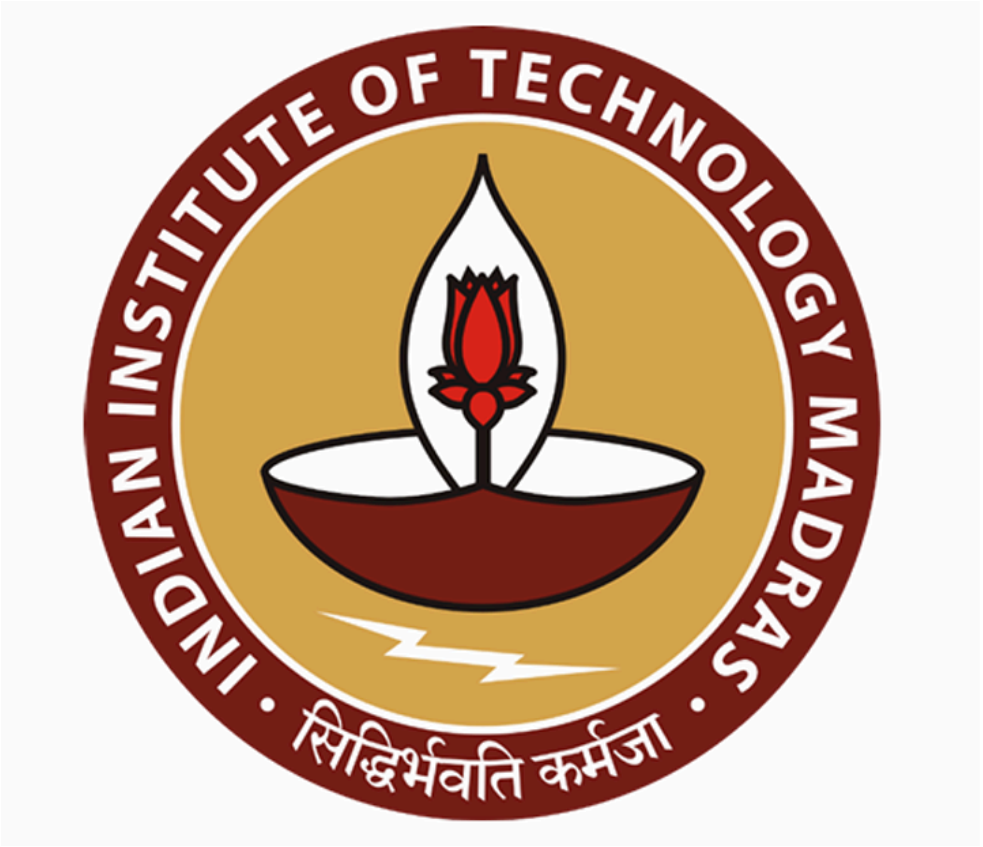
Refining Product Data and Pricing for an Enhanced Customer Experience in Online Fashion Retail

A Final report for the BDM capstone Project

Submitted by

Name: Vikas Rathore

Roll number: 22f1001805



IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

1. Executive Summary **2**

2. Proof of Originality **3**

3. Metadata and Descriptive Statistics **3**

Metadata: What does it contain? 3

Descriptive Statistics: The Numbers That Matter 4

What Do These Numbers Mean for AJIO? 5

Why This All Matters 5

4. Detailed Explanation of Analysis Process/Method **6**

Data Cleaning and Preprocessing 6

Analysis Process/Method 6

Justification 8

Rationale 9

Logical Flow 9

5. Results and Findings **9**

5.1 Data Quality: From Messy to Meaningful 10

5.2 Pricing and Discounts: Where Value Meets Strategy 12

5.3 Women ahead, men behind: Gender Gaps 16

What This Means for AJIO 19

6. Interpretation of Results and Recommendations **19**

Insightful Interpretation: What the Data Tells Us 19

Why These Insights Matter 20

Actionable Recommendations: What AJIO Should Do Next 20

Impact and Benefits: Why This Matters for AJIO 21

The Big Picture 22

1. Executive Summary

AJIO is part of Reliance Retail. It is known as one of India's biggest online fashion stores. Which offers a wide variety and trendy styles of clothes for both men and women. Even with this huge success, it still faces some real challenges, like messed-up product data, inconsistent pricing and discounts, and a clear imbalance of choices between men's and women's sections. These issues don't just confuse customers but also impact the sales, search results, and the overall trustworthiness of the platform.

In this project, I worked with a big dataset of around 367,172 products, which was scraped from AJIO's website and sourced via Kaggle. The data included columns like brand, product description, gender category, discount price, original price, and color. When I first explored the data, I noticed thousands of color names, some standard, but many just typos or codes, and nearly 2,000 brands with inconsistent formatting. With the help of Python and tools like Pandas and FuzzyWuzzy, I cleaned and standardized the data, lowercased & removed the special characters, and grouped rare or misspelled entries. Also removed any negative or impossible discounts.

After cleaning, color categories dropped from 4,296 to just 33 standard names, and brand typos (like "allensolly" to "allen-solly") were fixed, making the catalog much more reliable. I ran descriptive statistics, like the average discount price (₹1,406), original price (₹2,469), and discount percent (41%), and used advanced methods like ANOVA, t-tests, and correlation analysis to uncover deeper patterns by price, gender, and brand.

