CAPSTONE 2

RALS SUSTAINABLE CARE

RALITSA MIHAYLOVA

Table of contents

Introduction and Background

The proposed startup RALS aims to provide eco-friendly and vegan makeup and hair products. We plan to conduct market research using a data set of teh sephora sales and Google Trends a data to identify the most popular wellness products. Our goal is to create high-quality makeup and haircare that aligns with the growing demand for environmentally conscious and cruelty-free options. We believe that RALS can make a significant contribution to the beauty industry by offering sustainable makeup alternatives.

To conduct market research for RALS, we will use a combination of web scraping and data analysis. We will start by analyzing the sephora products , and after scraping data from Google Trends to identify the most searched keywords related to wellness and beauty products. This data will give us insights into the current trends and demands of the market.
We will also analyze Sephora sales data to determine which makeup products are the most popular among customers. This data will provide us with a better understanding of customer preferences and allow us to identify potential gaps in the market that RALS can fill with its eco-friendly and vegan makeup products.
By combining these two sources of data, we will be able to identify the most promising opportunities for RALS to launch its eco-friendly and vegan makeup products successfully. This approach will allow us to make data-driven decisions that maximize the potential success of the startup.

Objective/Goals of the Project

The beauty industry has seen a significant shift towards eco-friendly and vegan products in recent years. Consumers are becoming more conscious of the impact that their purchases have on the environment and are increasingly opting for sustainable and cruelty-free options. RALS is a startup that has been created with this trend in mind, and we aim to provide high-quality makeup and haircare products that align with these values.
At RALS, we understand the importance of market research in creating successful products. We believe that data analysis is crucial to identifying customer preferences and identifying gaps in the market that we can fill. Our aim is to create makeup and haircare products that not only meet the growing demand for environmentally conscious and cruelty-free options but also deliver on quality and effectiveness.
To conduct our market research, we will be using a combination of web scraping and data analysis techniques. Our initial focus will be on analyzing the Sephora sales data set, which will give us insights into the most popular makeup and haircare products. We will be looking at data such as sales volume, customer reviews, and ratings to identify which products are currently the most sought-after by consumers.
In addition to analyzing Sephora sales data, we will also be using Google Trends data to identify the most searched keywords related to wellness and beauty products. This data will give us an insight into current trends and customer preferences, which will help us to create products that meet the needs of our target audience.

We understand that analyzing and interpreting large amounts of data can be challenging, and that's why we plan to use data visualization techniques to help us better communicate our findings. We will be creating charts, graphs, and other visual aids that will help us to present our data in a clear and concise manner.
Overall, our goal is to create a startup that delivers high-quality, eco-friendly, and vegan makeup and haircare products that meet the needs and desires of modern consumers. We believe that our market research approach will help us to achieve this goal by providing us with the insights and data that we need to make informed decisions about our product development and marketing strategies.

Data Extraction and Data Exploration

To start our market research for RALS, we began by analyzing the Sephora data set, which contains information on various makeup and haircare products sold by Sephora. The data set includes columns such as brand, name, price, number of reviews, number of loves, review score, size, category, reviews to love ratio, returns, and price.
My reasoning to use this dataset is **two** folds: **(1)** the data set was compiled in late November 2022 by scaping Sephora's website, making it a very relevant to the problem at hand; and **(2)** Sephora is well known mass beauty retailer that carrys many products across brands.

In the second part of our research, we will be focusing on web scraping to gather more data on current trends and customer preferences in the beauty industry. We believe that by combining web scraping with our analysis of the Sephora data set, we will be able to gain a more comprehensive understanding of the market and create products that meet the needs and desires of modern consumers.

Here a short description of the columns, we have a total of 45 columns inside the dataset.
If the column is highlighted is one the ‘core’ columns, meaning that is an important source of our data, the non-highlighted ones are categories of a prouct, in case one of the products matches that category we will have a 1(yes), and 0(no) in every other category.
Products can be part of multiple categories, meaning we can have more ones per row inside the dataset.
Note: The name of the columns has been modified from the original data set in order to make it more understandable and clear.

