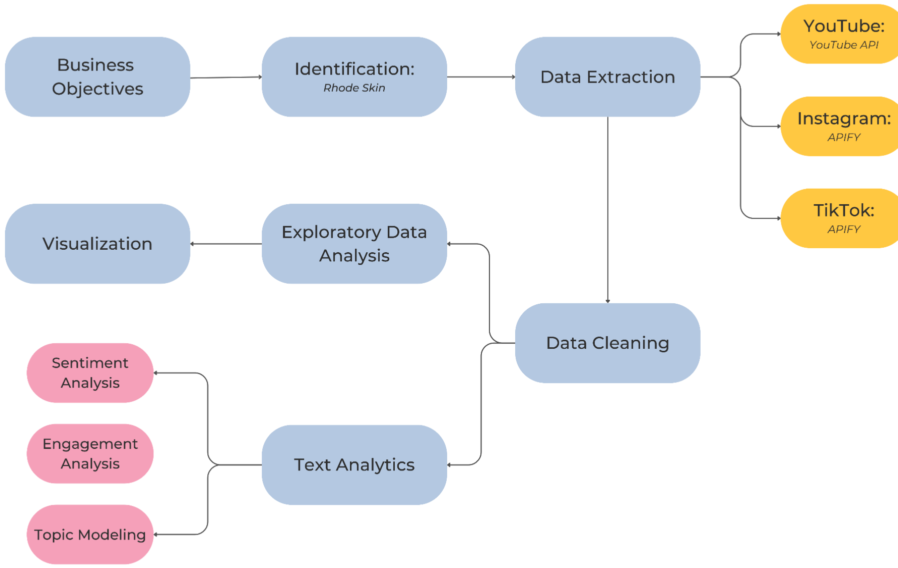



*Fig. 1: Social Media Analytics Process*

## 2.1 Business Objectives and Identification

In the introduction, one of the problems identified in an extremely competitive and ever-evolving beauty industry is that most small and medium beauty companies do not effectively utilize their social media analytics to establish themselves in a highly competitive landscape. Through understanding this problem, the business objective is identified: to aid these companies to fully understand their business nature, how their target audience interacts with their brand, how the company influences their customers online, creating marketing awareness and increasing brand loyalty using social media analytics and machine learning techniques. To conduct this research, Rhode Skin by Hailey Beiber was chosen.

## 2.2 Data Extraction

In this section, there were two different methods that were executed to scrape data from three social media platforms: Instagram, and TikTok.

### *2.2.1 Collecting Data Through External Scraper: Apify*

Because Instagram and TikTok have a strict policy against data scraping, no API is provided from the platform itself for users to scrape data from their content. However, Apify, an external web scraper service provider, provides the platform to scrape data from Instagram and TikTok (refer to Appendix A). In Apify, the developer of the scraper states that it is ethical and does not contain any private data of users unless it is publicly shown by the user. Since it is stated that is ethical, Apify will be utilized to scrape Instagram and TikTok data.

To collect data from Instagram posts and comments, and TikTok posts and comments, these scrapers were utilized:

- [Instagram Post Scraper]
- [Instagram Comment Scraper]
- [TikTok Post Scraper]
- [TikTok Comment Scraper]

The scraping process first starts off with the Post Scraper. In the Instagram and TikTok Post Scrapers, 'Rhode' was added into the input parameter and the settings for scraping the posts are from January 2024 onwards. After running the scraper, there were many variables in the dataset that included the post's captions, time created, number of likes, shares and comments, post url, etc. The data for Instagram and TikTok posts was last scraped on August 2, 2024.

From the scraped post data, the url for each post was copied and pasted into the input parameter for Instagram and TikTok Comment Scraper. This would scrape comments from each of the links provided. After running the scraper, variables like timestamp, comments, likes, etc were added into the dataset. The data for Instagram and TikTok posts was last scraped on August 2, 2024.

One limitation that occurs in this scraping process is that Apify scrapes the comments from Instagram and TikTok through their website platform without logging in into an account. Because of this, Instagram and TikTok only display a minimum of 50-100 comments for each post rather than displaying all of the comments.

### 2.3 Data Cleaning

In this section, all datasets for comments that were scraped followed a data cleaning template. Data cleaning templates for post dataset were slightly different from data cleaning for comments. Python was used for this process and the packages that were utilized for cleaning include: pandas, numpy, re, nltk, langdetect, and os. In this section, the data cleaning process will be explained.

For data cleaning in comments and post data, these are the steps taken the clean the variable 'Comment' and 'Caption':

1. Text normalization
2. Removing emojis
3. Removing non-english comments
4. Removing spam or bot patterns
5. Spelling check
6. Removing certain word patterns
7. Removing stop-words
8. Handling missing data, deleting duplicates

In the text normalization, all the comments were converted into lower-case letters. Next, emojis were removed and non-english comments like Spanish, Malay, French, etc were removed as well. Languages that contained non-latin characters/script like Japanese, Chinese, or Korean were removed. This is done using the package: Langdetect. Spam and bot patterns like 'get rich quick' were removed as well.

In the comments, there were many spelling mistakes, typos and slangs that could hinder the process of the analysis. To combat this, a function was used to check every row and correct certain words. Since it took the code a longer time to process every single row in a large dataset, only basic spelling checks were performed. Spelling mistakes like exaggerated words (eg. “loveeee”) was corrected into their original spelling (eg. “love”). In cleaning the dataset for post, not much cleaning was needed as Rhode’s captions contained no slang nor spelling mistakes.

For the post dataset, irrelevant columns were deleted, leaving only the important ones that are useful for the analysis. Missing and duplicated data were removed at the end before saving the cleaned dataset.

### 2.4 Exploratory Data Analysis

With exploratory data analysis, Python is also utilized. Packages like pandas, TextBlob, wordcloud, matplotlib were used to aid in the analysis. These processes include basic statistical techniques and visualization like descriptive analysis, user analysis, and generating word clouds. Category statistics and trends over time were analyzed only for the post dataset.

