



Decoding Market Segmentation: Utilising Cluster Analysis for Cosmetics Market Expansion

Ms. Arthi Reddy ^a, Ms. Aditi Gatkal ^b, Mr. Yogesh Kumar Lalwani ^c

^a M.Sc. Statistics Student, Department of Statistics, S K Somaiya College, Somaiya Vidyavihar University, Mumbai 400097, India

^b M.Sc. Statistics Student, Department of Statistics, S K Somaiya College, Somaiya Vidyavihar University, Mumbai 400097, India

^c M.Sc. Statistics Student, Department of Statistics, Indira Gandhi National Open University (IGNOU), New Delhi 110068, India

DOI: <https://doi.org/10.55248/gengpi.5.0624.1405>

ABSTRACT

Cosmetics have become a crucial part of everyday life, helping people express themselves and feel confident. People are becoming more beauty-conscious, and the trend is on the rise. Every day sees the arrival of new products and companies, each vying for attention, but only a few endure the test of time, securing their place in the market. Cosmetics market is facing serious challenges to meet the increasing expectations and demands of the consumers. To thrive in the cosmetic industry, the companies have to decode how certain crucial factors influence consumers' buying choices. This study delves deep into understanding the dynamics of consumers' cosmetics buying behaviour, analysing the impact of various influential variables. The research took place in March 2024 across six cities in North-Western Maharashtra. A questionnaire, utilising snowball sampling, was distributed to consumers aged between 16 to 30 years. The sample comprised 222 respondents, with 122 female and 100 male participants. Principal component analysis was first used to identify key variables influencing consumer behaviour and then the consumers were grouped using cluster analysis to uncover distinct consumer segments.

Keywords: PCs, Consumers' Cosmetics buying behaviour, Cosmetics Market, Segmentation Analysis

1. Introduction

Cosmetics hold a unique power to elevate happiness and confidence, skillfully enhancing one's appearance and strengthening self-assurance with a touch of glamour. These products forge a deeper connection between individuals and their sense of self, enabling them to present the version of themselves they feel most comfortable and confident in. The realm of cosmetics extends far beyond facial beauty care, encompassing skin care, hair care, fragrances, and various other aspects of personal grooming. Naturally, the diverse mindsets of consumers lead to a wide array of choices and preferences when it comes to purchasing cosmetic products.

The cosmetics market is a vibrant and dynamic sector, influenced by ever-changing trends and consumer demands, and boasting a global value in the billions. Constant innovation and diversification are hallmarks of this industry, as it strives to meet the evolving preferences and needs of consumers. This effort results in a broad range of products designed to cater to various tastes, budgets, and concerns. The fierce competition within the cosmetic market drives brands to outpace each other through innovation, effective branding, strategic pricing, and sustainability efforts. Various factors shape consumers' cosmetics buying behaviour, including product quality, brand reputation, price, product suitability, current trends, ingredient transparency, customer reviews, marketing campaigns, and celebrity endorsements.

The Indian cosmetics market, in particular, stands out as a rich and diverse segment within this global industry, reflecting a complex tapestry of cultural influences and consumer demands. As part of this dynamic global sector, the Indian market continuously innovates and diversifies its offerings to keep pace with the ever-evolving preferences and needs of its consumers.

This study aims to explore the intricacies of consumers' cosmetics buying behaviour, focusing on the various influential variables that impact purchasing decisions. To achieve this, cluster analysis has been employed to segment the cosmetic market effectively based on consumer preferences. Principal component analysis was initially used to identify key variables influencing consumer behaviour. Subsequently, consumers were grouped using cluster analysis to uncover distinct consumer segments. Through a detailed examination of consumer data and clustering techniques, this research seeks to provide a comprehensive understanding of how companies can harness these insights to drive market expansion and better meet consumer needs.

2. Objectives

1. To segment the cosmetic consumer market into the groups of similar preferences.

2. To segment the female consumer cosmetic market on the basis of similar preferences.
3. To segment the male consumer cosmetic market on the basis of similar preferences.