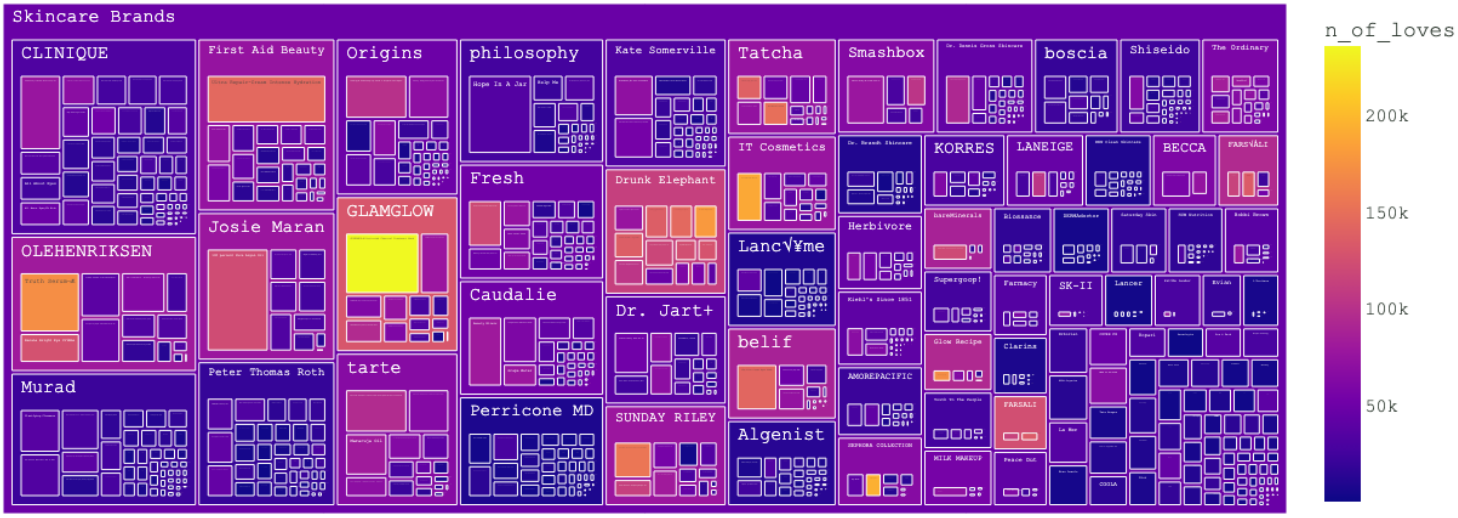
#	Column	Non-Null Count	Dtype
0	brand	1649 non-null	object
1	name	1649 non-null	object
2	price	1649 non-null	float64
3	n_of_reviews	1649 non-null	int64
4	n_of_loves	1649 non-null	int64
5	review_score	1649 non-null	float64
6	size	1427 non-null	float64
7	clean_product	1649 non-null	int64
8	Anti_Aging	1649 non-null	int64
9	BB_&_CC_Cream	1649 non-null	int64

```
10  Bath_&_Shower              1649 non-null   int64
11  Beauty_Supplements         1649 non-null   int64
12  Blemish_&_Acne_Treatments  1649 non-null   int64
13  Blotting_Papers            1649 non-null   int64
14  Body_Lotions_&_Body_Oils   1649 non-null   int64
15  Cellulite_&_Stretch_Marks  1649 non-null   int64
16  Decollete_&_Neck_Creams    1649 non-null   int64
17  Exfoliators                1649 non-null   int64
18  Eye_Creams_&_Treatments    1649 non-null   int64
19  Eye_Masks                  1649 non-null   int64
20  Face_Masks                 1649 non-null   int64
21  Face_Oils                  1649 non-null   int64
22  Face_Primer                1649 non-null   int64
23  Face_Serums                1649 non-null   int64
24  Face_Sunscreen             1649 non-null   int64
25  Wash_&_Cleansers           1649 non-null   int64
26  Facial_Peels               1649 non-null   int64
27  Foundation                 1649 non-null   int64
28  Hair_Oil                   1649 non-null   int64
29  Highlighter                1649 non-null   int64
30  Holistic_Wellness          1649 non-null   int64
31  Mini_Size                  1649 non-null   int64
32  Mists_&_Essences           1649 non-null   int64
33  Moisturizer_&_Treatments   1649 non-null   int64
34  Moisturizers               1649 non-null   int64
35  Night_Creams               1649 non-null   int64
36  Setting_Spray_&_Powder     1649 non-null   int64
37  Sheet_Masks                1649 non-null   int64
38  Skincare                   1649 non-null   int64
39  Tinted_Moisturizer         1649 non-null   int64
40  Toners                     1649 non-null   int64
41  Tools                      1649 non-null   int64
42  Value_&_Gift_Sets          1649 non-null   int64
43  reviews_to_loves_ratio     1649 non-null   float64
44  return_on_reviews          1649 non-null   float64
45  price_per_ounce            1427 non-null   float64
```

