Data Visualization Design Project

January '23 Term

Exploratory Data Analysis and Visualizing iHerb Face Masks Data

A Project Report by,

Group 6

Deeksha Chutani	21F1002583
Param Chordiya	21F1003953
Shagun Dwivedi	21F1001731
Shaifali Vashishtha	21F1003257

Objective:

The objective of this project was to analyze mask-related datasets to gain insights into the competition and help improve the marketing strategy and R&D.

Approach:

The project was categorized into four parts: selecting tools based on our needs, determining the extent of preprocessing required, selecting types of visualizations, and interpreting the visualizations into meaningful insights. We also performed sanity checks on the results at various stages of the process and built a presentable narrative around the data.

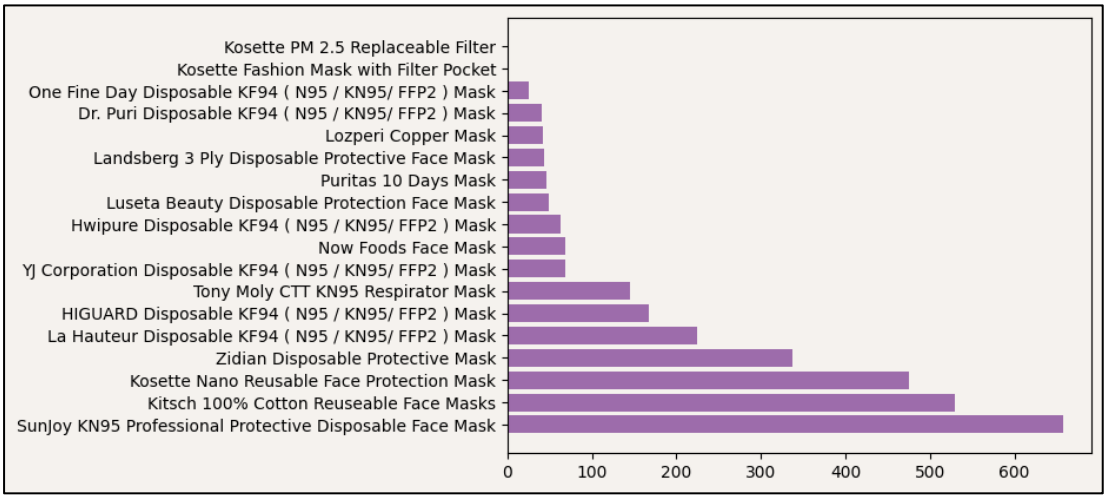
Problems worked on:

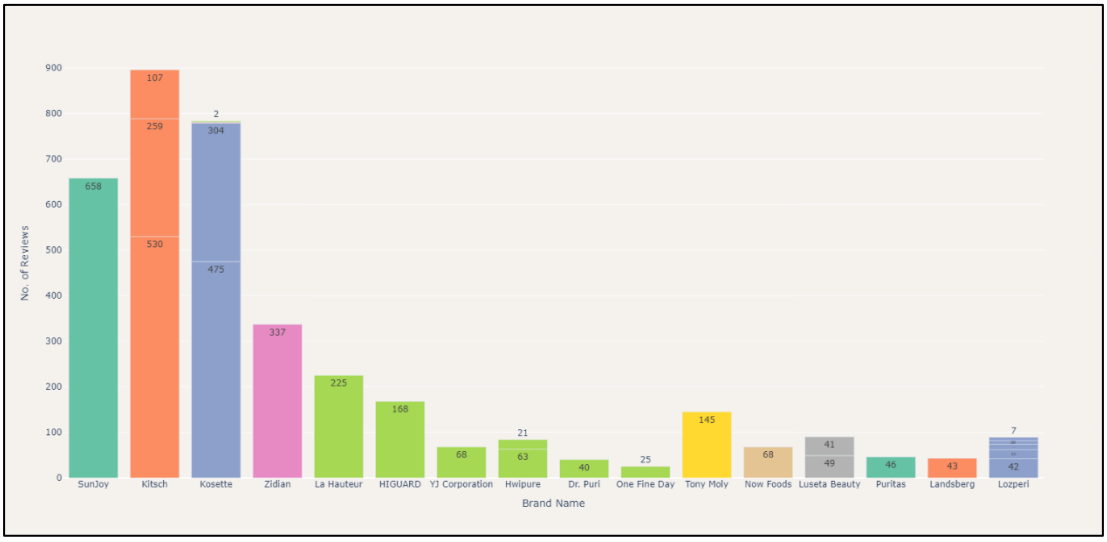
1. What are the most popular face masks in the store, and what are their key features?
2. What do consumers like about the most popular face masks, and why?
3. What are the different profiles of consumers who buy face masks, and how can we segment them based on their demographics, behaviors, and preferences?
4. How can additional data sources be used to gain more insights into the types of consumers buying different face masks?
5. What are the key findings and implications of this analysis for the client's marketing strategy and research and development department?
6. What are the potential constraints and improvements to the analysis, and how can we address them?
7. Identify and suggest a few extra features that could have helped gain valuable insights into the different types of consumers who purchase these face masks.

