***Data Visualisation Project (Jan-2023)***

Exploratory data analysis and visualization of face masks data

***By Group 10***

Imaad Ansari - 21f1004808

Harshitha Srikanth-21f1004861

Yatin Chug -21f1007066

Arram Nithish - 21f1003498

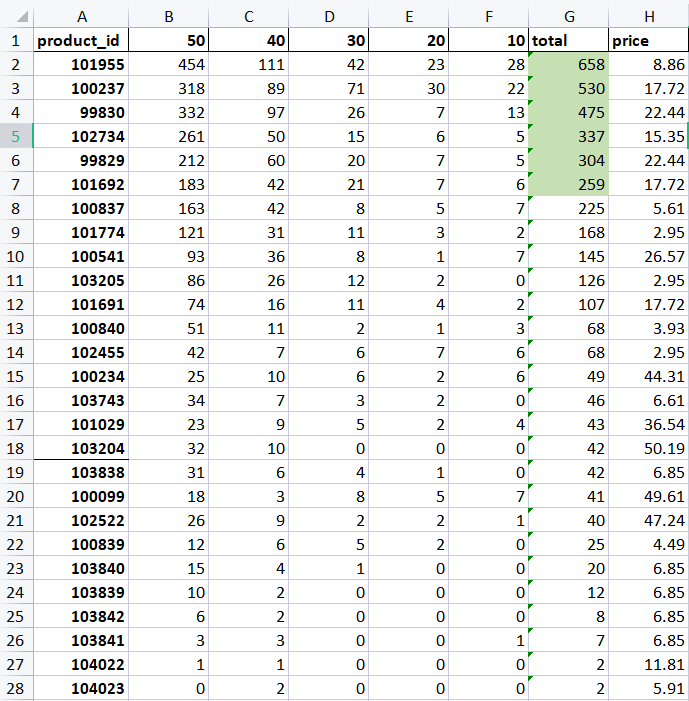
[Presentation link](https://www.canva.com/design/DAFgh2dJqVE/BoCZZ6cq3G2orhurzDaT-g/edit?utm_content=DAFgh2dJqVE&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton) : <https://www.canva.com/design/DAFgh2dJqVE/BoCZZ6cq3G2orhurzDaT-g/edit?utm_content=DAFgh2dJqVE&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton>

***Index***

| **#** | **Project Questions** | **Pg.No** |
| --- | --- | --- |
| 1 | What are the most popular face masks in the store and what are their key  features (e.g. price, material, design)? | 3 |
| 2 | What do consumers like about the most popular face masks and why? | 4 |
| 3 | What are the different profiles of consumers who buy face masks and how can we segment them based on their demographics, behaviors, and preferences? | 5 |
| 4 | How can additional data sources be used to gain more insights into the type of consumers buying different face masks? | 8 |
| 5 | What are the key findings and implications of this analysis for the client's marketing strategy and research and development department? | 9 |
| 6 | What are the potential constraints and improvements to the analysis and how can we address them? | 12 |
| 7 | Identify and suggest a few extra features which could have helped you to gain a few more valuable insights into the different types of consumers who purchase these face masks. | 12 |
|  | Code | 14 |

***1. What are the most popular face masks in the store and what are their key features (e.g. price, material, design)?***

With the data provided to us, we can find the most popular masks by counting the number of reviews and number of positive ratings provided by customers.



In this table, we can see the top Six most popular face masks with their product id.

**1.SunJoy, KN95, Professional Protective Disposable Face Mask, 10 Pack**

**2.Kitsch, 100% Cotton Reusable Face Masks, Leopard, 3 Pack**

**3.Kosette, Nano Reusable Face Protection Mask, Medium, 1 Mask**

**4.Zidian, Disposable Protective Mask, 50 Pack**

**5.Kosette, Nano Reusable Face Protection Mask, Large, 1 Mask**

**6.Kitsch, 100% Cotton Reusable Face Masks, Neutral, 3 Pack**

We can observe that the price of these most popular masks are so genuine that they can be bought by most people.

