A Study on Data Scraping of Cosmetic Products from E-Commerce Website

Nandini L[1] and Dr. Manjunath[2]

Student, Department of MBA[1]

Assistant Professor, Department of MBA[2]

RNS Institute of Technology, Bengaluru, Karnataka, India

Abstract: *The popularity of cosmetic products has snowballed over the past few years. Hence, there have been many more brands and makeup products on the market, especially in the online market. To analyze the online marketing trends in the Indian cosmetic industry, we have used some of the most popular Indian websites like Amazon, Flipkart, Ulta and Sephora to scrape the product data. This dataset also contains the ingredients of these products, which can be further analyzed..*

Keywords: Data mining, scraping, cosmetic products, websites, market trends, Consumer preferences, customer satisfaction

I. INTRODUCTION

Data scraping is the automated process of extracting vast amounts of information from websites. Online stores like Flipkart, Amazon, Ulta, and Sephora are important sources of information for the cosmetics sector. The market analyses the consumer behavior, consumption pattern, Consumer preferences and market trends. The research identifies the factors that influencing the ratings of cosmetic products. The cosmetics industry is a worldwide market that offers a vast array of items, from makeup to haircare and skincare to body care. This dataset offers a thorough overview of the diverse cosmetic items available for purchase in the USA and other locations. This business is home to tiny, niche brands as well as big, international corporations that all target different market niches. We can gain a comprehensive picture of the dynamics of the cosmetics market by examining product categories, subcategories, prices, brands, customer ratings, and other factors.

II. OBJECTIVES

- To examine the dynamics of cosmetic products ratings from a data analytics perspective.
- To identify the factors influencing ratings of cosmetic products.
- To analyze the impact of Price, website, Country and Category on ratings of cosmetic products.

III. LITERATURE REVIEW

According to V. Srividya and P. Meghala (2019) prices across various websites using various characteristics, including product name, price, and rating. iPhone product details are obtained by scraping e-commerce websites, and the resulting data is saved in a CSV file. Data mining-based real-time product analysis solves this problem for the customer by providing products from multiple e-commerce sites in one location, each with different prices, plans, and offers from different e-business companies. Extracting HTML data from URLs and using it for one's own use is known as web scraping. The ubiquitous use of interactive graphs and charts allows us to obtain business insights. Tools for visualization are used to make these charts.

According to Dr. M. Rajeshwari (2022) referred to Extract product information off of e-commerce websites use Python to extract the necessary data from e-commerce websites for comparison. With so many e-commerce websites available in today's expanding internet world, getting product specifics out of them can be difficult at times. Web scraping is one of the most significant techniques for information collection. Web scraping is one automatic method for gathering a lot of data from websites. Web scraping is used in a wide range of businesses, such as e-commerce, market research, and brand surveillance.

IV. RESEARCH METHODOLOGY

The methodology using data scraping, product information from leading e-commerce platforms was collected, capturing attributes such as ratings, prices, brands, and product categories. Descriptive analytics and correlation analysis were conducted with SPSS and Excel to understand the relationships between pricing, ratings, and other product attributes. The dataset was categorized to examine trends within specific segments (e.g., skincare, haircare) and to highlight region-specific preferences.

V. SAMPLING

The study employed convenience sampling by collecting data from selected e-commerce platforms (Amazon, Flipkart, Sephora, and Ulta), which may introduce bias as it captures only a segment of the cosmetics market available on these platforms.

VI. STATISTICAL TOOLS

- **Microsoft Excel:** Organization data, pivot table, statistical analysis and visualization, such as creating charts, tables and calculating descriptive statistics.
- **SPSS Software:** Multiple regression analysis is used to examine the relationship between independent variable and dependent variable.

VII. HYPOTHESIS FOR ANOVA

Null hypothesis (H0)
Ho- Category, Price, Country and Website components of cosmetic sale avenues do not influence the ratings given to the cosmetic products.

Alternative hypothesis (H1)
H1- Category, Price, Country and Website components of cosmetic sale avenues influence the ratings given to the cosmetic products.

ANOVA[a]

Model		Total squares	df	Mean square	F	Sig.
1	Regression	66.261	4	16.565	70.281	<.001[b]
	Residual	1355.519	5751	.236		
	Total	1421.780	5755			

Dependent variable: Rating

Predictors: (constant), Category(X4), Price(X1), Country(X3), Website(X2)

Inference: As SIG value of 0.000 is less than the standard P value of 0.05, we reject Ho(Category, Price, Country and Website components of cosmetic sale avenues do not influence the ratings given to the cosmetic products) and conclude that H1(Category, Price, Country and Website components of cosmetic sale avenues influence the ratings given to the cosmetic products).

COEFFICIENTS[a]

Model		Not Typical B	Ratios std Error	Not typical coefficients Beta	t	Sig.	Result
1	(Constant)	4.399	.024		181.92	.000	
	Price(X1)	.014	.006	.034	2.565	.010	Significant
	Website(X2)	.114	.016	.249	7.045	<.001	Significant
	Country(X3)	-.071	.035	-.071	-2.010	.044	Significant
	Category(X4)	.019	.004	.067	4.990	<.001	Significant

Dependent variable: Rating

VIII. DISCUSSION AND RESULTS

The study used data scraping of cosmetic products from popular e-commerce sites to examine market trends and consumer preferences within particular categories. It highlights the importance of product differentiation based on quality, sustainability, and branding in the cosmetics e-commerce space. Brands that respond to consumer demand for transparency, eco-friendly packaging, and ethical sourcing gain a competitive advantage. The findings suggest that understanding consumer preferences and pricing strategies is crucial for brands aiming to capture market share in an increasingly digital and competitive marketplace.

KEY FINDINGS

The study indicates a strong consumer preference for skincare and haircare products, particularly those with natural and sustainable ingredients. The study found that products with higher ratings often belong to brands that emphasize quality ingredients and ethical practices. Regional variations also emerged, with markets in Asia demonstrating a preference for affordable, locally produced brands, while Western markets showed stronger demand for premium, sustainable products.

This study finds the consumer preference by using

USA: Consumer preferences and demand for the USA based cosmetic products

INDIA: Consumer choices and preferences regarding Indian based cosmetic products

IX. CONCLUSION

The study provides a valuable framework for understanding the dynamics of the e-commerce cosmetics industry. It is also providing valuable insights into consumer preferences, product trends, and effective pricing strategies within the cosmetics industry. Brands that prioritize sustainability and transparency are better positioned to build loyalty and capture market share in the e-commerce landscape.

X. LIMITATIONS

- The data scraped from websites may contain missing fields or inconsistent formatting, potentially impacting the accuracy of the analysis.
- Potential biases in the scraped data, as it focuses on a limited number of websites and may not fully represent the market.
- The absence of detailed demographic data, such as age, gender, or geographic diversity within consumer reviews, limits the ability to analyze how preferences vary across consumer segments.
- As this study relies on scraped data, it may not capture recent changes in consumer behavior or new product releases, limiting the study's relevance over time.

REFERENCES

[1]. Dr. M. Rajeswari, D. Pavyaasree, 2022, Scrape the product details from ecommerce websites, Journal of emerging technologies and innovative research

[2]. P. Meghala (May 2019), Scraping and Visualization of Product Data from Ecommerce Websites, International Journal of Computer sports and engineering

Copyright to IJARSCT
www.ijarsct.co.in

DOI: 10.48175/IJARSCT-22141

ISSN
2581-9429
IJARSCT

213