The first step in our data analysis process was to upload the Sephora data set into our software program. Once uploaded, we performed a quick exploration of the data set using various commands such as head, tail, and columns to get a sense of what information was available. We also checked for null values in the data set to ensure that our analysis was accurate. After exploring the data set, we decided to remove some columns that we deemed not useful for our analysis, such as the 'Unnamed' column. This helped us to streamline our analysis and focus on the most relevant data points.

One of the key aspects of our analysis was creating a visualization of all the products in the data set. This allowed us to see which products were the most popular based on the number of 'loves' they received. The more loves a product had, the more people loved it, making it an inspiration for RALS. This visualization helped us to identify the most sought-after products in the market and better understand customer preferences.

In this visualization we can observe all the sephora Brands we can find, under each there are all the famous brands bevinded by category. On the right we can see a scale, the scale tells us how many loves (click on the heart of the products on the sephroa website in order to add it to your favorites), with yellow meanin really loved and blue, not much.
The visualization is interactive, by using inside python will show you each product under brands, with the category, the number of loves and number of reviews.



The visualization that we created by plotting the number of 'loves' for each product in the Sephora data set can provide valuable insights for our market research for RALS. By observing all the brands and categories available in Sephora and their corresponding popularity, we can gain a better understanding of the current trends in the beauty industry.
For instance, if we notice that a particular brand or category has a high number of 'loves', we can infer that it is popular among customers and may be a potential area for RALS to explore. Similarly, if we notice that a particular brand or category has a low number of 'loves', we can infer that it may not be as popular among customers and may not be worth investing in for RALS.
Additionally, the color scale that we used in our visualization can help us to quickly identify which products are the most loved by customers (yellow) and which products are not as loved (blue). This can help us to focus our attention on the most popular products and identify potential gaps in the market that RALS can fill with its eco-friendly and vegan makeup and haircare products.

Overall, the visualization of the 'loves' data can provide valuable insights for our market research and help us to make data-driven decisions for RALS. By leveraging this data, we can

create products that align with the current trends and preferences of modern consumers and maximize the potential success of our startup.

Data Analysis

Data analysis is a critical aspect of our market research for RALS. In this phase, we are focused on examining the Sephora dataset to identify trends, patterns, and insights that can help us make data-driven decisions for our eco-friendly and vegan makeup and haircare products.
One of the key areas of analysis that we are focused on is identifying the top brands in the Sephora dataset.

1. ***Most loved brands***

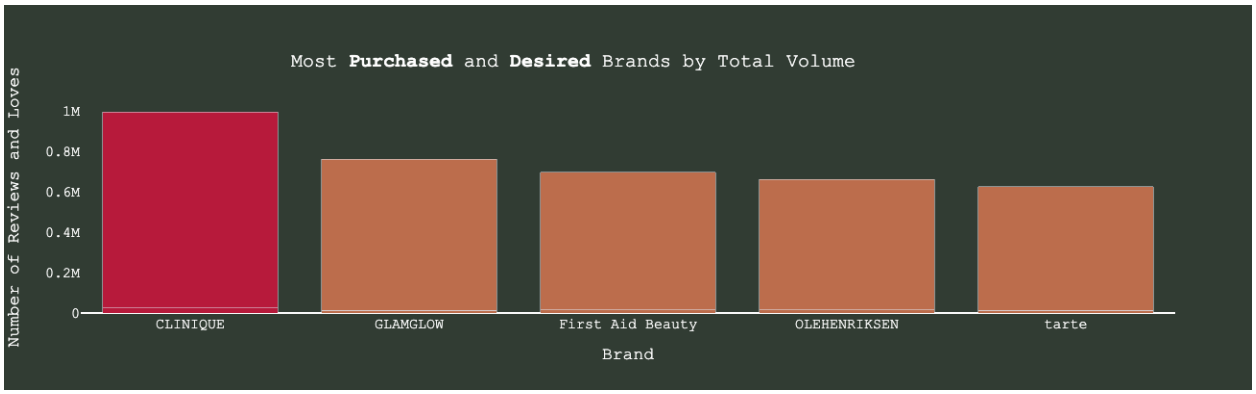
By analyzing the data, we can identify which brands are the most popular among customers and which brands have the highest sales volumes. This information can be used by RALS to make decisions about which brands to partner with, which products to focus on, and how to position our own brand in the market.
For instance, if we identify that a particular brand is consistently among the top performers in terms of sales and popularity, we may choose to partner with that brand to create co-branded products or simply analyze their marketing strategy or the ingredient they use.
This can help us to leverage the existing customer base of the brand and increase the visibility of our own brand.