* The maximum price the customer can pay is 22.44AUD and the masks are reusable.
* Some of the masks(Zidian and SunJoy) are disposable and come with packs of 10 or 50.they are quite cheap(0.5 -0.8 AUD per mask) for people to use in daily life.
* Kosette masks are 100% polyester and washable easily and with adjustable strap.
* Kitsch masks are very soft and 100% cotton and come with packs of 3 with different designs.
* Zidian and SunJoy masks are disposable and made of non-woven fabrics.

***2. What do consumers like about the most popular face masks and why?***

**

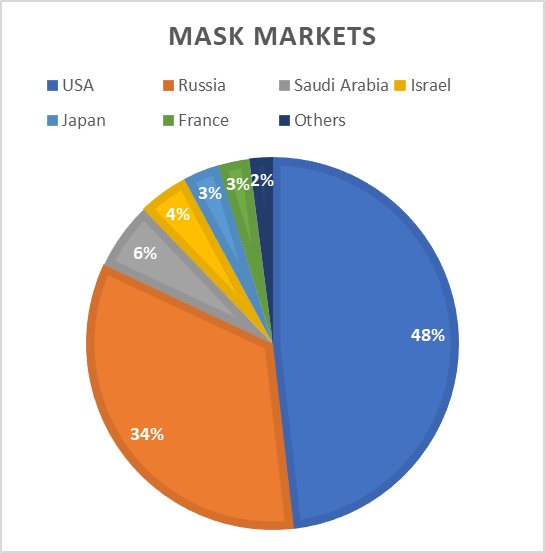
Things that customers like about most popular masks are:

* **COMFORTABLE**: more than 20% customers in the comments talk about comfortability and easy to breathe masks. As this is obvious that everyone wants a mask which is more comfortable and can be worn easily for a long time without irritation.
* **PRICE**: price is another important factor with which these masks are most popular. Disposable masks are cheaper than local pharmacies. Zidian masks come with a pack of 50 that per mask price is quite less.
* **WASHABLE**: around 10% of customers like the material that after many washes the mask color and the elastic remains the same.
* **PROTECTABLE**: some customers just want to protect themselves from COVID and want a mask that covers their face. They like masks with more layers. Kosette masks come with nano filters which protect them more.
* **SIZE**: size is the most common issue everyone gets while ordering products online.Smaller size masks are uncomfortable as it causes pain around ear and nose.Kosette masks have adjustable straps which are liked by customers.Kitsch masks are cotton masks which come with one size which fits all and these masks are stretchable.

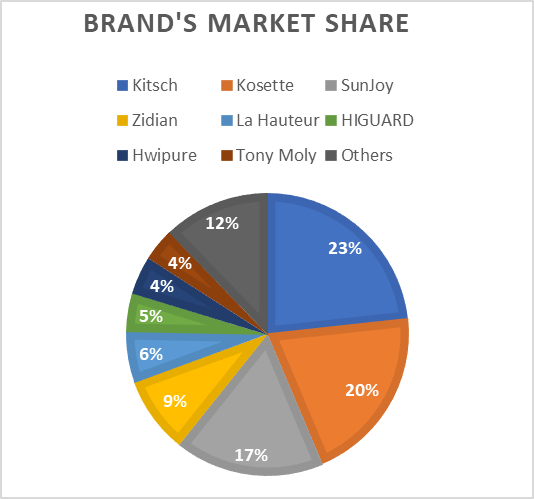
***3. What are the different profiles of consumers who buy face masks and how can we segment them based on their demographics, behaviors, and preferences?***

· To get the data about the region of the consumer the language of the review is used to determine the region of the buyer.

