
Amaan Ul Haq Siddiqui

[linkedin.com/in/amaanulhaqsiddiqui](https://www.linkedin.com/in/amaanulhaqsiddiqui)

github.com/amaan-igs

Amazon Sales Analysis

31st August 2024



Introduction

This report presents an analysis of Amazon product data, stored in AWS S3, preprocessed using PySpark on AWS EMR, and further analyzed with Pandas. The goal is to derive insights into product pricing, ratings, manufacturers, and customer reviews to address business challenges related to pricing strategies, customer engagement, and product popularity.

Overview

Source: Amazon product data stored in AWS S3.

Tools Used:

- **Preprocessing:** PySpark on AWS EMR.
- **Analysis and Visualization:** Pandas, Plotly, Seaborn, and Matplotlib.

Problem Statement

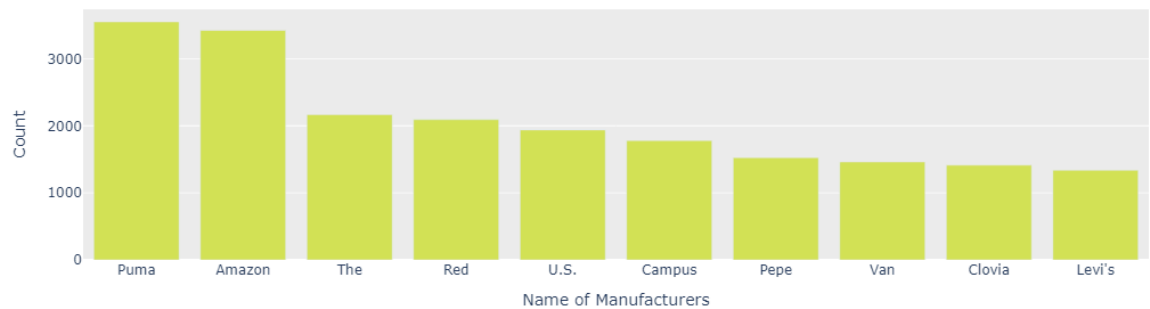
The primary business problem addressed in this analysis is the identification of optimal pricing strategies and understanding the impact of pricing and discounting on customer engagement, as reflected by ratings and reviews. The goal is to help Amazon optimize pricing to maximize sales and customer satisfaction.

Key Insights

Popular Manufacturers:

- **Puma** emerged as the most popular brand by count, followed by **Amazon**. However, Amazon holds the highest average rating among popular manufacturers.

Popular Manufacturers Category



Top Categories:

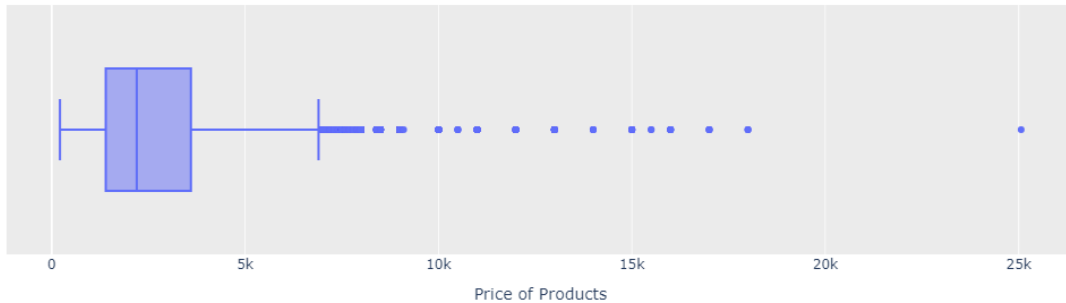
- The most popular main category is "Men's Clothing," with significant representation among the top-selling products.

main_category	
accessories	1279
men's clothing	6965
men's shoes	4054
stores	2571
women's clothing	3572

Pricing Analysis:

- The average price of popular products is skewed towards the lower end, with most products priced under ₹3600.