In the descriptive analysis for both comment and post dataset, descriptive statistics like count, mean, std, min, max, quartiles were calculated for the ‘Likes’ column. This will show a summary of the distribution of likes on comments and posts. For post dataset, ‘Collect’, ‘CommentCount’, ‘ShareCount’, ‘PlayCount’, and ‘VideoDuration’ were also used for descriptive statistics to show their summary and distribution for the post.

For user analysis, the code creates a bar graph visualizing who are the top users according to the amount of likes they have obtained in their comment. This visualization can help us identify which comment receives the most positive engagement or resonates strongly with other users.

To analyze the dataset's overall word frequency, a word cloud is generated to visually represent the most occurring words in the comments. On a side note, stop words were added to the word cloud graph as there were words that were not relevant to the analysis like “actually”.

In the exploratory data analysis for the post dataset, categorical analysis was performed to quantify the number of unique users that are mentioned in Rhode’s posts since January 2024. In addition, categorical analysis was also used to analyze the number of each post’s format (MP4 or JPEG). Lastly, trends over time line graphs were generated for the post dataset to visualize the number of post uploads that were published over time.

# 3 Analysis

This section details the analytical techniques that are applied to the cleaned datasets from Instagram, and TikTok. The analysis is focused on three areas: sentiment analysis, engagement analysis, and topic modeling. Each subsection will explain the process conducted for each analytical technique.

### 3.1 Sentiment Analysis

In order to gauge the overall sentiment in the comments for each social media platform, sentiment analysis was applied using the TextBlob library. From this library, we can assign the sentiment score for each comment. The dataset used for this analysis are comments data from Instagram and TikTok.

The sentiment analysis starts off with a function that would go through every comment and calculate its polarity. Once it has calculated the polarity, the function then assigns a sentiment score from a scale of -1 (negative) to 1 (positive). In addition, another function is created to categorize these scores into five sentiment categories: strongly positive, positive, neutral, negative, and strongly negative.

When reviewing what sentiment categories should be used, we concluded that certain words in ‘negative’ were due to words that were assigned a negative score when in context, it was meant for positive remarks (eg. “obsessed” was categorized as negative, but in the context of social media, it shows the user’s love with the brand). Hence, we have categorized ‘strongly negative/positive’ and ‘negative/positive’ to account for cases of negative words where it is meant positive comments. ‘Strongly negative/positive’ would show that the entire comment would potentially express a clear and overwhelming negative or positive sentiment rather than a small amount that is categorized in ‘negative/positive’.

After assigning each comment its sentiment score and category, the distribution of sentiment, sentiment over time, and sentiment by topic can be further explored.

#### *3.1.1 Sentiment Distribution*

To understand and gauge the overall user’s emotions and perception toward the brand, sentiment distribution is utilized to provide insight on this collective sentiment.

In order to visualize the sentiment distribution, the code first calculates the total number of comments that is delegated to each sentiment category. From this, a bar graph is formed to illustrate the proportion of each sentiment category relative to the total number of comments.

#### *3.1.2 Sentiment Over Time*

By analyzing how user’s sentiment evolves over time, it is possible for brands to understand how perceptions and attitudes towards a brand or topic evolve.

To visualize this analysis, the timestamp of a comment is used together with the average sentiment score. This would result in a line graph illustrating the average sentiment score monthly from January 2024 to our current present day (date where the data was last scraped).

#### *3.1.3 Sentiment by Topic*

Analyzing sentiment by topic would allow the brand to fully understand in which specific area or aspect that evokes positive or negative feelings about the brand. In this section for analyzing sentiment by topic, our team has performed topic modeling before conducting sentiment by topic. Hence, this analysis utilizes a dataset that has been preprocessed in the topic modeling (refer to section 4.3).

In the code, a loop was implemented to generate a bar graph for each topic that shows the sentiment category distribution (positive, neutral, negative) within that specific topic. This loop was given for each comment dataset per social media platform.

On a side note, while preprocessing the data in topic modeling with the sentiment category variable, the team has decided to merge the sentiment category values: ‘strongly negative/positive’ to ‘negative/positive’. This is because the section deals with generating bar graphs for every topic and it would be time consuming for the analysis.

## 3.2 Engagement Analysis

Engagement Analysis shows how the consumers interact and communicate with the brand or company through engagement. It measures the interactions of a company's content. Some of the popular engagement that we choose to include in this analysis are likes, comments, and views.  A product or services posted out on social media platforms tend to show if the consumers are attracted and interested. A good product tends to show out more likes count, comments count, and engagement on the posts depending on whether your posts are interactive, interesting, and attractive enough.

The engagement analysis would list out the most likes count and comments count of posts at selected options such as highest likes count and comments count of April, highest videos view of 2024, and more. We would then monitor the results through the analysis to conclude the engagement analysis.

After assigning each setting of engagement category, the topic average engagement rate, top performing content, engagement over time, and correlation with sentiment by topics could be further discussed and explored.

### *3.2.1 Average Engagement Rate*

By observing the average engagement rate for each post, it is easier for brands to further understand their customer base towards different types of posts.

For Average Engagement Rate, we used the '.mean()' function from pandas to calculate the overall average value of the column 'engagement_rate'. The result is then stored into another column under 'avg_engagement_rate'.

## 3.3 Topic Modeling

In this section, topic modeling will be conducted to identify the key themes and underlying topics in the 'captions' of the post dataset and 'comments' of the comment dataset. Latent Dirichlet Allocation (LDA) is a topic modeling algorithm that is chosen for this research. Once each topic has been assigned to each caption and comment, topic distribution, and topic evolution over time will be analyzed.