3. Methodology

Sample Description: The sample collected contains 222 responses from six cities in Maharashtra: Mumbai, Thane, Palghar, Pune, Nashik, and Ahmednagar (now Ahilya Nagar). Out of these responses, 122 were from females and 100 were from males, all aged between 16 to 30 years. The sampling method employed was Snowball sampling.

Data Collection: For this study, primary data was collected from individuals through a structured questionnaire to understand the level of influence a particular variable has on the consumer's cosmetics buying behaviour. The questionnaire was divided into two sections: Demographic information of the responders and Level of influence section having Likert scale ranging from 1-5 where 1 implied "No Influence" and 5 implied "Extreme Influence" corresponding to 20 attributes. To clarify what was considered as cosmetics, separate lists of products for males and females were provided. A multiple-choice grid format was utilised to reduce the complexity of the questionnaire.

Statistical Tools: The RStudio 2024.04.1+748 with R version 4.4.0 was used for analysing the data.

This study is divided into three sections, with the collected data segmented into three parts. The first dataset encompasses all respondents, while the second and third datasets specifically include female and male respondents respectively. The analysis involves Cronbach's Alpha Reliability Test, Principal Component Analysis (PCA), Cluster Analysis and ANOVA.

4. Analysis and Results

4.1. Reliability Statistics

Reliability analysis evaluates the correlation of each item with the entire scale, determining the appropriateness of the variables for Principal Component Analysis.

Table 1 - Reliability statistics

Dataset	Cronbach's Alpha	Cronbach's Alpha Based on Standardised Items	N Of Items
Entire	0.880	0.88	20
Female	0.870	0.87	20
Male	0.890	0.89	20

For the present study, these high values indicated strong reliability, confirming that the data was suitable for Principal Component Analysis.

4.2. Principal component analysis (PCA)

PCA serves various purposes, including exploratory data analysis, dimensionality reduction, and visualisation of high-dimensional data. In this study, a threshold of 90% variability was established to pinpoint the most significant variables that explained 90% of the total variation within the cosmetic consumer market. Across the entire dataset, 14 principal components (PCs) were distinguished, collectively representing 90% of the variation present in the data. These 14 principal components were then used to select the top variables based on their higher loadings, ensuring a comprehensive representation of the dataset's underlying structure. For Male Dataset, 14 principal components were identified. Similarly, for the Female Dataset, 13 principal components were identified.

4.3. Cluster analysis (CA)

Similar to Principal Component Analysis (PCA), cluster analysis operates by transforming the original dataset into a new form that highlights similarities and differences among the observations. In this study, to obtain an optimal number of clusters, the NbClust package was used to compute various clustering indices and then, the majority rule criterion was used on that set of indices. For Entire and Female datasets, K-Means Clustering was performed to segment the cosmetic consumer market on the basis of the variables retained using Principal Component Analysis for each dataset respectively. However, in the case of Male dataset, K-Means Clustering gave imbalanced clusters. Therefore, Hierarchical Clustering was employed on the dataset with 14 PCs using Complete Linkage Method with Manhattan distance was utilised to get balanced clusters. Finally, Analysis of Variance was carried out to validate the clusters acquired.

Table 2 - Cluster Centers (Entire)

Variables	Cluster 1	Cluster 2
Recommendations	2.8476	3.8803
Origin/Source	3.1048	4.0342
Current Trend	2.5048	3.5128
Ingredients	3.3048	4.4274
Product Design	2.8381	3.6838
Ratings & Reviews	3.3048	4.2991
Seller Reputation	2.5714	3.3590
Availability	3.1429	4.1538
Product Description	2.9429	4.1026
Website Store	2.6000	3.9915
Advertisements	2.3810	3.5556
Brand Name	3.1619	4.1111
Quality	4.0095	4.6838
Deals & Discounts	2.9238	4.0684

