**Ecommerce Cosmetic Dataset Analysis Report**

Introduction

The Ecommerce Cosmetic Dataset, sourced from Kaggle.com, comprises a comprehensive collection of data pertaining to various cosmetic products. This dataset forms the basis of a personal project aimed at analyzing trends and insights within the cosmetics industry. As the analysis is intended for dissemination on LinkedIn, Github , the potential audience encompasses professionals, enthusiasts, and stakeholders interested in consumer behavior, market trends, and the cosmetics sector..

Methodology

The analysis utilized SQL Server Management Studio (SSMS) for data exploration and manipulation, leveraging its querying capabilities. Power BI was employed for visualization, allowing for the creation of insightful charts and dashboards. Correlation analysis was conducted to identify relationships between variables and extract meaningful insights.

Data Exploration

The dataset comprises 15 columns, including website, form, ingredients, brand, price, and category. Initial exploration revealed inconsistencies in the data, necessitating cleaning and preprocessing to ensure data accuracy.

Key Insights

Product Distribution Across Different Websites:

The distribution of products across different websites reveals that Amazon has the highest number of products, followed by Flipkart, Sephora, and Ulta.

Category and Subcategory Analysis:

The most prevalent categories include lips, eyes, and body, with a wide range of subcategories such as perfume, eyeshadow, and lipstick indicating diverse product offerings.

Price Distribution:

The average price distribution highlights variations across categories and subcategories, with body perfume being the highest-priced category on average.

Brand Analysis:

Sephora Collection emerges as the brand with the highest market share, despite having a slightly lower average rating compared to other brands like Clinique and Maybelline.

Formulation Analysis:

Liquid and cream formulations are the most common, suggesting consumer preferences for these product forms.

Ingredient Analysis:

Ingredients such as Glycerin and Dimethicone are frequently used in cosmetic products, indicating their importance in formulation.

Color Analysis:

The majority of products have no specified color, followed by black and multicolor options, indicating a wide range of color preferences among consumers.

Rating Analysis:

The correlation analysis reveals a positive but relatively weak correlation (0.1835) between price and customer ratings, suggesting that higher-priced products tend to have slightly better ratings.

Interpretations

The analysis of the Ecommerce Cosmetic Dataset yields several important inferences and interpretations. The dominance of Amazon in the online cosmetics market suggests its strong market presence and consumer trust. The popularity of lips, eyes, and body categories underscores the significance of these product segments in consumer preferences. The higher average price of body perfume indicates potential luxury positioning and consumer willingness to invest in fragrance products. Despite Sephora Collection's market share leadership, its slightly lower average rating highlights the importance of brand reputation and customer satisfaction. The prevalence of liquid and cream formulations reflects consumer demand for easy-to-use and versatile products. Additionally, the widespread use of ingredients such as Glycerin and Dimethicone suggests their importance in cosmetic formulations. The absence of specified colors in the majority of products underscores the diversity of consumer preferences and the demand for versatile cosmetics. The weak positive correlation between price and ratings suggests that consumers may prioritize factors other than price when evaluating product quality and satisfaction. Overall, these inferences provide valuable insights into consumer behavior, market dynamics, and competitive strategies within the cosmetics industry.

Limitations and Caveats

While the analysis of the Ecommerce Cosmetic Dataset offers valuable insights into the cosmetics industry, it is imperative to acknowledge certain limitations and caveats. Firstly, the dataset's scope is restricted to products available on specific e-commerce platforms, potentially excluding a comprehensive representation of the broader cosmetics market. This limitation may introduce bias towards online consumer behavior and overlook data from brick-and-mortar stores or other online retailers. Additionally, variations in data quality and accuracy across different sources may result in inconsistencies or errors in the analysis. Moreover, reliance on the dataset's information may lead to limitations in granularity or detail, particularly in areas such as product descriptions or ingredient lists. Furthermore, the correlation analysis conducted assumes linear relationships between variables, which may not always hold true in practical scenarios. It's important to note that correlations do not imply causation, and other unaccounted factors may influence observed relationships. Despite these constraints, the insights gleaned from the analysis offer valuable perspectives on industry trends, but a cautious approach is warranted when interpreting the findings."

Conclusion

In conclusion, the analysis of the Ecommerce Cosmetic Dataset provides valuable insights into the dynamics of the cosmetics industry. Through exploration of product distribution, category trends, brand performance, and correlation analysis, I gained valuable understanding of consumer preferences and market trends. While acknowledging the limitations and caveats inherent in our analysis, the findings offer actionable insights for businesses, marketers, and industry stakeholders seeking to understand and navigate the complexities of the cosmetics market. Moving forward, continued research and analysis, coupled with robust data collection methods, will be essential in further refining our understanding of this dynamic and ever-evolving industry landscape.

**DASHBOARD**