### *3.3.1 Applying Topic Modeling Algorithm: Latent Dirichlet Allocation (LDA)*

To discover the topics within the dataset, Latent Dirichlet Allocation (LDA) algorithm is applied using the Gensim library. In addition to this, few packages in Python were used to help in conducting the algorithm: pandas, matplotlib, nltk, pyLDAvis. The post and comment datasets from Instagram, and TikTok were used to conduct topic modeling. Throughout this process, there were a few key steps that were taken, starting with data preprocessing, followed by model training, and topic inference and visualization.

Before applying the LDA algorithm, data preprocessing was conducted to remove any stopwords and tokenize the comment and caption. Irrelevant columns were removed as well.

After the preprocessing, a dictionary was created using the comments and captions. This is to map out each word to a unique ID. The corpus, which is a list of bag-of-words representations for each comment or caption, was then generated using this dictionary.

To start off the model, a function is created to find out the optimal number of topics for the specific dataset. This is achieved by calculating the coherence scores for different numbers of topics ranging from 5 to 40. The results of this coherence score measures how interpretable the topics are. In the code for computing coherence score, a line graph was created to visually identify the optimal number of topics for the chosen dataset.

Once the optimal number of topics was determined, the LDA model was trained on the corpus using the computed optimal number of topics. The model was created with 10-15 words and a random seed added for reproducibility. The model is then saved into a Gensim file for future usage when categorizing the dataset's 'comment' and 'caption' according to its topic. To visualize this model, pyLDAvis library was utilized to create a HTML file that shows interactive graphs to explore the relationships between topics.

With the saved dictionary and LDA model Gensim file, the code can then assign the topic's ID to each 'comment' and 'caption' in the delegated dataset. Additionally, a topic name for each topic ID is added into the dataset for easier interpretation and analysis.

Through this application of the LDA model, the team is able to further analyze topic evolution over time, and visualize occurring words according to each topic.

### *3.3.2 Topic Distribution*

In this subsection, we analyze the topic distribution across Instagram and TikTok posts and comments to determine the frequency and significance of each topic within each platform. By examining topic distribution, we gain valuable insights into which topics are most represented and discussed by users on each platform.

The topic distribution for the comments and posts of each platform was done in Python, with the help of the pandas and matplotlib packages. The process for analyzing topic distribution begins with calculating the frequency of each topic, and sorting them accordingly. A bar chart is then created to visualize the distribution of topics. This visualization helps in identifying and comparing the prevalence of different topics across the platforms.

### *3.3.3 Topic Evolution Over Time*

Within this part of the analysis, we explore how topics evolve over time to gain insights into everchanging user interests and engagement patterns on Instagram and TikTok. By analyzing the frequency of topic mentions on a weekly basis, trends and emerging interests can be spotted.

To perform this analysis, we employed the use of the pandas and matplotlib packages. First, we performed data preparation by converting timestamps in our datasets to a datetime format. The data were then aggregated by week, and the number of mentions for each topic was counted. Line plots were then created to visualize these weekly counts, highlighting trends and fluctuations in topic popularity over time.

# 4 Results

## 4.1 Sentiment Analysis

Sentiment analysis shows that most reviews in this section are quite positive. Some comments were marked as a specific word that is a negative context, but the analysis was interpreted as expressing love for the product. Therefore, these comments have been labeled “Strong Negative/Positive,” indicating that users truly enjoy using the product. A bar graph representing this data shows that the majority of the comments fall under ‘Neutral,’ ‘Positive,’ and ‘Strongly Positive.’ This analysis therefore helps us understand people’s attitudes towards the brand over the years.

### *4.1.1 Total Number Of Comments*

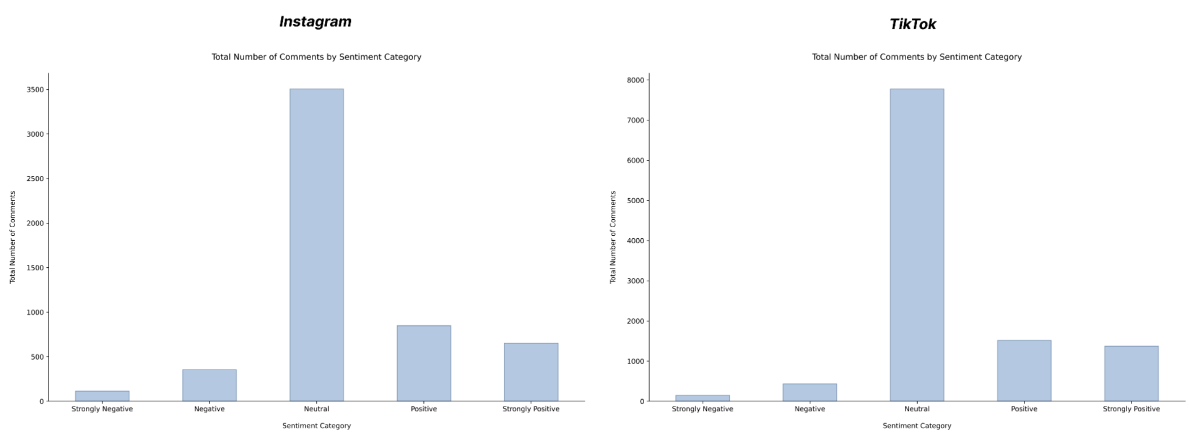



*Fig. 2: Total number of comments by sentiment category*

In Fig. 2, we can observe that most of the comments are categorized under ‘Neutral’ followed by ‘Positive’ and ‘Strongly Positive’ having a slightly smaller portion than ‘Negative’ and ‘Strongly Negative’. ‘Negative’ and ‘Strongly Negative’ are reported to have the smallest distribution portion for both Instagram and TikTok graphs. In the sentiment distribution for Instagram, ‘Neutral’ category contains approximately 3,500 comments while for the TikTok graph, ‘Neutral’ category contains more than that approximately 7,500 comments.

Next, ‘Positive’ category for Instagram contains more than 500 comments whereas for TikTok, ‘Positive’ category contains more than 1,000 comments. Similarly to the ‘Positive’ category for sentiment distribution for Instagram, ‘Strongly Positive’ category contains more than 500 comments, only slightly less comments than the ‘Positive’ category. This is also true for the TikTok dataset. ‘Strongly Positive’ category contains more than 1,000 comments, but slightly less comments than ‘Positive’.