brand	n_of_reviews	n_of_loves
CLINIQUE	23993	972394
OLEHENRIKSEN	16636	645625
First Aid Beauty	15630	680855
GLAMGLOW	12524	748385
tarte	12319	612764

Another aspect of our data analysis is the creation of an interactive visualization of the top 5 most purchased products. By visualizing the data, we can gain a deeper understanding of which products are the most popular among customers, and how they compare to each other. The visualization can also help us to identify any patterns or trends in the data that may not be immediately apparent from just looking at the raw numbers.
In the interactive visualization, we can see the number of saves and purchases for each of the top 5 most purchased products. This information can be used by RALS to create products that

are similar to those that are currently popular, or to identify gaps in the market that RALS can fill with its eco-friendly and vegan makeup and haircare products.

Overall, data analysis is a crucial part of our market research for RALS. By analyzing the Sephora dataset, we can gain valuable insights into customer preferences and behavior, identify trends and patterns in the market, and make data-driven decisions that maximize the potential success of our startup.



Analyzing the most purchased and desired brands in the Sephora dataset can provide valuable insights into why certain brands are more attractive to consumers than others. Factors such as brand image, marketing strategies, price points, ingredients, and overall popularity can all play a role in this analysis. In the case of Clinique, which emerged as the top brand in our analysis, several key factors contribute to its success.

One of the primary marketing messages of Clinique is that their products are doctor-recommended and safe to use, which helps to build trust with consumers. Additionally, the brand has been established for many years, giving it resources to invest in marketing and product development. While Clinique is not an entirely vegan brand, they do offer many products that are fragrance-free, phthalate-free, and paraben-free, aligning with the growing demand for clean and sustainable beauty options.
Sustainability is also a priority for Clinique, with a goal to make 75% of their packaging recyclable, refillable, reusable, recycled, or recoverable by 2025. They also aim to use Forest Stewardship Council certified materials for all of their secondary packaging. These efforts to reduce waste and promote eco-friendliness align with the values of many consumers who are increasingly seeking out sustainable beauty products.
Understanding the factors that contribute to Clinique's success can help inform RALS' approach to developing and marketing their own eco-friendly and vegan makeup and haircare products. By identifying the key drivers of consumer preference, RALS can position themselves to offer similar benefits while addressing gaps in the market that are not being served by existing brands.

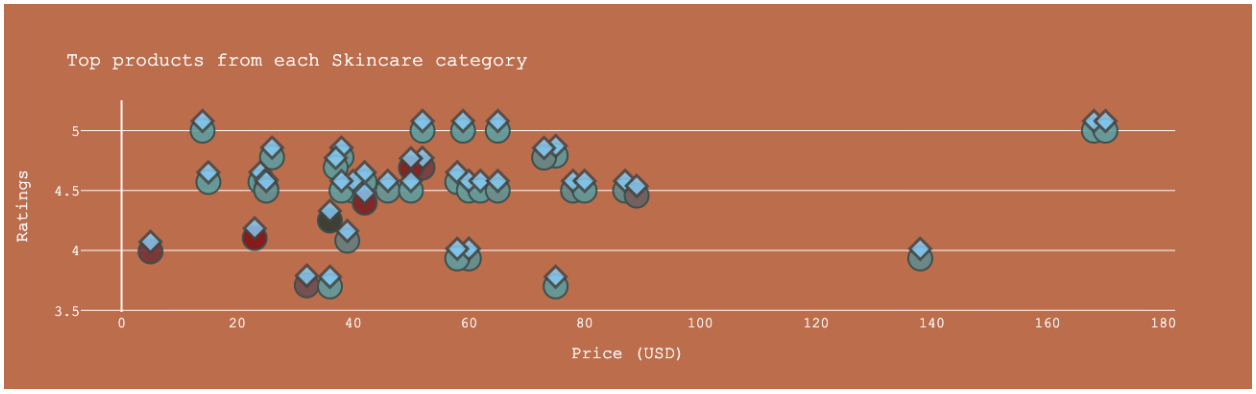
In addition to brand analysis, we also created an interactive visualization of the top 5 most purchased products in the dataset, providing valuable insights into consumer behavior and