Cluster 1 : Moderate Influence Across Most Factors (N=105)

Individuals in this group were moderately influenced by various factors when making decisions. The quality of the product held the highest influence (4.0095), indicating that individuals in this cluster placed significant importance on the product's quality. Ingredients and Ratings & Reviews also had a considerable influence (both at 3.3048), suggesting these consumers valued detailed information about what the product contained and other customers' experiences. Origin/Source (3.1048), Availability (3.1429), and Brand Name (3.1619) moderately affected their decisions as well. Components such as Product Design (2.8381), Product Description (2.9429), and Deals & Discounts (2.9238) had a moderate influence, reflecting a balanced approach where these elements were considered but were not primary decision drivers. Advertisements (2.3810), Current Trend (2.5048), Seller Reputation (2.5714), Website/Store (2.6000), and Recommendations (2.8476) exerted slight to moderate influence on this cluster, indicating that personal endorsements, trends, seller details, online store interface, and marketing efforts were less critical in their purchasing decisions.

Cluster 2 : High to Extreme Influence Across All Factors (N=117)

This cluster of individuals demonstrated a higher sensitivity to all the factors considered, reflecting a more discerning and demanding consumer group. Quality stood out as the most influential component (4.6838), suggesting that this cluster had a near-extreme concern for product excellence. Ingredients (4.4274), Ratings & Reviews (4.2991), Availability (4.1538), and Product Description (4.1026) also held high to extreme influence, highlighting the importance of comprehensive product information and accessibility. Brand Name (4.1111) and Origin/Source (4.0342) were also highly influential, underscoring a preference for reputable and well-known brands with a clear product Origin/Source. Recommendations (3.8803), Current Trends (3.5128), Product Design (3.6838), and Seller Reputation (3.3590) showed that this cluster valued social proof, trendy items, aesthetic appeal, and Seller Credibility more than Cluster 1. Website/Store (3.9915) and Advertisements (3.5556) had a high influence, indicating that the online shopping experience and marketing efforts were crucial in their decision-making. Deals & Discounts (4.0684) also played a significant role, suggesting that this cluster was keen on finding value for money. Overall, Cluster 2's decisions were shaped by a comprehensive assessment of various factors, reflecting a more meticulous and information-driven approach.

Table 3 - ANOVA Table (Entire)

Variables	Df	Sum Sq	Mean Sq	F value	Sig. Value
Recommendations	1	59.02	59.02	59.59	4.1e-13
Origin/Source	1	47.8	47.8	39.88	1.5e-09
Current Trend	1	56.23	56.23	47.31	6.2e-11
Ingredients	1	69.74	69.74	72.75	2.4e-15
Product Design	1	39.57	39.57	34.89	1.3e-08

Ratings & Reviews	1	54.72	54.72	70.49	5.7e-15
Seller Reputation	1	34.32	34.32	25.8	8.1e-07
Availability	1	56.56	56.56	62.82	1.1e-13
Product Description	1	74.42	74.42	85.98	<2e-16
Website/Store	1	107.1	107.14	117.7	<2e-16
Advertisements	1	76.35	76.35	68.38	1.3e-14
Brand Name	1	49.86	49.86	56.02	1.7e-12
Quality	1	25.16	25.156	33.28	2.7e-08
Deals & Discounts	1	72.49	72.49	74.23	1.4e-15

The ANOVA table for the entire dataset showed that all 14 variables selected based on PCA were significantly different across the two clusters. There was a significant difference in the level of influence these variables had on the consumers' buying decisions.