In the ‘Negative’ category, the sentiment distribution for Instagram contains approximately more than 250 comments, while for TikTok, the distribution for ‘Negative’ shows that the count is around 500. For ‘Strongly Negative’ category, sentiment distribution for Instagram shows a slightly lesser portion than ‘Negative’. For TikTok’s graph, it is observed that the ‘Strongly Negative’ category contains less than 500 comments.

### *4.1.2 Total Likes By Sentiment Category*

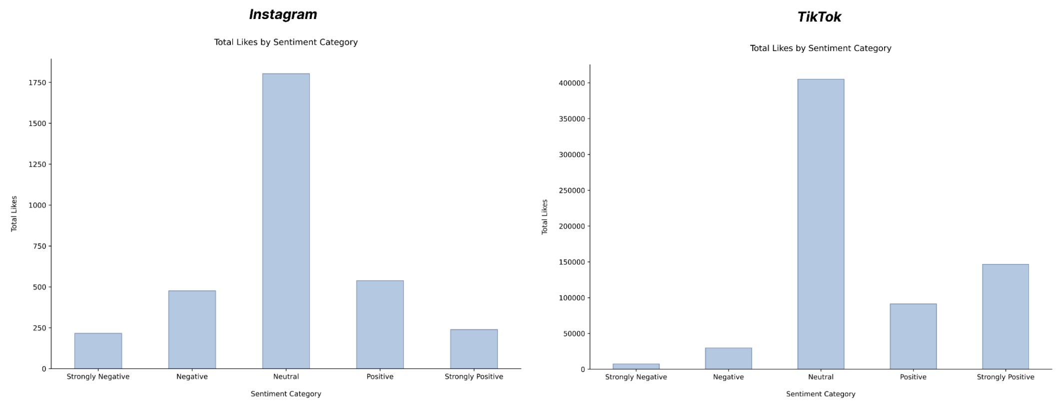



*Fig. 3: Total Likes By Sentiment Category*

Similarly to Fig. 3, the distribution of likes for sentiment categories exhibits a similar trend. In the graph for Instagram and TikTok, we can observe that the 'Neutral' category has the highest like counts among all the other categories.

In Instagram's sentiment distribution for likes, 'Positive' category has the second highest distribution with more than 500 like counts. The third highest distribution would be the 'Negative' category. This category has reached close to 500 like counts. In 'Strongly Positive', like counts account for 250 likes whereas the like counts for 'Strongly Negative' falls below 250.

For TikTok's sentiment distribution, 'Strongly Positive' category has the second highest distribution accounting for almost 150,000 likes. The third highest distribution would be the 'Positive' category with almost 10,000 like counts. 'Negative' category contains less than 50,000 like counts and for 'Strongly Negative', less than 25,000 likes.

### *4.1.3 Average Sentiment Score Over Time*

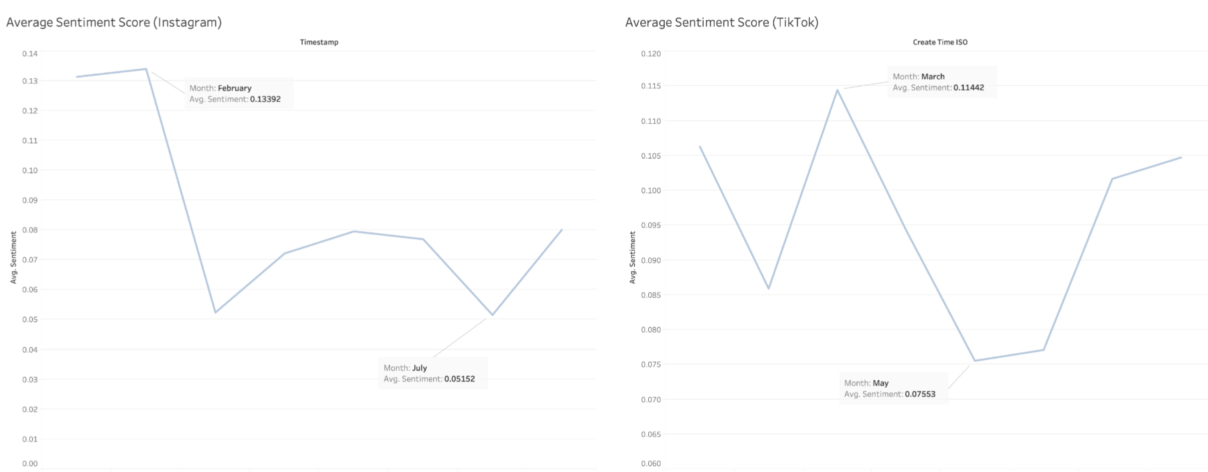



*Fig. 4: Average Sentiment Score Over Time*

In this section, the average monthly sentiment score for Instagram and TikTok was calculated from 1st January to 5th August. According to Fig. 4 for Instagram, the month of January to February had the highest average sentiment score around 0.13 throughout the year. In March, there was a sharp decline in the sentiment score falling to around 0.05. However, there was a gradual increase throughout April to June. In July, there was another decline falling back to the sentiment score around 0.05. In August, it is observed to have an increase again at 0.08.

Next, in TikTok's average monthly sentiment score, it can be observed that February had a decline in sentiment score falling to 0.085. In March, there was a sharp increase in sentiment score reaching up to 0.11. The months of April and May had a sharp decrease as well, having a sentiment score of around 0.075. The sentiment score had been observed to have a gradual increase again during the month of June.