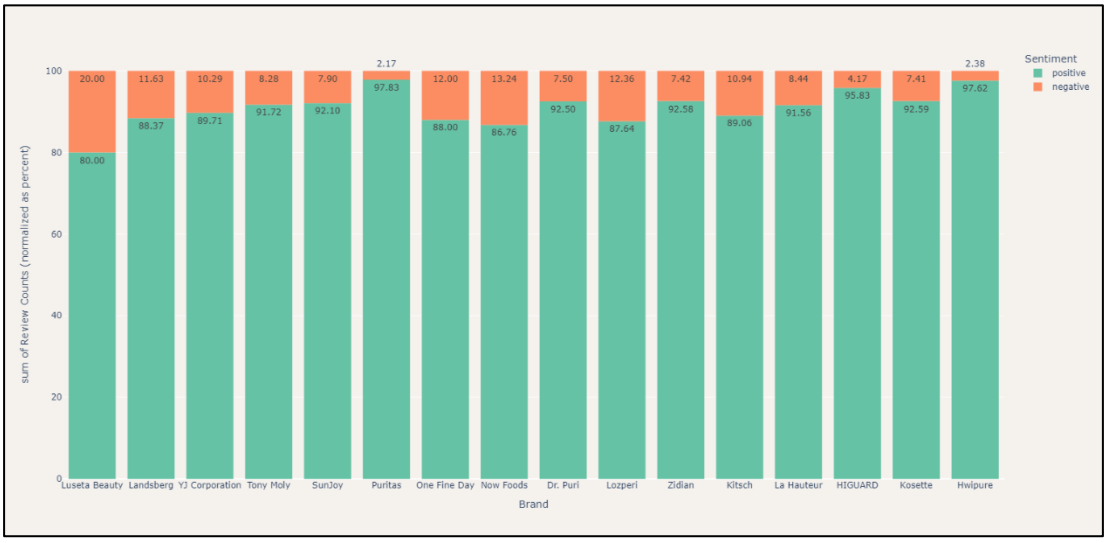
Key findings:

1. **Key Features of Five Most Popular Masks**:
   100% Cotton Reusable Face Masks, Nano Reusable Face Protection Mask, KN95 Professional Protective Disposable Face Mask, Disposable Protective Face Mask, and Disposable KF94 (N95/KN95/FFP2) Mask are the top five selling masks in the store. Blush, neutral, and leopard are the most popular shades for 100% Cotton Reusable Face Masks.

Here are a few illustrations:





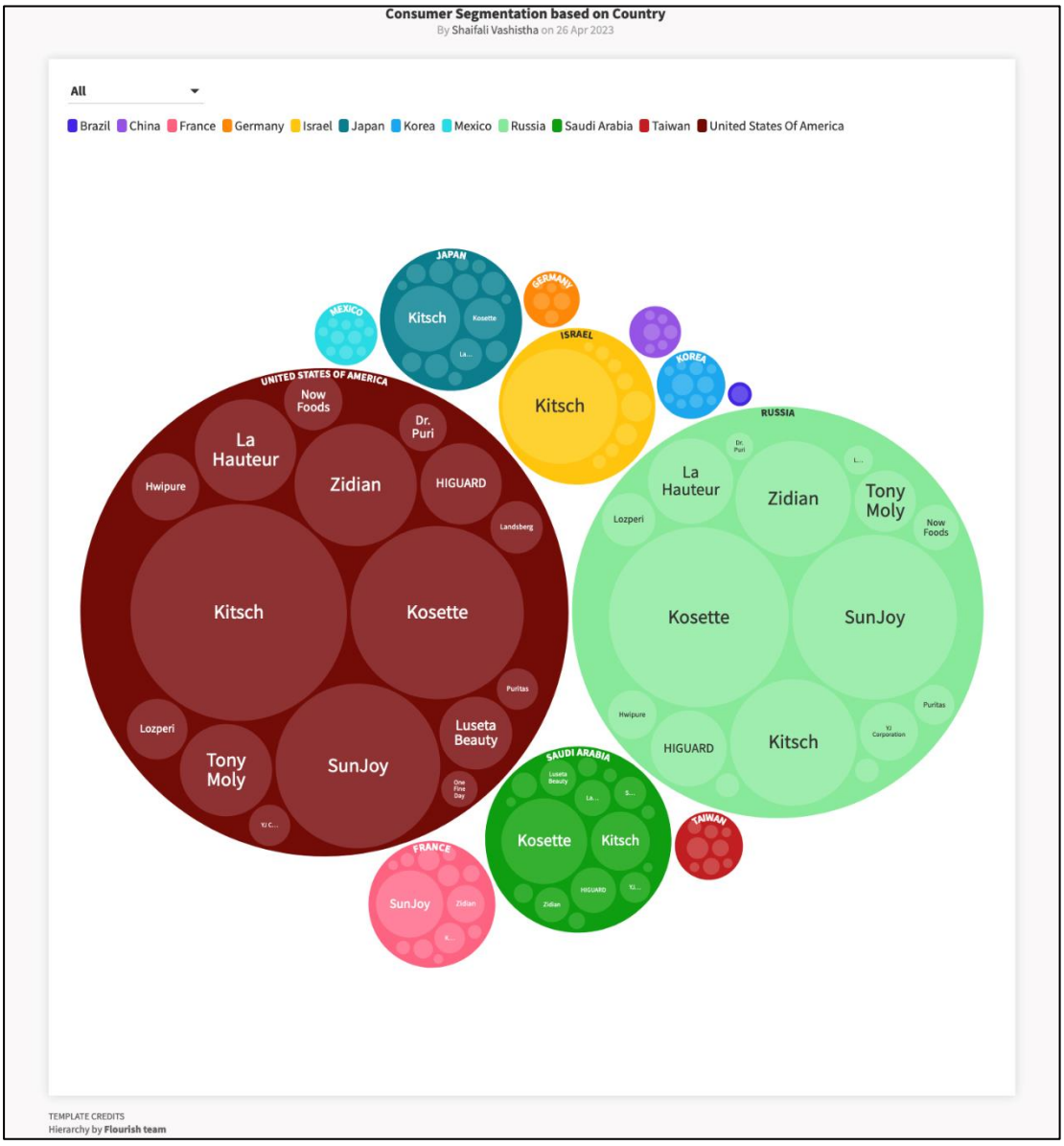


2. **What do consumers like about the most popular face masks, and why?**
   Through **Text Analysis** on positive user reviews, we deduced that consumers like face masks that have comfortable fitting on the face, nose, ear, and chin, are made of cotton or cloth, are breathable, washable, and reusable, and have a low price.