Table 4 - Cluster Centers (Female)

Attribute	Cluster 1	Cluster 2
Recommendations	2.8413	4.0339
Product Design	2.7778	3.8305
Current Trend	2.4127	3.6441
Deals & Discounts	3.2222	4.2712
Product Description	3.4444	4.0169
Website/Store	2.8571	3.9492
Refund Policy	3.2381	4.3051
Ratings & Reviews	3.4127	4.2203
Seller Reputation	2.5873	3.3898
Brand Name	3.1111	4.0678
Availability	3.3175	4.2203
Origin/Source	3.3968	3.9831
Advertisements	2.3492	3.7458

Cluster 1 : Slight to Moderate Influence Across Most Factors (N=63)

The Cluster 1 females were moderately influenced by various factors in their purchasing decisions. The most influential factors for this group were Product Description (3.4444), Origin/Source (3.3968), Availability (3.3175), and Ratings & Reviews (3.4127). These consumers prioritised clear, detailed descriptions of products, the source of the products, the ease of access to those products, and feedback from other buyers. Deals & Discounts (3.2222), Refund Policy (3.2381), and Brand Name (3.1111) were moderately influential on female consumers' buying behaviour. Website/Store (2.8571), Recommendations (2.8413), and Product Design (2.7778) also had a slightly moderate influence, indicating that financial incentives and post-purchase assurances were important. However, factors like Seller Reputation (2.5873), Current Trend (2.4127), and Advertisements (2.3492) had lesser influence, suggesting that these consumers were less swayed by trends and marketing efforts. Overall, this cluster was moderately influenced by practical and financial aspects of shopping.

Cluster 2 : High to Extreme Influence Across All Factors (N=59)

The females of Cluster 2 were influenced by numerous factors, indicating that they were perceptive consumers who were highly receptive to various influences. The most critical factors for this cluster were Refund Policy (4.3051), Deals & Discounts (4.2712), Ratings & Reviews (4.2203), and Availability (4.2203), highlighting the importance of post-purchase satisfaction, financial savings, social proof, and the availability of products in their purchasing decisions. Recommendations (4.0339), Product Design (3.8305), Brand Name (4.0678), and Product Description (4.0169) also had a high influence, suggesting that this group valued expert opinions, aesthetic and functional aspects of products, and the reputation of brands. Even factors such as Seller Reputation (3.3898), Current Trend (3.6441), Advertisements (3.7458), Website/Store (3.9492), and Origin/Source (3.9831) had moderate to

higher influence on purchasing decisions compared to Cluster 1. This indicated that Cluster 2 was much more sensitive to a broad range of factors, including ethical considerations and brand prestige.

Table 5 - ANOVA Table (Female)

Variable	Df	Sum Sq	Mean Sq	F value	Sig. Value
Recommendations	1	43.34	43.34	47.13	3.13e-10
Product Design	1	33.77	33.77	33.99	4.77e-08
Current Trend	1	46.2	46.2	41.13	2.94e-09
Deals & Discounts	1	33.52	33.52	33.93	4.88e-08
Product Description	1	9.99	9.986	12.41	0.000604
Website/Store	1	36.33	36.33	36.77	1.59e-08
Refund Policy	1	34.69	34.69	31.55	1.28e-07
Ratings and Reviews	1	19.87	19.873	25.53	1.58e-06
Seller Reputation	1	19.62	19.622	15.36	0.000148
Brand Name	1	27.89	27.885	30.43	2.02e-07
Availability	1	24.84	24.837	28.17	5.16e-07
Origin/Source	1	10.47	10.47	8.051	0.00534
Advertisements	1	59.42	59.42	65.12	6.11e-13

The ANOVA table for the female dataset indicated that all 13 PCA-selected variables differed significantly between the two clusters. It also showed a significant difference in how these variables influenced females' cosmetics buying choices within each cluster.