## 4.2 Engagement Analysis

### *4.2.1 Monthly Engagement Analysis*

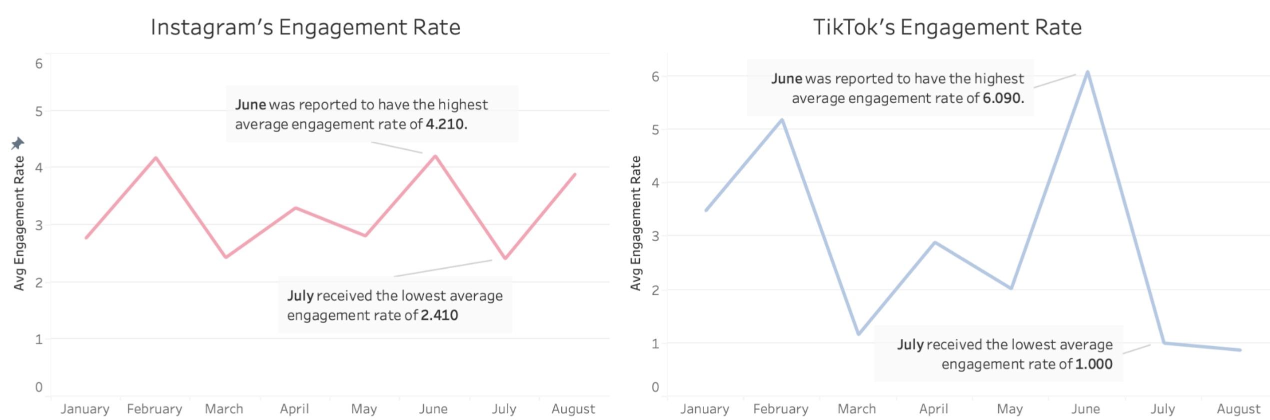



*Fig. 5: Monthly Average Engagement*

As we can see from Fig. 5, the month of June shows the highest average engagement rate. This could be due to the total posts made in June which was also higher than the other months. For the average engagement rate of months, June reached the highest rate of 4.21% in Instagram and 6.09% in TikTok.

For both Instagram and TikTok, July is also observed to have the lowest engagement rate.

### *4.2.2 Daily Engagement Analysis*

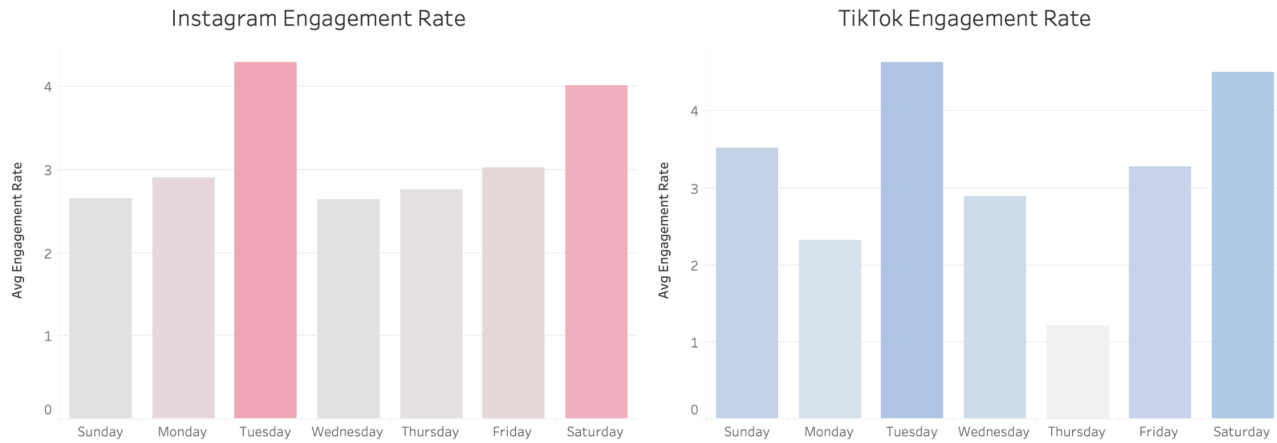



*Fig. 6: Average Engagement for Day of the Week*

According to the results for the day of the week in Fig. 6, the highest average engagement rate can be seen on Tuesday and Saturday for both Instagram and TikTok.

## 4.3 Topic Modeling

### *4.3.1 Topic Distribution*

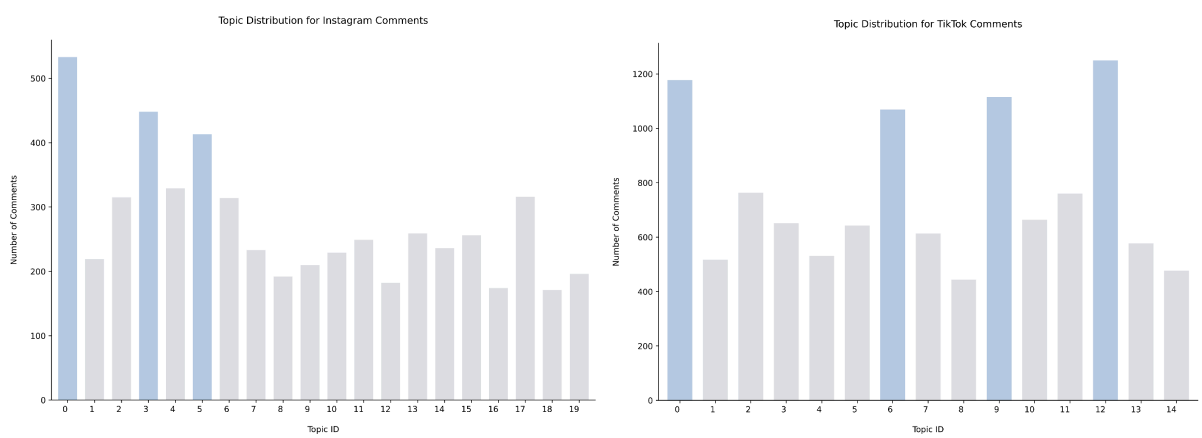



*Fig. 7: Topic distribution for Instagram and TikTok Comments*

In this section, the distribution of topics is presented in bar graphs to show which topic among the social media platform’s comments is most prevalent. In Fig. 7: Topic Distribution of Instagram Comments, it's evident that topic IDs 0, 3, and 5 have the highest comment counts, each exceeding 400. From these topic IDs, the corresponding topic names are:

- **Topic 0:** Petite Lip Tint Restock Request
- **Topic 3:** Rhode Product Availability Request
- **Topic 5:** Product Restock and Expansion Request

On the other hand, Fig. 7: Topic Distribution of TikTok Comments highlights four topics that exceed 1000 comment counts. These topics include 0, 6, 9, and 12.