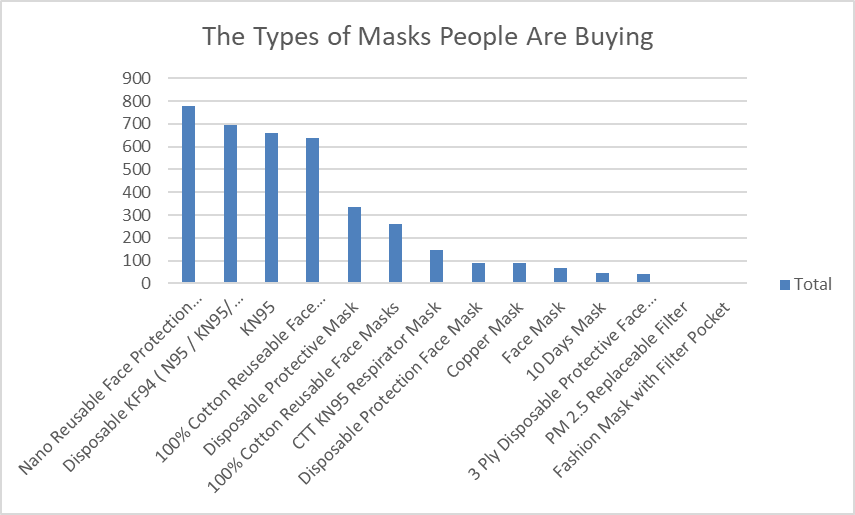
· To get data about the brands the products.csv file is used, and the data is mapped to reviews.csv files using the productid.



***Demographics:*** The data shows that the majority of sales were made in the USA and Russia, suggesting that these two countries may have the highest demand for face masks. It is possible that this is due to a combination of factors, such as population size, government regulations, and public perception of the importance of wearing masks.



**Brand preferences:** The data shows that Kitsch, Kosette, and SunJoy were the top-selling brands, indicating that consumers may prefer these brands for their quality, price, or marketing strategies. Kitsch has more sales overall, however Kosette and SunJoy have a bigger market in Russia than Kitsch.



***Type of mask preferences:***The data shows that the most popular types of masks were the Nano Reusable Face Protection Mask, Disposable KF94 (N95/KN95/FFP2) Mask, and KN95, suggesting that consumers may prefer masks that offer high levels of protection and are convenient to use. However, it is important to note that different types of masks may be preferred for different situations or personal preferences.

***4. How can additional data sources be used to gain more insights into the type of consumers buying different face masks?***

*To gain more insights into the type of consumers buying different face masks, you can use additional data sources such as demographic data, social media data, and website analytics. Here are some ways to use these data sources:*

***Demographic Data:*** You can use demographic data such as age, gender, income, and location to understand the preferences of different consumer segments. You can analyze the reviews of each face mask by different demographic groups and see if there are any patterns or differences in their feedback. For example, you might find that younger consumers prefer face masks with colorful designs, while older consumers prefer face masks that are more plain and simple.

***Social Media Data:***You can use social media data to gain insights into the conversations and sentiment around different face masks. You can monitor social media platforms such as Twitter, Facebook, and Instagram to see what people are saying about different face masks. You can use sentiment analysis tools to understand the overall sentiment around each face mask, and see if there are any trends or patterns in the feedback.

***Website Analytics:*** You can use website analytics to understand how consumers are interacting with your face mask product pages. You can analyze data such as page views, bounce rates, and conversion rates to see how different types of consumers are engaging with your products. For example, you might find that consumers in certain locations are more likely to view your face mask products, or that consumers who visit your site on mobile devices are more likely to make a purchase.

***Sales Data:***You can use sales data to understand which face masks are selling the most and which consumer segments are buying them. You can analyze sales data by region, channel, and other parameters to see if there are any patterns or trends in consumer behavior. This can help you identify which types of face masks are most popular among different consumer segments and adjust your product offerings accordingly.

***Competitor Data:*** You can use competitor data to understand how your face masks are performing relative to those of your competitors. You can analyze pricing, product features, customer feedback, and other parameters to see how your face masks compare to those of your competitors. This can help you identify opportunities to differentiate your products and improve your marketing strategy.

***Industry Reports:***You can use industry reports and research to gain insights into the broader trends and dynamics of the face mask market. You can analyze reports on consumer behavior, market size, and other industry metrics to understand where the market is headed and what challenges and opportunities lie ahead.

By combining the insights from these additional data sources with the reviews of different face masks, you can gain a deeper understanding of the preferences and behaviors of different consumer segments. This can help you make more informed decisions about your product offerings, marketing strategies, and customer engagement tactics.

***5. What are the key findings and implications of this analysis for the client's marketing strategy and research and development department?***

· For finding and implications for the analysis department, the reviews from the dataset have been analyzed.

· Python’s TextBlob library is used to determine the sentiment of each review.