preferences. By examining the number of saves and purchases for each product, RALS can gain a better understanding of the features and benefits that are most important to consumers, and use this information to develop their own product offerings.

2. Most loved products

As we continued our data analysis, we took a closer look at the top-rated products in each category, without any consideration for price. Our criteria for including a product in this list was that it had to be either rated 5 stars more than 10 times or the highest rated in its category with more than 10 reviews. Categories without any qualifying products were excluded from the list. In addition to this, we also identified the most loved category based on the number of "loves" received by the products in that category. The top three categories were face serums, beauty supplements, and face oils.

To make this information more accessible, we created a visualization of the top-rated product for each category, represented by a diamond marker. By clicking on each diamond, users can view key information such as brand, rating, category, and price.



	brand	name
827	GLAMGLOW	SUPERMUD¬Æ Activated Charcoal Treatment Mask
13	SEPHORA COLLECTION	Sleeping Mask
2	IT Cosmetics	CC+ Cream with SPF 50+
802	Drunk Elephant	TLC Sukari Babyfacial Mask
807	OLEHENRIKSEN	Truth Serum¬Æ

This information can be very valuable for RALS as it can help them identify the most popular and top-rated products in each category. This can assist them in making informed decisions when selecting which products to feature on their platform. Additionally, understanding which categories are the most loved can help them focus their marketing efforts and attract more customers to their platform.

At the end we also looked for the best products without category constraints, in order to see which are the products taht perform the best overall, in order to see which products RALS should focus the most.
Here the top 5 overall products.
Which are all related to skin care, in partciulary face masks.

Web scaping GOOGLE TRENDS

In the second part of our project, we delved into web scraping using Google Trends. Specifically, we focused on scraping data from Google Trends in New York using the keyword "Sephora". The aim of this exercise was to gain insights into what potential customers are searching for when it comes to Sephora.

1. ***General analysis of the make up products and sephora products***

We then went a step further and performed a more detailed scraping exercise using additional keywords such as "makeup", "skincare", and "Sephora products". This allowed us to gain a deeper understanding of what customers are specifically searching for within these categories. Overall, this exercise in web scraping Google Trends can provide valuable insights for RALS in terms of understanding consumer demand and trends within the beauty industry. By analyzing the search data, RALS can identify areas of opportunity for new product development, identify popular products and brands, and make informed decisions regarding marketing strategies. Furthermore, web scraping Google Trends can aid in forecasting trends and predicting consumer behavior, allowing RALS to stay ahead of the competition and meet the needs of their target market.