Table 6 - Cluster Centers (Male)

Variable	Cluster 1	Cluster 2
Origin Source	2.8182	4.0179
Ingredients	3.0455	4.2321
Recommendations	2.9091	3.7143
Product Description	2.6136	3.9286
Quality	3.7045	4.7857
Current Trend	2.7273	3.3393
Refund Policy	2.6591	3.6964
Availability	3.0000	4.0357
Seller Reputation	2.3636	3.5000
Price	3.3409	3.5536
Environmental Impact	2.5682	3.9107
Product Suitability	3.2955	4.5536
Brand Name	3.3182	4.1250

Cluster 1 : Slight to Moderate Influence Across Most Factors (N=44)

The males within this group were moderately affected by most factors, leaning towards a slight influence overall. Quality still had a moderately high influence (3.7045), though to a lesser extent than in Cluster 1. Price (3.3409), Brand Name (3.3182), and Product Suitability (3.2955) had a moderate influence, indicating some concern for the cost of the product, product fit, and reputation. Ingredients (3.0455) and Availability (3.0000) were moderately influential, showing these factors were considered but were not that critical. Recommendations (2.9091), Origin/Source (2.8182), and Current Trends (2.7273) had a moderate influence, suggesting these consumers were less driven by personal endorsements, product origin, and trends. Refund Policy (2.6591), Product Description (2.6136), and Advertisements (2.6364) also had a slightly moderate influence, indicating these factors were less critical in

their decision-making. Seller Reputation (2.3636) and Environmental Impact (2.5682) exerted a slight influence, highlighting that the credibility of the seller and sustainability were less significant.

Cluster 2 : High to Extreme Influence Across Most Factors (N=56)

This group of males exhibited a high to extreme influence from various factors in their decision-making processes. Quality was the extremely influential factor (4.7857), indicating a paramount concern for product excellence. Product Suitability (4.5536), Ingredients (4.2321), and Origin/Source (4.0179) were also highly influential, showing that these consumers valued detailed information about what the product contained and where it came from, as well as how well it fit their needs. Availability (4.0357) and Brand Name (4.1250) also had high influence on purchasing decisions of male consumers, reflecting a preference for readily accessible products and reputable brands. Product Description (3.9286), Recommendations (3.7143), Price (3.5536), and Refund Policy (3.6964) also had a considerable influence, suggesting the importance of detailed product information, personal endorsements, product cost, and purchase security. Environmental Impact (3.9107) and Seller Reputation (3.5000) had moderately high influence on this cluster, indicating a balanced concern for sustainability and credibility of the seller. Current Trends (3.3393) and Advertisements (3.2321) exerted a moderate influence, suggesting these factors were considered but were not primary drivers of their purchasing decisions.

Table 7 - ANOVA Table (Male)

Variable	Df	Sum Sq	Mean Sq	F value	Sig. Value
Origin/Source	1	35.46	35.46	32.32	1.35e-07
Ingredients	1	34.7	34.7	29.34	4.33e-07
Recommendations	1	15.98	15.98	14.62	0.000231
Product Description	1	42.6	42.6	38.61	1.26e-08
Quality	1	28.8	28.8	42.39	3.21e-09
Current Trend	1	9.23	9.23	7.106	0.00899
Refund Policy	1	26.51	26.51	17.83	5.41e-05
Availability	1	26.43	26.43	27	1.11e-06
Seller Reputation	1	31.82	31.82	25.52	2.03e-06
Price	1	1.11	1.11	0.675	0.413
Environmental Impact	1	44.41	44.41	33.13	9.86e-08
Product Suitability	1	39	39	58.8	1.28e-11
Brand Name	1	16.04	16.04	17.53	6.18e-05
Advertisements	1	8.75	8.75	5.945	0.0166

The ANOVA table for the male dataset revealed significant differences across both clusters for all 14 variables selected using PCA, with the exception of Price variable. There was no significant difference between the levels of influence that the cost of cosmetic products had on male consumers. It also showed a significant difference in how these variables (except Price) influenced cosmetics buying decisions of males within each cluster.

5. Conclusion

This study provides valuable insights into the cosmetic market by segmenting consumers based on their preferences.