· The stop words are removed from each review so that we can the keywords from each phrase

***What People Like?***

Most of the reviews that are analyzed turned out to be positive (90.6%) and some of qualities that people like the most are:

**· Comfortable:** The term "comfortable" appears frequently in positive reviews, indicating that customers find the masks comfortable to wear for extended periods. This could be because the masks are made from soft, lightweight materials that are gentle on the skin and do not cause irritation.

**· Quality:** The term "quality" suggests that customers value masks that are well-made and durable. Positive reviews frequently mention the quality of the materials, stitching, and overall construction of the masks.

**· Fit:** The term "fit" indicates that customers appreciate masks that fit well and provide a secure seal around the edges. Positive reviews frequently mention the importance of a good fit for both comfort and effectiveness.

**· Breathable:** The term "breathable" suggests that customers prioritize masks that are easy to breathe in and do not cause discomfort or restriction. Positive reviews frequently mention the importance of ventilation and air flow to prevent overheating and reduce fogging of glasses.

**· Price:** The term "price" suggests that customers are interested in finding masks that are affordable and provide good value for their money. Positive reviews frequently mention masks that are priced competitively and offer a high level of performance and comfort.

***What people Dislike? – RnD on the weak points***

Although most reviews about the masks are positive, the are some negative reviews (9.4%). These issues should be addressed by research and development department. Some of the major issues people have with masks are:

**· Small:** The term "small" seems to be a major concern for customers in their negative reviews about face masks. This could indicate that the masks may not be available in a range of sizes, or the sizes available may not fit the customer's face comfortably, leading to discomfort while wearing the mask.

**· Breathe**: The term "breathe" indicates that the customers found it difficult to breathe while wearing the masks. This could be because the masks are too thick or poorly ventilated.

· **Little:** The term "little" could suggest that the masks do not provide enough coverage, leading to concerns about their effectiveness. This could be because the masks are not designed to cover the entire face or have loose fits around the edges.

**· Size:** The term "size" could indicate that the customers had difficulty finding masks that fit their face size and shape. This could suggest that there is a lack of variation in the sizes available or that the masks are not designed to fit a diverse range of face shapes.

**· Thick:** The term "thick" suggests that the masks may be made from a material that is too heavy or dense, making it difficult to breathe while wearing them. Alternatively, it could suggest that the masks do not have sufficient ventilation, leading to discomfort and difficulty breathing.

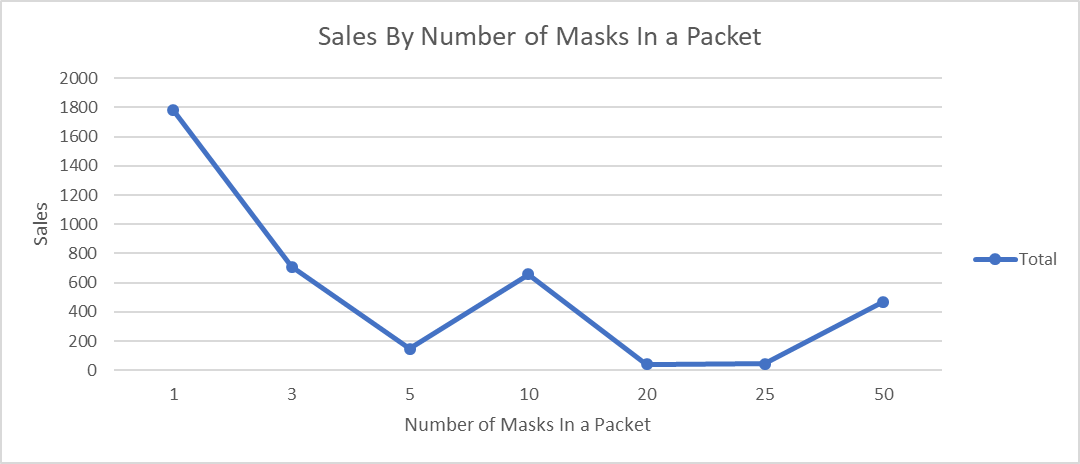
*Based on the concerns raised in the keyword analysis, the research and development team can consider the following suggestions to improve upon the issues:*

·  **Size:** To address the issue of size, the research and development team can consider offering a wider range of sizes to accommodate different face shapes and sizes. The team could also explore incorporating adjustable features such as straps or nose wires to allow for a more customized fit.