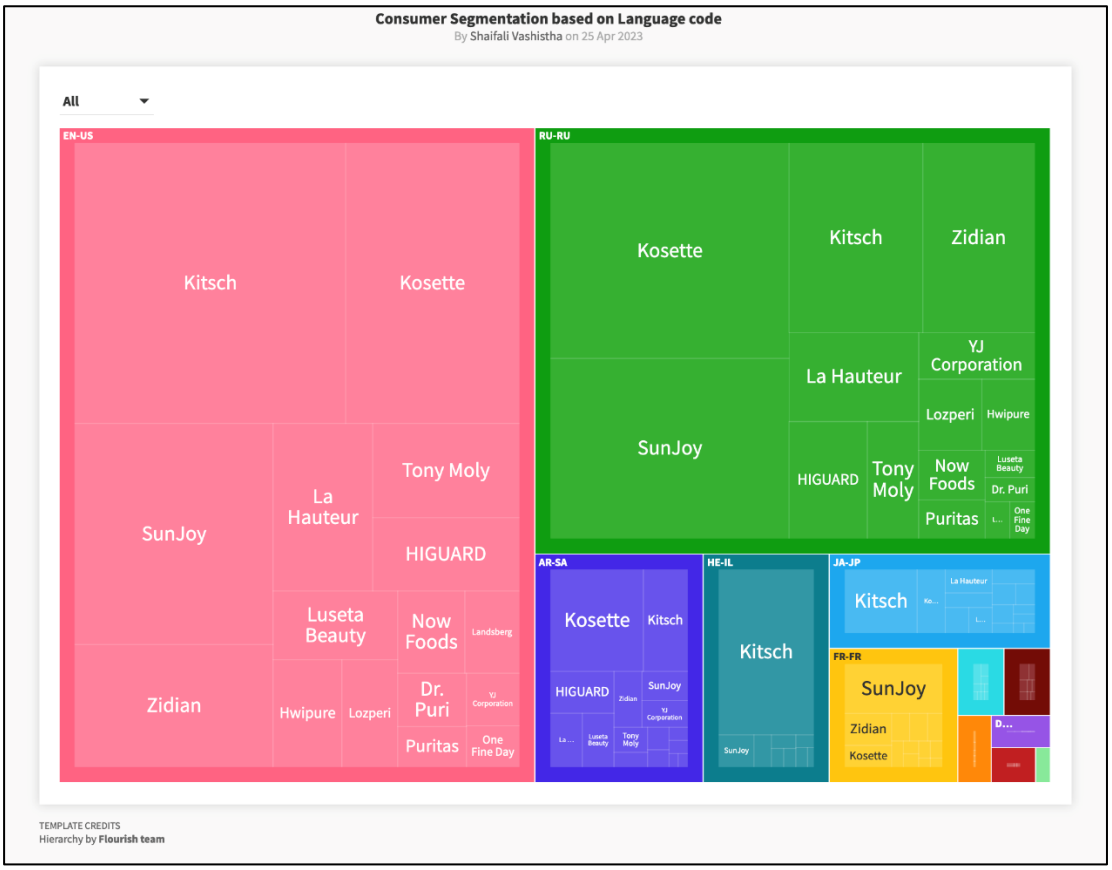
3. **Consumer Segmentation:**

Consumers were divided into groups based on their demographics (age, gender, marital status, family size, income, education, race, occupation, nationality, language, and religion), behavior (how much they spend on goods, how frequently they buy them, what features they look for, what buying patterns they use, and what decisions they make while shopping), and preferences (preferred product attributes, brands, prices, and purchasing habits). To understand consumer behavior, we employed behavioral segmentation plots like heatmap and rating flow plots, as well as demographic segmentation plots by countries (ethnicity) and languages.

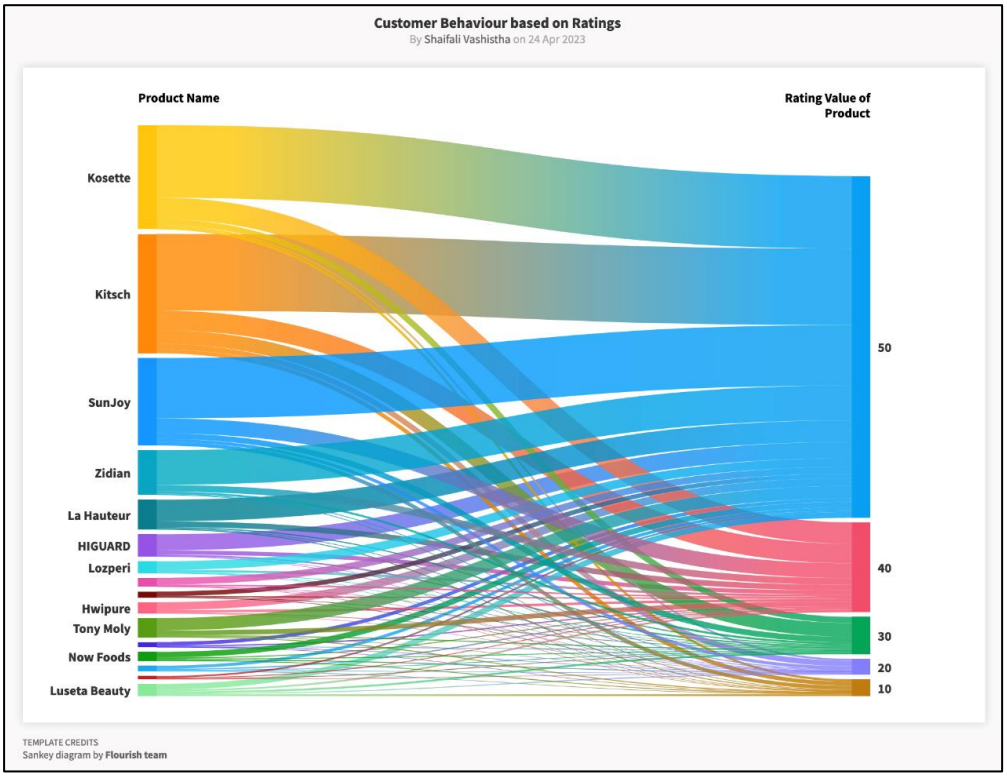
Segmentation by Countries (Region):