For the entire sample, variables such as Price, Product Suitability, Variety, Refund Policy, Celebrity Endorsement, and Environmental Impact exhibit similar effects across the clusters, leading to minimal variability in their influence. As a result, these variables were excluded from the segmentation or clustering analysis. The resulting clusters reveal that:

- Cluster 1: Moderately prioritises product quality, ingredients, and ratings & reviews.
- Cluster 2: Displays extreme sensitivity to all factors, particularly emphasising product excellence, comprehensive information, and the online shopping experience.

For female consumers, variables such as Price, Quality, Variety, Ingredients, Product Suitability, Celebrity Endorsement, and Environmental Impact also show comparable effects across the clusters, leading to their exclusion from the segmentation analysis. The clusters identified are:

- Cluster 1: Moderately influenced by practical aspects and somewhat indifferent to trends and advertising.

- Cluster 2: Highly influenced by a wide array of factors, especially financial benefits, product quality, and brand reputation.

For male consumers, variables including Ratings & Reviews, Deals & Discounts, Product Design, Variety, Celebrity Endorsement, and Website/Store show consistent effects across the clusters, resulting in their exclusion from the segmentation analysis. The clusters identified are:

- Cluster 1: Prioritises product authenticity, quality, and environmental impact.
- Cluster 2: Considers availability, brand reputation, and current trends to varying degrees in their purchasing decisions.

Overall, the study highlights that certain variables consistently influence consumer preferences across different segments, leading to their exclusion from cluster analysis, while other variables help distinguish between the different consumer groups within each gender.

6. Limitations

The study was limited to the participants of age ranging from 16 to 30 years. The sample size is small. Snowball sampling method was employed which might not accurately reflect the views of the broader population. The study specifically focuses only on the area of North-Western Maharashtra.

7. Scope for Further Research

The study offers several paths for future research. Additional Geographical regions can be included to get a better perspective of the consumer market. A large sample size and different sampling methods can be opted. The researchers can explore more variables to get a deeper understanding of the consumer's cosmetics buying behaviour. Predictive modelling and a cross-industry approach can forecast trends and provide comparative insights. Also, examining psychographic factors such as personality and lifestyle can enhance understanding of consumer motivations.

References

- Mohanty, S. (2013). Cosmetic segmentation and clustering. *Management Research Paper*, 2(1), ISSN 2277-8179.
- Sahrash, M., & Rao, H. (2022). The cosmetic industry in India: A PESTLE analysis. In *Perspectives on Business Management & Economics* (Vol. VI, pp. 46). St. Francis College for Women. ISBN: 978-81-951151-0-5.
- Majumdar, S. (2015). Analysis of factors influencing consumer preferences for green cosmetic and food products: A study in and around Kolkata (West Bengal, India) (Doctoral dissertation, ICFAI University Jharkhand).
- Sankaranarayanan, K. G., & Mekoth, N. (2014). Cosmetic consumption pattern among male college students: A cluster analytic segmentation approach. *International Journal of Research in Commerce & Management*, 5(6), ISSN 0976-2183.
- Sundarambal, K., & Edhayavarman, D. (2018). A study on consumer preference on cosmetics among college girls with reference to Pudukottai District. *International Journal of Creative Research Thoughts (IJCRT)*, 6(1), 366. ISSN: 2320-2882.
- Winter, E.A., Litvinova, T.M., Babaskin, D.V., Babaskina, L.I., & Krylova, O.V. (2018). Consumer preferences in cosmetic products within the market's drugstore segment. *Journal of Pharmaceutical Sciences and Research*, 10(9), 2232-2235.
- Gayathri, A., & Deepika, A. (2019). A study on consumer preference towards cosmetic brand among college girls. *Journal of Emerging Technologies and Innovative Research (JETIR)*, 6(3), 67. ISSN: 2349-5162.
- Tripathi, N., Chaubey, D. S., & Tripathi, D. M. (2016). Factors identification for merchandising of cosmetics products using cluster analysis. *International Journal of Multidisciplinary Educational Research*, 5(8/4), ISSN 2277-7881.