·  **Breathe and Thick:** To address the issues related to breathability and thickness, the research and development team could explore using lighter-weight materials that still provide the necessary level of protection. They could also consider incorporating ventilation features such as air vents or mesh panels to improve air flow and reduce discomfort.

·  **Little:** To address concerns about coverage, the research and development team could consider designing masks with more comprehensive coverage that effectively covers the nose, mouth, and chin. They could also explore the use of materials that are more pliable and conform better to the contours of the face to improve the seal around the edges.

*Pack of how many masks people are buying?*

**

The data on the number of masks in a packet indicates that the most popular pack sizes are those containing a single mask or a larger quantity of masks, such as 10 or 50. Packs containing 3, 5, 20 and 25 masks are less popular, with relatively low sales numbers. This information can be useful for the marketing and research and development departments in several ways:

·  **Packaging:** The data on the number of masks in a packet can help inform packaging decisions. For example, since packs containing a single mask are popular, the marketing department could consider designing attractive single-mask packaging that appeals to customers who want to buy just one mask at a time. Alternatively, the company could focus on offering bulk packs of masks containing 50 or more masks for customers who want to buy in bulk.

· **Pricing:** The data on pack sizes can also inform pricing decisions. For example, since packs containing a single mask are popular, the company could consider pricing single masks at a slightly higher price point than masks sold in larger packs. This could help increase revenue and profit margins while still offering customers a variety of purchasing options.

· **Product Development:** The data on pack sizes can inform product development decisions. For example, if the sales numbers for packs containing 3, 5, and 25 masks are low, the research and development team could consider discontinuing these pack sizes or exploring ways to make them more appealing to customers. Alternatively, the team could focus on developing and promoting pack sizes that are more popular, such as packs containing 1, 10, or 50 masks.

***6. What are the potential constraints and improvements to the analysis and how can we address them?***

* The analysis that has been done on the given data is appropriate. However there were some assumptions, the most important one is that we have judged the most popular mask by just the number of reviews. As some of the masks were introduced late to the market. It can also be the reason that those masks have less reviews.
* The limitations we have observed are that we are given a lack of information to judge the popular masks, as the reviews of the products from only 1 website.These things are necessary to gain more meaningful information and insights for products.
* Biased data: Sentiment analysis models are only as good as the data they are trained on, and if the training data is biased, the analysis results will also be biased. To address this issue, it is important to use a diverse and representative training dataset and to regularly evaluate and update the model to ensure that it remains accurate and unbiased.

***7. Identify and suggest a few extra features which could have helped you to gain a few more valuable insights into the different types of consumers who purchase these face masks****.*