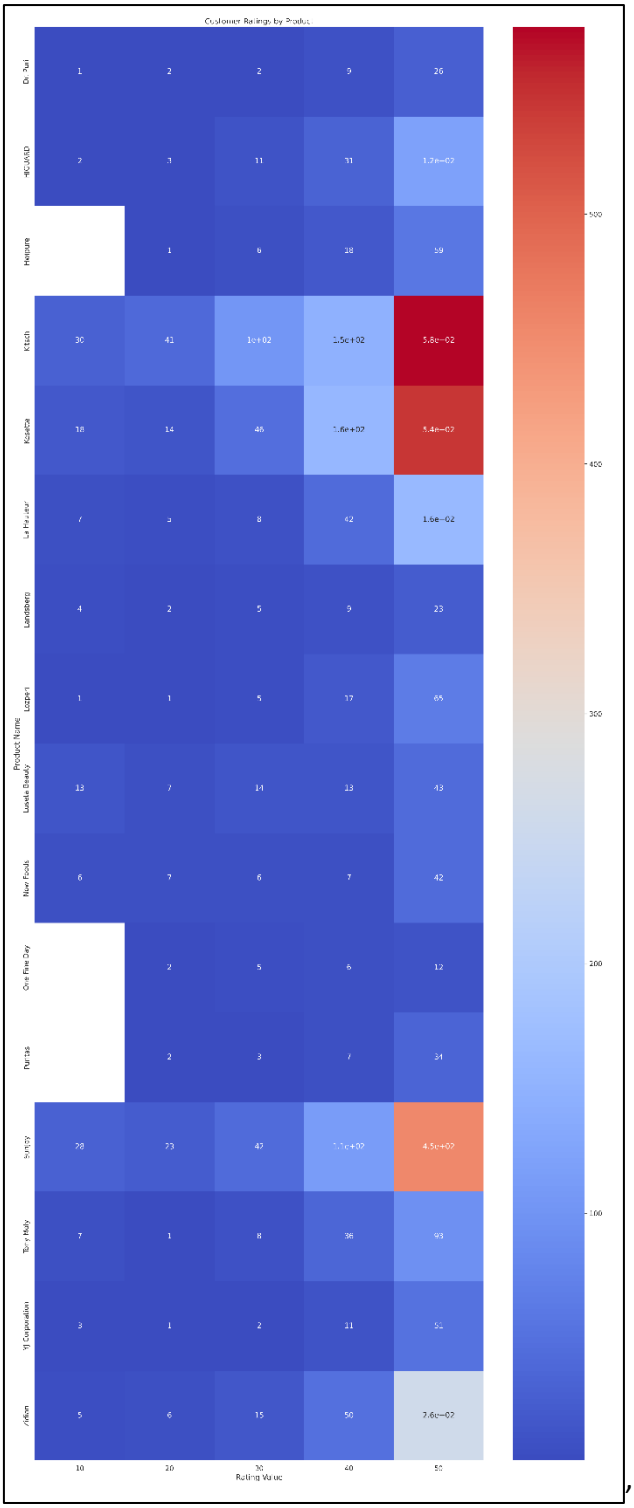


Segmentation by Language :



Behaviour of Customer to buy a product based on its product ratings:




Customer Ratings by Product
Product Name
Rating Value
10
20
30
40
50
100
200
300
400
500

Behaviour of Customer to buy a product based on its product ratings:

Behaviour of Customer to buy a product based on abuse count

abuseCount / Brand name	0	1	2	3	4
Dr. Puri	36	4			
HIGUARD	158	8		2	
Hwipure	77	6	1		
Kitsch	871	22	3		
Kosette	742	35	3	2	1
La Hauteur	217	8			
Landsberg	37	5	1		
Lozperi	86	2	1		
Luseta Beauty	83	6	1		
Now Foods	66	2			
One Fine Day	23	2			
Puritas	44	2			
SunJoy	651	6	1		
Tony Moly	133	10	2		
YJ Corporation	66	1	1		
Zidian	331	6			