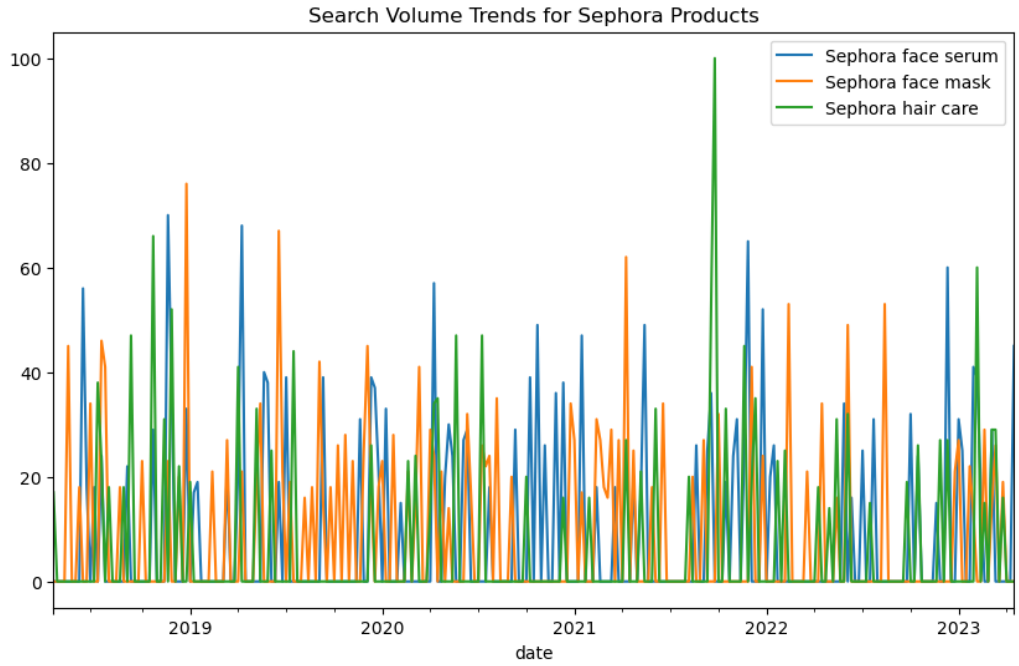
	query	value	keyword
0	clinique	100	makeup
1	clinique makeup	97	makeup
2	eye makeup	73	makeup
3	best makeup	71	makeup
4	makeup artist	41	makeup
..	...	...	...
64	drunk elephant	3	Sephora products
65	amika hair products	2	Sephora products
66	wow hair products	1	Sephora products
67	dermstore	1	Sephora products
68	amika products	1	Sephora products

From this study we can see how one of the most searched words is clinque which we saw previously is considered the ‘top’ brand.
We can also see how hair care is searched amog sephora products (amika and wow are hair care brands).

2. ***Analyzing the most popular products from part 1 of the project***

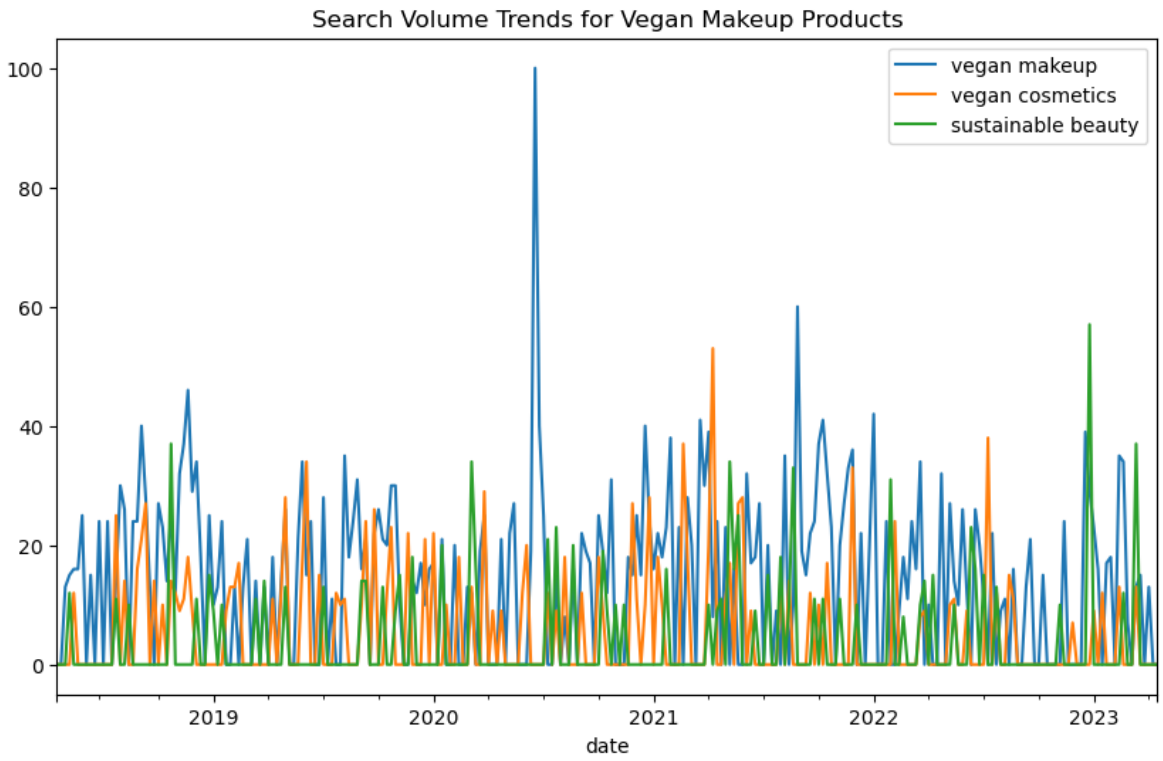
After I decided to use the most popular products from our previous analysis and see how they perform on google trends.
Specifically, we looked at the search volume trends for three key products: face serum, face mask, and hair care.