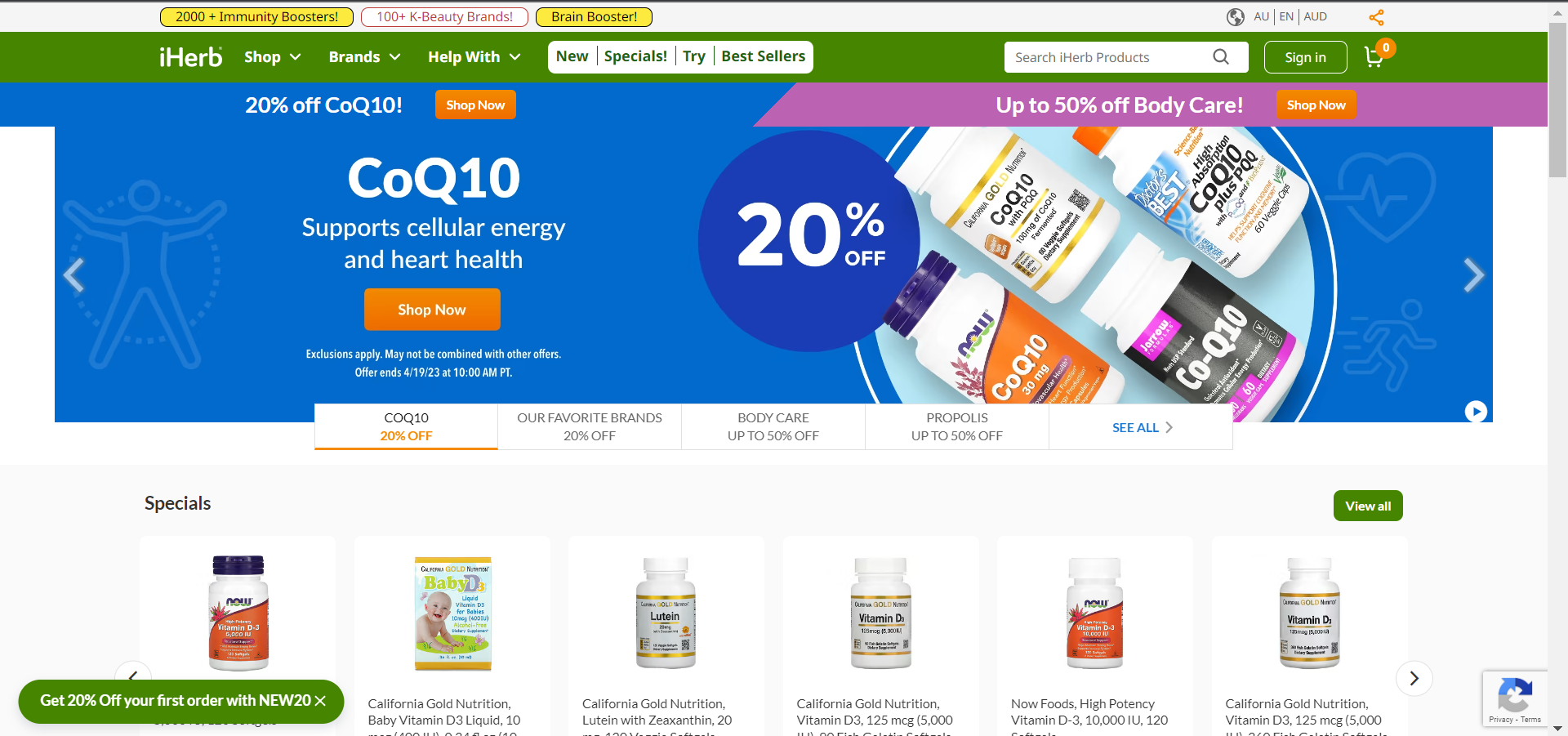
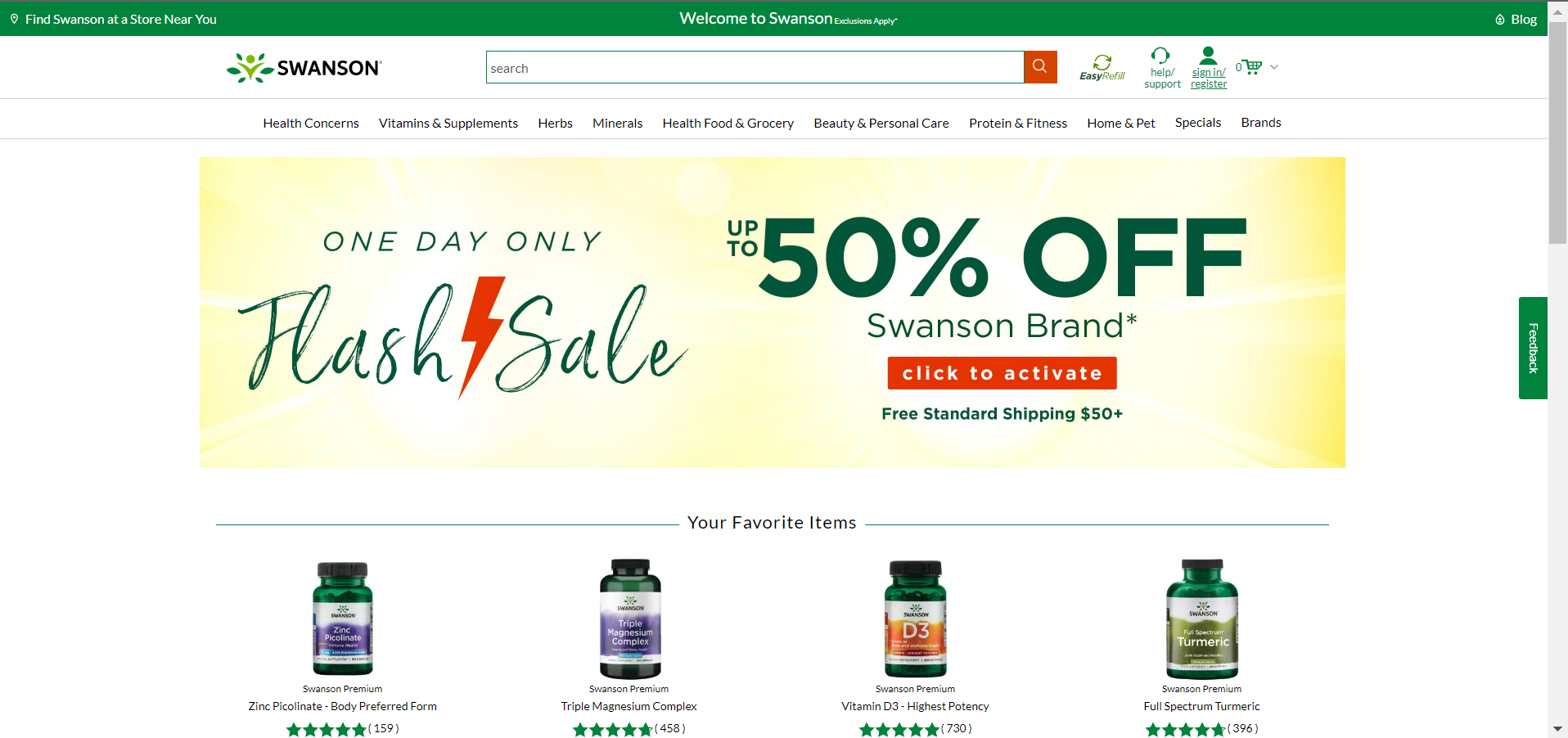
* Valuable insights into customer preferences and opinions can be obtained by collecting demographic information, which includes age, gender, and location, among other factors. This can also aid in better understanding the target audience.
* Information about a customer's purchase history can offer insights into their buying patterns, preferences, and previous experiences with similar products.
* Customer loyalty and level of engagement with the brand can be determined by analyzing their order size and frequency.
* Gaining knowledge about a customer's sentiment towards a product, brand, or customer service can provide insights into areas that require improvement or areas where the brand excels.
* Customer engagement, such as social media interactions or customer service inquiries, can provide valuable insights into customer satisfaction and loyalty.
* A crucial metric in determining the success of a product in the market is the repurchase rate, which indicates how often a customer repurchases a specific product.
* Instead of an overall product rating, specific ratings for major product attributes such as color and fit can be collected to analyze the product's strengths and weaknesses.