- **Topic 0:** Product Love and Request
- **Topic 6:** Hailey Bieber and Rhode
- **Topic 9:** Need for Rhode Products and Brand Love
- **Topic 12:** Request to Bring Back Lip Tint

Next, we have the bar graphs showing the distribution of topics for Instagram and TikTok's posts. In the Topic Distribution for Instagram Post, Fig. x reports that there are three topics that are mentioned more than 20 times in Rhode's Instagram post captions. These topics IDs are: 0, 4, and 13. The topic names according to their topic ID are:

- **Topic 0:** Refreshing Pineapple Cleanser
- **Topic 4:** Rhode Summer Lip Product
- **Topic 5:** Rhode Pocket Blush

In Fig. 8: Topic Distribution for TikTok Post, topic 2, 13, and 14 were highlighted as they were mentioned by Rhode's TikTok post captions more than 50 times. These topics include:

- **Topic 2:** Rhode Pop-Up Event & New Products
- **Topic 13:** Lip Tint & Case Waitlist
- **Topic 14:** Lip and Skincare Products

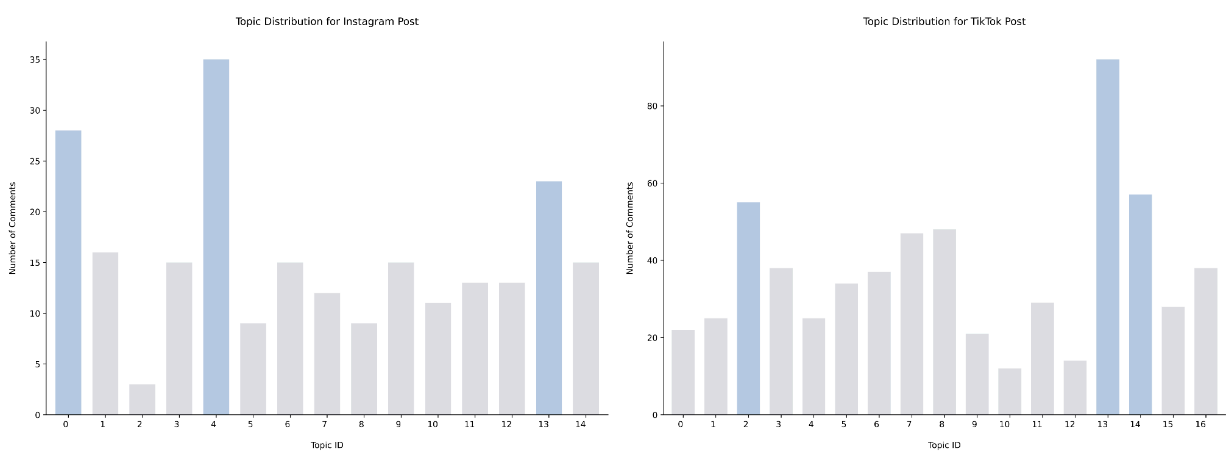



*Fig. 8: Topic distribution for Instagram and TikTok Posts*

***4.3.2 Topic Evolution Over Time***

With the line graph created to track the evolution of each topic across 1st January 2024 to 5th August, we are able to monitor the topic trends and identify at which time they reached their highest number of mentions. To visualize the mentions of each topic, multiple line graphs were generated using the TikTok post dataset. However, with the large amount of line graphs, we have selected prominent topics with at least 10 mentions. These topics include:

- **Topic 7:** Pocket Blush Shades

- **Topic 8:** Pineapple Refresh Cleanser
- **Topic 13:** Lip Tint and Case Waitlist

In Fig. 9, we can observe that topic 7, 8, and 13 have reached at least 10 mentions within the year. From the graph, topic 7 peaked in week 24 with 10 mentions. It can be seen that the number of mentions starts from 2 and slowly increases in week 11. After week 24, its peak period, it shows a small decline of less than 5 mentions. It slowly increases after week 26, however, there are slight fluctuations between week 26 and 31. In topic 8, its peak period can be seen in week 4 with 14 mentions. Its gradual increase can be seen from week 2 to week 6. After week 4, there is a sharp decline in the number of mentions. However, the following weeks showed a constant number of postings. Next, topic 13 can be observed to have a fluctuation in their number of mentions from week 6 to week 30. Its highest peak is seen at week 9 and 25 with 10 mentions.