Using the Python package pytrends, we connected to the Google Trends API and defined the search terms as 'Sephora face serum', 'Sephora face mask', and 'Sephora hair care'. We also set the geographic location to US-NY and the time range to the past 5 years.
The resulting graph shows the search volume trends for these three products over the past 5 years. From the graph, we can see that the search volume for Sephora face serum has been consistently higher than the other two products, with occasional spikes in interest for face masks and hair care. This suggests that face serum may be a more popular product among potential customers in the New York area, and Rals could consider focusing their marketing efforts on this product to attract more customers.



3. Analyzing the performance of vegan and sustainable make up

In the final analysis, we explored the search trends for 'vegan makeup', 'vegan cosmetics',and 'sustainable beauty' in New York over the past five years using Google Trends. The search volume for all three terms peaked in 2020, which could be attributed to the growing awareness of sustainability and ethical consumerism in the beauty industry. This suggests that there is a growing market for sustainable and vegan makeup products, which is in line with RALS's mission as a sustainable makeup company using clean and vegan ingredients.
The analysis of search trends can help RALS identify consumer preferences and make data-driven decisions about their product offerings. By incorporating sustainable and vegan ingredients, RALS can tap into the growing market of conscious consumers and differentiate themselves from competitors. Additionally, by staying up-to-date with search trends, RALS can adjust their marketing strategy and product development to meet changing consumer demands.



Conclusions and recommendations

Based on the analysis performed on the Sephora dataset and Google Trends data, we can make several conclusions and recommendations for RALS.
Based on the analysis of the data set, we can conclude that the most popular product categories among Sephora customers are face serums, beauty supplements, and face oils.

Additionally, we found the highest-rated products in each category, which can provide insights for RALS to develop and promote similar products to target customers.
We also identified the top-rated products within each category, which could be used as a benchmark for RALS to create high-quality products.
Furthermore, the analysis of the customer ratings and reviews can give an idea of the factors that make a product successful in the market, such as its effectiveness, packaging, and price.

As a recommendation, RALS can use the insights from the data analysis to inform its product development, marketing strategies, and customer targeting. For example, they can focus on developing clean and vegan products in the face serum, beauty supplement, and face oil categories, as these are popular among Sephora customers. Additionally, RALS can emphasize the effectiveness, packaging, and value for money of its products in its marketing messages, based on the factors that customers appreciate in the highest-rated products.

Secondly, the analysis of Google Trends data showed that the search volume for vegan makeup, vegan cosmetics, and sustainable beauty peaked in 2020, indicating a growing trend

towards sustainability and ethical consumerism in the beauty industry. Therefore, it would be wise for RALS to focus on creating sustainable and vegan beauty products to cater to this growing demand.

Lastly, based on the analysis, we recommend that RALS should focus on creating sustainable and vegan beauty products, especially in the skincare and makeup categories.
RALS could also use the top-rated products identified in the Sephora dataset as a benchmark for creating high-quality products. Additionally, RALS could leverage the growing trend towards sustainability and ethical consumerism by marketing their products as sustainable and vegan-friendly.

In conclusion, the analysis of the Sephora dataset and Google Trends data can provide valuable insights for RALS to create high-quality, sustainable, and vegan beauty products that cater to the growing demand for ethical consumerism in the beauty industry.

Bibliography

https://www.sephora.com/
https://www.forbes.com/sites/kristinlarson/2021/04/30/vegan-beauty--brands-like-typology-the-ordinary-luneaster-lead-the-way/?sh=29adc8152e0a
https://www.mordorintelligence.com/industry-reports/vegan-cosmetics-market
https://www.kaggle.com/datasets/raghadalharbi/all-products-available-on-sephora-website