Preference of Customer Price or Price/Mask:

Preferences of Customer Price or Price/mask

Price of Package		Customer Count
AUD	2.95	299
AUD	3.93	68
AUD	4.49	25
AUD	5.61	225
AUD	5.91	2
AUD	6.61	46
AUD	6.85	89
AUD	8.86	658
AUD	11.81	2
AUD	15.35	337
AUD	17.72	896
AUD	22.44	779
AUD	26.57	145
AUD	36.54	43
AUD	44.31	49
AUD	47.24	40
AUD	49.61	41
AUD	50.19	21

Price/Mask		Customer Count
AUD	0.31	337
AUD	0.73	43
AUD	0.89	658
AUD	0.89	49
AUD	0.99	41
AUD	2.01	21
AUD	2.20	46
AUD	2.36	40
AUD	2.95	299
AUD	3.93	68
AUD	3.94	2
AUD	4.49	25
AUD	5.31	145
AUD	5.61	225
AUD	5.91	896
AUD	5.91	2
AUD	6.85	89
AUD	22.44	779

4. **Additional Data Sources:**
   We can leverage additional data sources, such as income, gender, and age group, to learn more about the types of consumers that purchase different face masks. We can determine which types of masks are more popular among specific populations by examining consumer data on gender.

5. **Implications for the client's marketing strategy and research and development department:**
   According to our analysis, the client can introduce more affordable or expensive masks to appeal to market segments, depending on their income levels. To appeal to consumer preferences, we also advise the client to create items with soft fittings, materials made of cotton or cloth, breathability, washability, and reusability.
6. **Constraints and improvements:**
   I. Data quality is crucial for accurate analysis, and there is always room for improvement in cleaning and pre-processing data to enhance its quality.
   II. Sample size can impact the accuracy of insights and a larger sample size can reduce the impact of outliers and provide more accurate insights.
   III. The data may have inherent biases which can affect the accuracy of the analysis and lead to incorrect conclusions.
   IV. Relevant variables that could impact the analysis may not be included in the data, so it's important to consider additional factors that may influence the results.
   V. External factors that can impact the analysis, such as changes in the economy or social trends, should be considered.

   Possible improvements to address these constraints include:
   I. Collecting more data from additional sources to increase the sample size and improve accuracy and reliability of the analysis.
   II. Further cleaning of the data to remove any inconsistencies and errors that may affect the analysis.
   III. Use of advanced analytics like machine learning algorithms and natural language processing models to provide more insights into consumer behavior and patterns.
   IV. Use of other visualization methods that are more complex and can give a better understanding of the dataset.

7. **Suggestions and Additional Features**

   I. **Age and gender**: Demographic information such as age and gender can provide insight into the preferences of different age groups and genders for certain types of face masks. For example, younger individuals may prefer more fashionable face masks, while older individuals may prefer more functional ones.

II. **Occupation**: Occupation can be a significant factor in determining the type of face mask consumers purchase. For example, healthcare workers may require more protective masks than someone who works in an office setting.
III. **Income level**: Income level can also play a role in determining the type of face mask purchased. Higher-income individuals may be more willing to invest in higher-quality, more expensive masks.
IV. **Reason for purchase**: Knowing the reason behind the purchase can provide valuable insight into the needs and preferences of consumers. For example, someone purchasing a mask for outdoor activities may prefer a mask with better breathability, while someone purchasing a mask for medical reasons may prioritize masks with higher levels of protection.
V. **Purchase frequency**: Analyzing how often consumers purchase face masks could provide insights into their level of concern and engagement with public health issues. For example, consumers who purchase face masks frequently may be more concerned about the spread of respiratory illnesses, while consumers who purchase face masks infrequently may only purchase them when they perceive an immediate threat.

Conclusion:

In conclusion, the purpose of this project was to analyze two datasets pertaining to face masks to offer insights to a manufacturer of personal care products. We were able to categorize consumers based on demographics, behaviors, and preferences as well as discover the most popular face masks and important qualities that customers look for through our analysis. The client's research and development department and marketing strategy benefited from our findings' implications as well. However, we also noted several limitations that might compromise the analysis's accuracy and offered potential solutions to deal with them. This project's overall results demonstrated the value of exploratory data analysis and visualization in getting insightful information and guiding business decisions.