

Amazon Beauty Product Review and Analysis Summary

Tools & Software

The analysis was conducted using Python in a Jupyter Notebook environment. Core packages utilized included: **pandas** for data processing, **seaborn** and **matplotlib** for data visualization, and **TextBlob** for sentiment analysis

Research & Analysis Questions

The four questions explored in the study are as follows:

1. Do verified purchase reviews differ in rating compared to unverified ones?
2. Are there temporal trends in reviews — do ratings trend higher or lower over time?
3. Which brands receive the most consistent ratings?
4. Is there a correlation between the sentiment of review text and the star rating?

Key Finding/Insights

Firstly, I analyzed whether verified purchase reviews differed in rating compared to unverified ones. The results indicated that verified reviews had higher average ratings (4.81) than unverified reviews (4.58), suggesting that verified buyers tend to be more satisfied or offer more trustworthy feedback.

Second, the analysis investigated temporal trends in ratings. Ratings remained relatively consistent between 2004 and 2016, generally within the 4.8–5.0 range. However, a notable decline was observed during 2017–2018, which may reflect changing customer expectations, shifting product quality, or external factors affecting perception.

The third question examined which brands received the most consistent ratings by calculating the standard deviation of review scores. Brands such as Yardley, Fruits & Passion, and Sexy Hair exhibited the lowest rating variability, suggesting uniform customer satisfaction and potentially strong brand loyalty.

Finally, in the fourth question, the analysis explored whether sentiment in review text correlated with star ratings. A moderate positive correlation of 0.30 was identified, indicating that more positively worded reviews generally, but not always, correspond with higher star ratings.

Recommendations (if presenting to a client)

Based on these findings, several recommendations can be made. Businesses should highlight verified reviews in product listings to enhance consumer trust. Brands should closely monitor changes in rating trends over time to address potential quality issues early, to avoid drastic dips in their ratings, as seen in the dip in 2017-2018. Consistently rated brands can be positioned as reliable choices for new or hesitant buyers. Additionally, companies may benefit from incorporating sentiment scoring into their review analysis pipelines to proactively flag dissatisfaction.

Future Work - For future work, this analysis could be expanded by incorporating more recent review data to reflect current market dynamics. Specifically, UCSD released another amazon

dataset review study from 2023, but the issue with this set was the expansive amount of data—more than I could feasibly analyze, and too large to the point of damaging my computer, and maximizing storage. In future work, I would like to analyze metadata fields, such as product style, category, and pricing, as it could give a deeper insight or consumer preferences. I would also like to analyze categories such as also bought, or frequently bought together, similar items, or also view, and possibly see if users bought those recommended products, and then direct companies to use better marketing and prediction strategies. Using more advanced natural language processing models for sentiment analysis or performing topic modeling could further enhance understanding of customer feedback. These next steps would allow for a more comprehensive assessment of brand performance and customer satisfaction, and allow for potential clients to maximize profits.