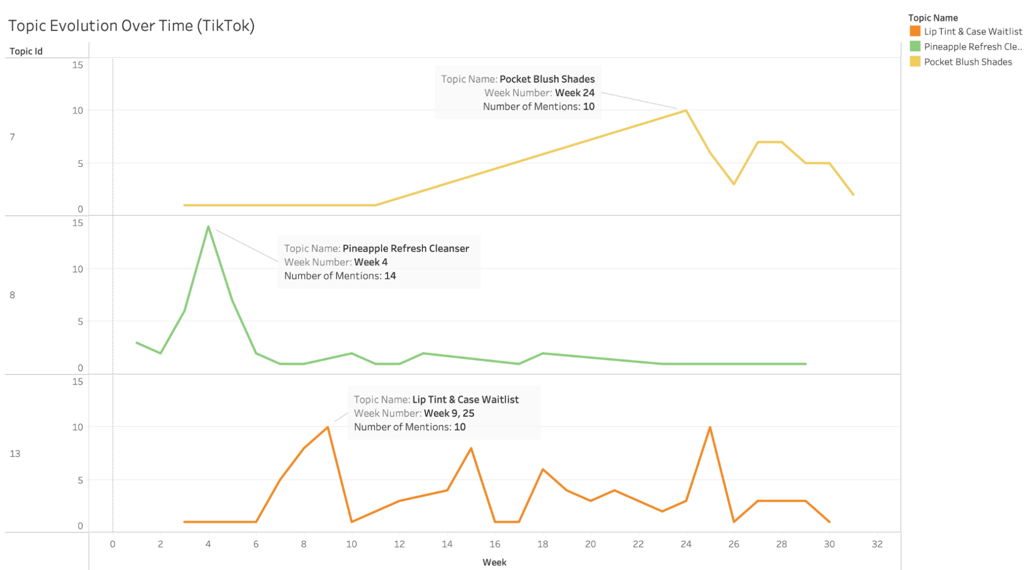



*Fig. 9: Topic evolution over time in TikTok*

On a side note, it is observed that topic 7, 8 and 13 have a low number of mentions during week 26. To investigate this, the peak period in week 26 for each topic was examined. Upon inspection, topic 2 and 4 was found to have a peak period in week 26 with 6 mentions as can be seen in Fig. 29.

- **Topic 2:** Rhode Pop-Up Event & New Products
- **Topic 4:** Rhode Pocket Blush Shades in NYC

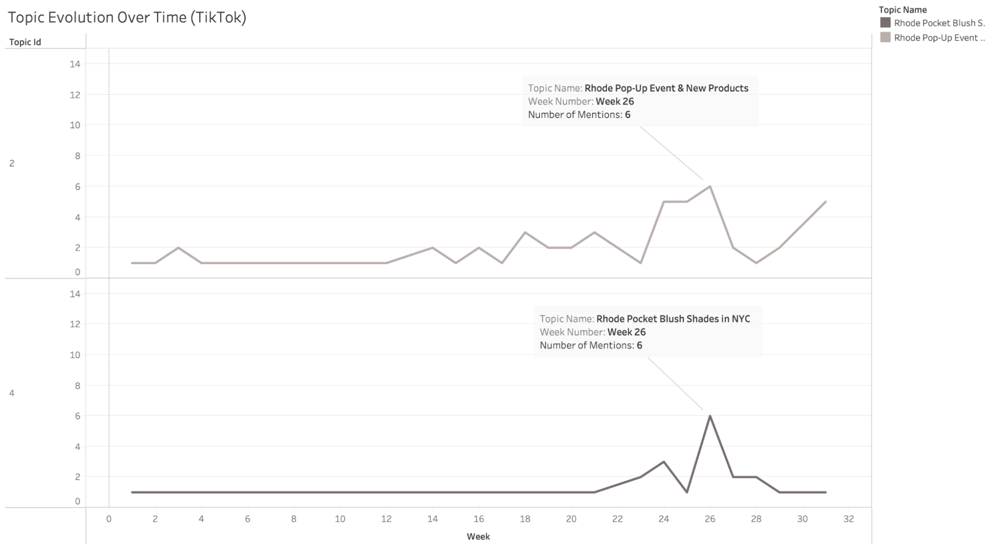



*Fig. 10: Topic evolution over time in TikTok (topic 2 and 4)*

Using the Instagram dataset, multiple line graphs were generated to visualize the number of mentions for each topic. Unlike the graphs in TikTok, most of the topics in Instagram have shown little to none mentions across the year 2024 (refer to Appendix C). Hence, only topics with at least 5 mentions were chosen for analysis. This includes topic 0 and 4. In topic 0, it can be observed that week 4 is the peak period with 7 mentions. After week 4, it can be shown that there is a gradual decline in the number of mentions. During week 13 and 14, it showed an increase with 3 mentions. However, after week 14, there was little to none mentions. In topic 4, peak periods can be seen during week 24 to 25 with 9 mentions. In week 26, there was a sharp decline in mentions. Week 27 showed a gradual increase onwards.

- **Topic 0:** Refreshing Pineapple Cleanser
- **Topic 4:** Rhode SummerLip Product

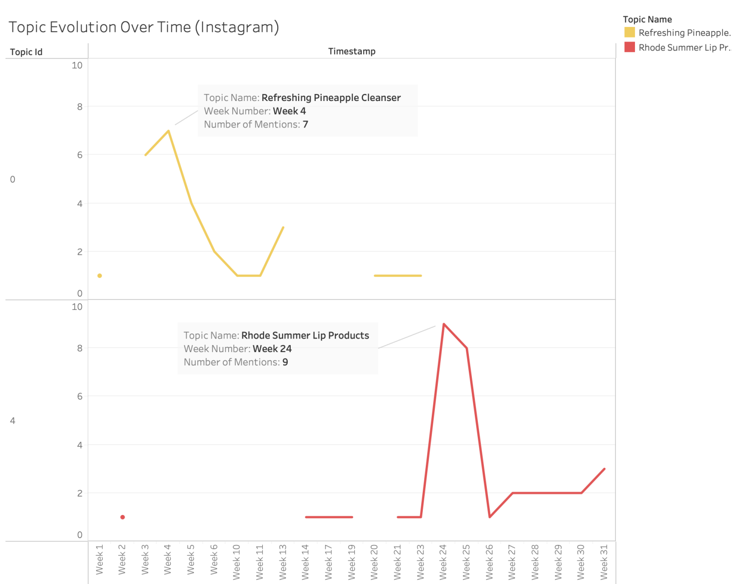



*Fig. 11: Topic evolution over time in Instagram*

# References


[1] S. McLachlan, “Social Media for Business: A Practical guide,” *Social Media Marketing & Management Dashboard*, Jun. 27, 2024. https://blog.hootsuite.com/social-media-for-business/