Analyzing the application/website and suggestions for the same:

* Better UI/UX designs and color schemes to attract the customer
* Personalized recommendations based on their purchase history and preferences
* Interactive product images that allow customers to see products from different angles or in different settings
* Virtual try-on tools for products
* Better social media presence and advertisements
* Loyalty program or rewards system for repeat customers

Some design improvements for the website that can help provide a better customer experience:

A quick comparison of the design of iHerb.com and a very similar swansonsvitamins.com website design

Inferences of the comparison:

* Simplify the Navigation: iHerb's website has a lot of products, and the navigation can be overwhelming for some users. To simplify it, the designers could consider grouping similar items and creating subcategories that are easier to browse.
* Use of High-quality Imagery: Using high-quality, consistent product imagery can enhance the shopping experience and help customers make informed purchasing decisions. Larger images and more detailed zoom features could be implemented to improve the visual appeal of the site.
* Improve the Color Scheme: While iHerb's current color scheme of white and green works well, it could benefit from additional colors or shades to provide more visual interest. For example, incorporating calming or earthy tones may be suitable for a health and wellness brand.
* Optimize for Mobile: With an increasing number of users browsing on their smartphones, it's essential to ensure that the website is optimized for mobile devices. A mobile-first approach to design would consider how users interact with the site on smaller screens and make necessary adjustments to enhance their experience.
* Use Consistent Typography: The typography on iHerb's website could be improved by using a consistent font family and size across all pages. This would make it easier to read and improve the overall visual appeal of the site.

