Mining Online Customer Reviews: What Do Customers Mention About Cotton Jeans?

A. Rose Downs, Ph.D. Student University of Missouri, Yuhui (Jessie) Liu, Ph.D. Candidate, University of Missouri, and Dr. Li Zhao, Associate Professor University of Missouri

Keywords: data mining, topic modeling, denim jeans, and customer reviews

Introduction. Cotton jeans hold a significant place in the global fashion industry, with the denim jeans market valued at $67.72 billion in 2023 and projected to reach $95.2 billion by 2030 (Smith, 2023). Despite their popularity, consumers often encounter difficulties in selecting jeans that meet their expectations in fit, quality, and color. Moreover, the increasing consumer focus on sustainability has brought cotton products to the forefront, making jeans an ideal case study to explore sustainability issues in the textile industry. Customer reviews have emerged as a valuable resource for understanding consumer opinions and their challenges. The significance of these reviews is underscored by findings that a majority of consumers consult reviews before making purchases (Brandes et al., 2023). This study, therefore, aimed to analyze customer reviews of cotton jeans to uncover the attributes most frequently mentioned by consumers. Through this analysis, the research seeks to offer insights into consumer behavior and preferences regarding cotton jeans, providing valuable information for both academics and retailers.

Literature Review. Customer reviews provide an important avenue for research as they summarize both the positive and negative perspectives of the product and potentially the experience as well. In 2021 alone, the predicted global revenue impact from reviews was $3.8 trillion (Brandes et al., 2023). Additionally, customer reviews provide a source for potentially uncovering consumers' underlying opinions, needs, and desires (Camacho-Otero et al., 2019; Li et al., 2023; Wu et al., 2021). In order to pull the important information discussed, data mining and topic modeling through a Latent Dirichlet Allocation (LDA) algorithm (Blei et al., 2003) was chosen as the method for this study. There is limited research using online customer reviews (OCR) to understand consumer experiences with cotton jeans. Therefore, this study intended to uncover common topics and themes that emerge from online customer reviews (OCR) on cotton jeans.

**Method.** To fulfill our research objective, we initially enlisted the assistance of two college students to gather students' perspectives on denim jeans brands.

Figure 1 *Distribution of Data Among Denim Jeans Brands*

Figure 2 *Colors used by Denim Jeans Brands*

Brand	Color names counts
AG Jeans	282
American	148
Diesel	16
Everlane	31
GAP	69
H&M	28
Levi's	184
Madewell	68
Rag & Bon	117
ragnbone	44
Uniqlo	38

We identified ten of the most popular denim jeans brands among college students, including Diesel, Levi's, Rag & Bone, American Eagle, H&M,
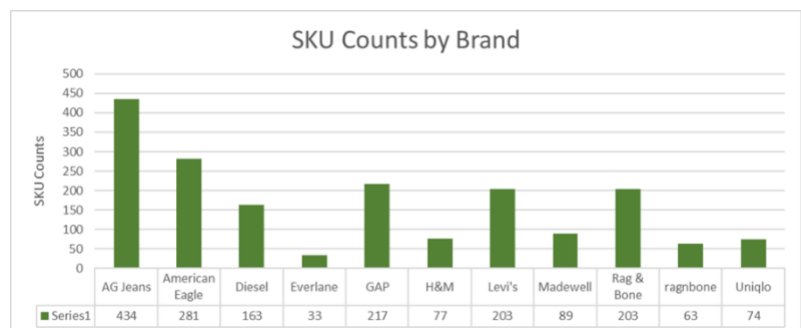
AG Jeans, GAP, Uniqlo, Madewell, and Everlane. Subsequently, the data analytic procedure was comprised of four phases: data collection, data cleaning, data analysis, and data interpretation. We collected product data and customer reviews from the official online stores of these brands, resulting in an accumulation of over 2000 SKUs (refer to Figure 1), with each brand offering products in various color options (refer to Figure 2). Following this, we proceeded to clean the data and conduct text mining on the customer reviews. During this phase, we employed LDA topic modeling analysis, which involved tasks such as removing numbers and non-meaningful characters, performing lemmatization, and filtering out stop words using the NLTK library in Python. We then applied the LDA algorithm to identify latent topics within the customer review dataset.

Results. In Table 1 below, there are five example topics that emerged from analyzing the online customer reviews (OCR). Each of these topics focus on a different aspect of denim jeans. For instance, the first topic centers around the perfect pair of jeans. Within the reviews, customers mention the fit, comfort, and color as important aspects in their denim jeans. With the second topic, the customer reviews mention the quality and how the jean feel and look. In comparison, the third topic that surfaced focuses more on the negative themes such as issues with size and returning the jeans. The fourth topic that appeared uncovered that the reviewers talk about the specific details of the jeans such as the hem or hip. Lastly, the fifth topic brought to light that customer reviews center around the denim material itself such as whether it is light, dark, or warm. The price is a relevant theme within this fifth topic as well. These results align with our desire to research denim jeans specifically as they are favored product but there are varying perceptions regarding size issues, color, quality, and material. Each topic highlights the importance of these characteristics of a pair of jeans to customers.

Table 1 *Topics and Keywords from Reviews*

Topic	Key Words
Perfect Jean	fit, jean, low, wear, like, color, perfect, comfortable, buy, pair
Style & Quality	great, look, pant, good, quality, nice, well, soft, make, feel
Size Issues	size, would, order, big, small, return, run, large, one, ever
Details	think, even, though, hem, bra, button, wide, take, hip, show
Material	light, white, price, ankle, blue, sleeve, heavy, dark, lot, warm

Conclusion. Overall, this type of research study provides value to research and the industry because of the importance of cotton and denim, jeans, and customer reviews. By analyzing customer reviews for the ten most popular jeans brands for college students, this uncovered their experience, both positive and negative, as well as their preferences and potential reasons for returning jeans. This information is important as it provides businesses insight into what potentially drives customers to purchase and then to return. This research has a scholarly impact in that a similar method can be used to analyze other product categories and compare markets.

References

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.

Brandes, L., Godes, D., & Mayzlin, D. (2023, March 23). *Research: The Pros and Cons of Soliciting Customer Reviews.* Harvard Business Review. https://hbr.org/2023/03/research-the-pros-and-cons-of-soliciting-customer-reviews

Camacho-Otero, J., Boks, C., & Pettersen, I. N. (2019). User acceptance and adoption of circular offerings in the fashion sector: Insights from user-generated online reviews. *Journal of cleaner production, 231*, 928-939. https://doi.org/10.1016/j.jclepro.2019.05.162

Li, M., Zhao, L., & Srinivas, S. (2023). It is about inclusion! Mining online reviews to understand the needs of adaptive clothing customers. *International Journal of Consumer Studies*, *47*(3), 1157–1172. https://doi.org/10.1111/ijcs.12895

Smith, P. (2023, August 29). *Value of the denim jeans market worldwide from 2022 to 2030 (in billion U.S. dollars).* Statista. https://www.statista.com/statistics/734419/global-denim-jeans-market-retail-sales-value/

Wu, R., Wu, H. H., & Wang, C. L. (2021). Why is a picture 'worth a thousand words'? Pictures as information in perceived helpfulness of online reviews. *International Journal of Consumer Studies, 45*(3), 364–378.