[2] A. Schmidt, “How an Emphasis on Self-Care is Transforming the Beauty Industry | Clarkston Consulting,” *Clarkston Consulting*, Nov. 27, 2023. https://clarkstonconsulting.com/insights/self-care-transforming-beauty-industry/

[3] Influencer Marketing Hub, “How Social Media is Shaping the Beauty Industry (+5 Social Media Strategies for Beauty Brands),” *Influencer Marketing Hub*, Jun. 24, 2024. https://influencermarketinghub.com/social-media-beauty-industry/

[4] C. Malis, “How celebrity make-up artists dominated the beauty scene: with the rise of Instagram and TikTok, gurus like Charlotte Tilbury, Anastasia Soare and Pat McGrath have launched brands and become icons too,” *South China Morning Post*, Jun. 04, 2023. [Online]. Available: https://www.scmp.com/magazines/style/luxury/article/3222728/how-celebrity-make-artists-dominated-beauty-scene-rise-instagram-and-tiktok-gurus-charlotte-tilbury

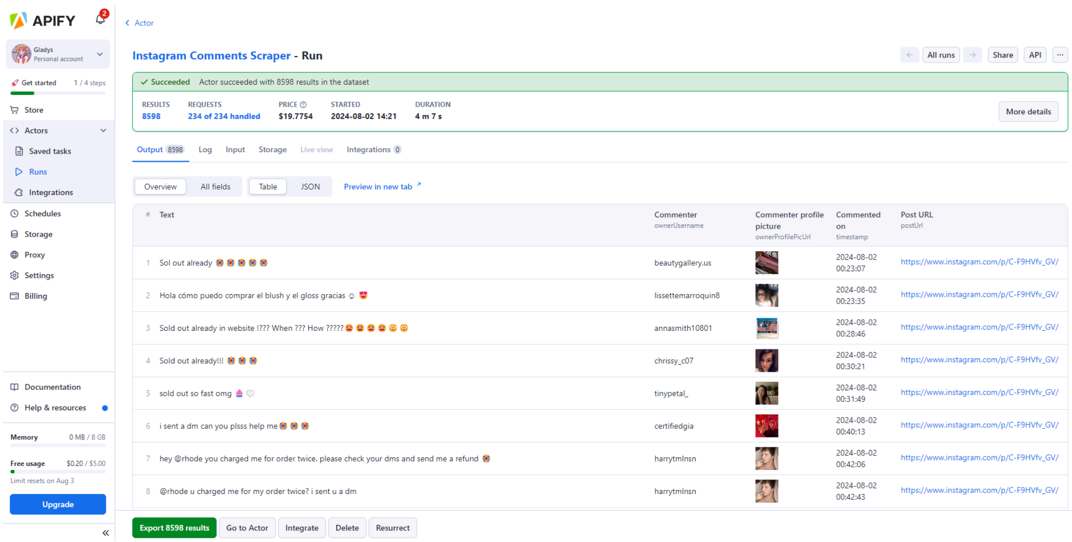
[5] C. Dawson, “The Beauty Industry performance on TikTok,” *ChannelX*, Nov. 20, 2023. [Online]. Available: https://channelx.world/2023/11/the-beauty-industry-performance-on-tiktok/

[6] D. Shah, “The ultimate digital marketing strategy for the beauty industry - 2Stallions,” *2Stallions*, Jun. 05, 2024. https://2stallions.com/blog/the-ultimate-digital-marketing-strategy-for-the-beauty-industry/

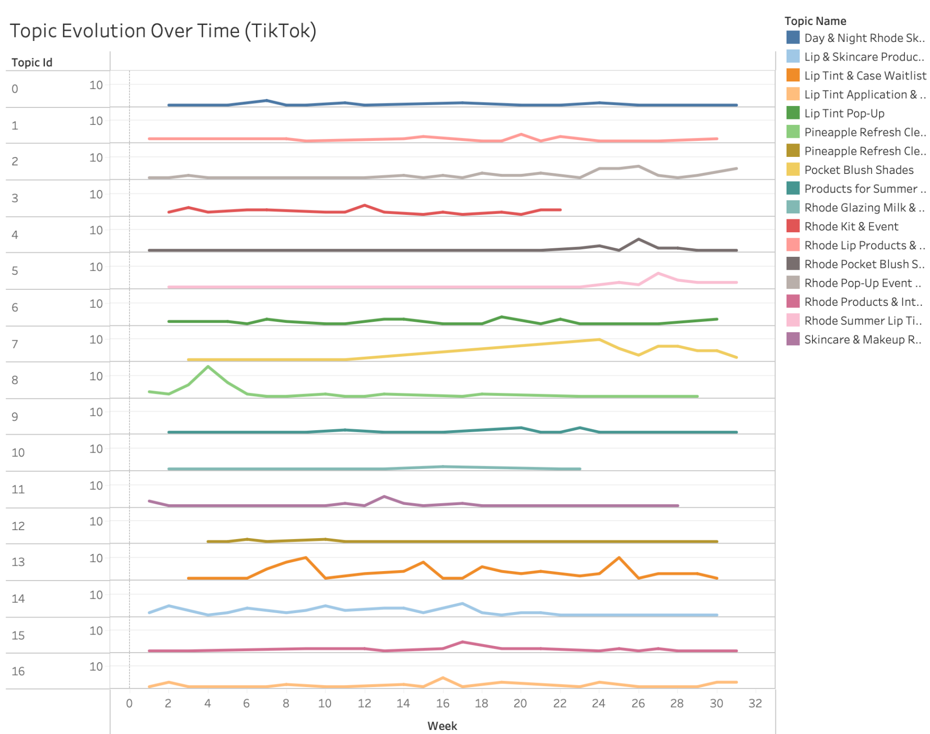
[7] “What is social media Analytics? | IBM.” https://www.ibm.com/topics/social-media-analytics

# Appendix

Appendix A: Data Scraping with APIFY




Appendix B: Topic Evolution Over Time (TikTok Post)



# Appendix C: Topic Evolution Over Time (Instagram Post)