Price Data Distribution

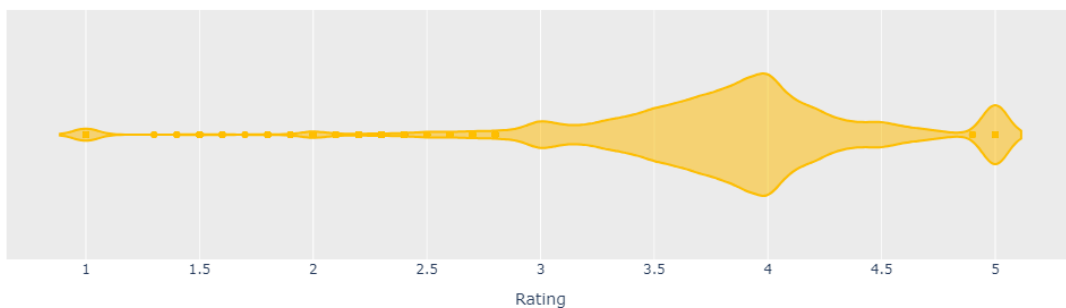


- **Outlier Pricing:** About 5.32% of products are outliers in pricing, with brands like Puma and Amazon showing higher outlier prices in specific subcategories.

Customer Engagement:

- Ratings are predominantly 4 or 5 stars for popular products, indicating high customer satisfaction in these categories.

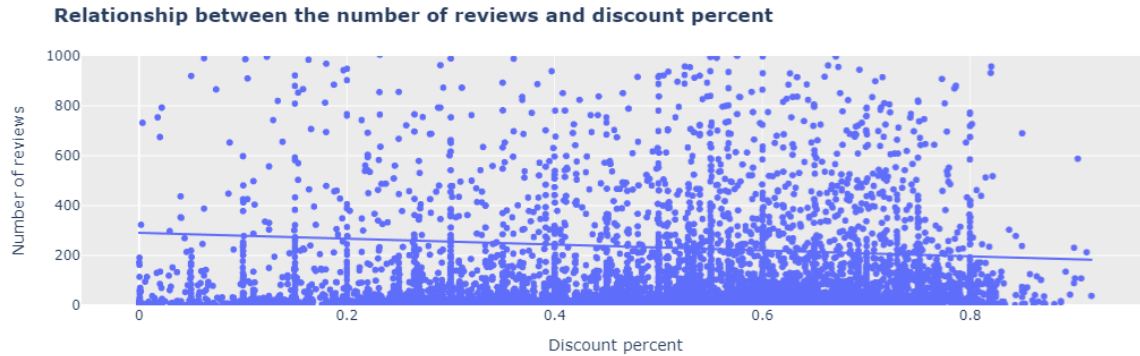
Rating Distribution of the Popular Products



- Most products received between 0-49 reviews, with a steep drop-off after 1,000 reviews, suggesting possible fake reviews for some outlier products.

Discounting Patterns:

- The distribution of discount percentages across products shows a right-skewed distribution, with spikes at multiples of ten, indicating strategic discounting by sellers.



- A slight decrease in the number of reviews was observed with increasing discount percentages, suggesting that heavy discounts might be perceived as a reduction in product value.

Business Implications

Pricing Strategy: Products priced under ₹3600 are more likely to be popular, especially from well-known brands like Puma and Amazon. Amazon should consider pricing products within this range to maximize sales while monitoring the performance of outlier-priced items.

Discounting: While discounting is a common strategy, extreme discounts may negatively impact the perceived value of a product. A balanced approach, with discounts that do not exceed certain thresholds, could help maintain product value while still driving sales.

Customer Engagement: High ratings correlate with popular products, indicating that quality and customer satisfaction are critical. Encouraging authentic customer reviews, especially for high-end products, could enhance trust and drive further sales.

Conclusion

This analysis uncovers critical patterns in Amazon's product pricing, customer preferences, and engagement dynamics. By applying these insights, Amazon can refine its pricing algorithms, strengthen its competitive positioning, and foster deeper customer loyalty. Exploring additional layers of data, such as seasonal trends and demographic influences, could further enrich Amazon's ability to anticipate market shifts and adapt its strategies for sustained growth.

Architectural Diagram