CODE:

Code used to extract the TSV file as an excel spreadsheet:

import pandas as pd

# Read the TSV file as a dataframe

df = pd.read\_csv('reviews.tsv', delimiter='\t', quotechar='"')

# Save the dataframe as an Excel file

df.to\_excel('new\_reviews.xlsx', index=False)

[Code For Q3 & Q5](https://colab.research.google.com/drive/1wF-KwxbUCxaDkMyWN_dYzcGfk_-AZaqI?usp=sharing)

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

from nltk.corpus import stopwords

from textblob import TextBlob

df=pd.read\_excel("new.xlsx")

prd\_df=pd.read\_excel("products.xlsx")

prd\_df.set\_index('product\_id')

#Making a common column of translation and english reviews.

df["all\_reviews"]=df["translation.reviewText"]

for i in range(len(df)):

if type(df["all\_reviews"][i])==float:

df["all\_reviews"][i]=df["reviewText"][i]

#Data Cleaning

df['word\_list']=df['all\_reviews']

df['word\_list']=df['word\_list'].apply(lambda x: [item for item in x.replace(".","").replace("!","").replace(",","").strip().lower().split(" ") if item not in stop])

#Sentiment Analysis

df['polarity\_score']=df['all\_reviews']

df['polarity\_score']=df['polarity\_score'].apply(lambda x: TextBlob(x).sentiment.polarity)

#Adding region based on the language used in the review.

lang={'ar-SA': 'Saudi Arabia',

'de-DE':'Germany',

'en-US': 'USA',

'es-MX':'Mexico',

'fr-FR':'France',

'he-IL':'Israel',

'ja-JP':'Japan',

'ko-KR':'South Korea',

'pt-BR':'Brazil',

'ru-RU':'Russia',

'zh-CN':'China',

'zh-TW':'Taiwan'}

df['region']=df['languageCode'].apply(lambda x: lang[x])

df['word\_list']=df['all\_reviews']

df['word\_list']=df['word\_list'].apply(lambda x: [item for item in x.replace(".","").replace("!","").replace(",","").strip().lower().split(" ") if item not in stop])

#Partiotning Good and Bad reviews

bad\_rev=df[df['polarity\_score']<0]

good\_rev=df[df['polarity\_score']>=0]

d={}

for i in good\_rev['word\_list']:

count\_words(i)

g\_d=d

d={}

for i in bad\_rev['word\_list']:

count\_words(i)

b\_d=d

print(sorted(g\_d.items(), key=lambda x:x[1],reverse=True))

print(sorted(b\_d.items(), key=lambda x:x[1],reverse=True))

#Count of good and bad reviews

print('Good Reviews',len(df[df['polarity\_score']>0]))

print('Bad Reviews',len(df[df['polarity\_score']<0]))

#WordCloud

#Making csv files for good reviews and bad reviews

gw=pd.DataFrame()

gw['word']=''

gw['count']=''

for i in gr:

gw.loc[len(gw.index)] = [i[0],i[1]]

bw=pd.DataFrame()

bw['word']=''

bw['count']=''

for i in br:

bw.loc[len(bw.index)] = [i[0],i[1]]

gw.to\_csv("good\_words.csv")

bw.to\_csv("bad\_words.csv")

#Making text files for words used in good reviews and bad reviews

def find\_reviews(word):

for i in range(len(df)):

if word in df['word\_list'][i]:

print(df['all\_reviews'][i])

print()

find\_reviews('small')

f = open("good\_words.txt", "w")

gw=pd.read\_csv('good\_words.csv')

for i in range(len(gw)):

for j in range(gw['count'][i]):

f.write(str(gw['word'][i])+" ")

f = open("bad\_words.txt", "w")

bw=pd.read\_csv('bad\_words.csv')

for i in range(len(bw)):

for j in range(bw['count'][i]):

f.write(str(bw['word'][i])+" ")