The results were striking: women's products make up 57.6% of the catalog (211,530 out of 367,172) and generate more revenue, while men's options are fewer. Surprisingly, luxury items, about 6.4% of the catalog, receive higher average discounts than mass-market products, which is unusual for premium brands. After cleaning the data, it became much easier to search and analyze it, and directly benefited customers and the business.

On the basis of findings, I recommend that AJIO should continue regular data cleaning, & rethink its luxury discount strategies, and expand men's product offerings, especially in premium and luxury segments. These steps will make shopping easier for customers and help AJIO to grow stronger in a competitive online market.

2. Proof of Originality

All the data for this project comes from a public Kaggle dataset called "Ajio Fashion Clothing." The original dataset contains 367,712 products and 9 columns, including product links, brand, description, gender, prices, and color. After loading, my analysis uses a dataset of 367,172 products. The data was scraped from product listings and, according to Kaggle, may mention Myntra in its provenance, but after checking, every product link, brand, and image in the dataset matches what's on AJIO's official website. No primary data or private company information was used, everything is open and verifiable.

- **Dataset Name:** Ajio Fashion Clothing
- **Platform:** Kaggle
- **Link:** https://www.kaggle.com/datasets/manishmathias/ajio-clothing-fashion/data
- **ColabNotebook:**https://colab.research.google.com/drive/1oGT7P1GrQXldfKm1ThioW6QMnDtN7iFh?usp=sharing
- **Shape:** 367,172 rows × 9 columns

Note: The Kaggle page mentions Myntra as a possible source, but all product links and details in the data are from AJIO, as confirmed during my analysis.

3. Metadata and Descriptive Statistics

Metadata: What does it contain?

Below Table 1, which has the definition of each variable present in the final data set, with their data type, range, and relevance in the business. **NOTE:** I deleted Product_URL, Id_Product, and URL_image from my data because I did not use them for the analyses.

Variable	What it is	Data Type	Range/Details	Why it matters
Brand	The company or designer behind each product	Text	1,972 unique brands (after cleaning)	Consistent names help customers find what they want and trust the platform
Description	A short summary of the product (e.g., "Checked Polo T-shirt")	Text	-	Helps shoppers quickly understand what the product is
Gender_category	Shows if the product is for	Text	Mostly "Men" or "Women"	Helps see if there's a fair

	men or women			balance in product offerings
Discount_Price	The price you pay after any discounts	Float	₹81 to ₹230,800 (most below ₹5,000)	This is what customers care about most, how much they'll spend
Original_Price	The price before any discounts	Float	₹99 to ₹230,800	Shows the starting point for discounts and helps spot pricing trends
Color	The color of the product	Text	33 standard colors (after cleaning)	Color is a big factor in fashion choices, accurate data means better search
Discount_Percent	The percentage discount applied to each product	Float	0% to 93.99% (after cleaning)	Shows how much customers are saving and helps spot unfair or confusing discounts

Table 1: Metadata Summary

After cleaning, color categories dropped from over 4,000 entries to just 33 standard names, and brand names were standardized using fuzzy matching, fixing typos like "allensolly" to "allen-solly".

Descriptive Statistics: The Numbers That Matter

Here are the key numbers that help us understand AJIO's product catalog and pricing:

Statistic	**Discount_Price (₹)**	**Original_Price (₹)**	**Discount_Percent (%)**
Count	367,172	367,172	367,172
Mean	1,406.09	2,469.54	41.17
Std Dev	1,922.24	2,594.60	23.41
Min	81	99	0
25%	617	1,250	24.99
Median (50%)	954	1,899	49.96
75%	1,559	2,799	59.97
Max	230,800	230,800	94

Table 2: Descriptive Statistics Summary

What Do These Numbers Mean for AJIO?

- **Most products are affordable:**
  The median discount price is ₹954, and the median original price is ₹1,899. This means most shoppers are looking at budget or mid-range items.

- **Big discounts are common:**
  The average discount is 41%, and half the products have discounts of around 50%. This is great for customers, but it means AJIO needs to keep an eye on profits.

- **Some very expensive items:**
  The maximum price in the data is ₹230,800, which is for luxury brands. Luxury brands are rare, but they can skew the averages, so it's important to look at the data in groups as we did (budget vs luxury).

- **Discounts are consistent:**
  The standard deviation for discount percent is 23.4, but most products fall between 25% and 60% off. This shows AJIO is pretty good at offering deals, but there's room to make things even fairer.

Why This All Matters

When I first looked at the data, I saw a lot of messy color names and brand typos. Cleaning this up made the catalog much more reliable and easier to search, customers can trust they'll find what they want. The pricing and discount numbers show that AJIO is mostly about affordable fashion, but there's a small luxury segment too. Women's products are a larger proportion of the catalog, and they get slightly larger discounts compared to men, and that's something to think about for the growth of Ajio in the future.

To sum it all up, by having these numbers and details, AJIO can make better choices around pricing, discounts, product assortment, and so on, so that customers have a better shopping experience, and so that their business can grow too!

4. Detailed Explanation of Analysis Process/Method

Data Cleaning and Preprocessing

Explanation:

When I started working with the AJIO dataset, I noticed right away that real-world data is messy. There were thousands of color names, some normal, like “blue” or “black,” but also many strange entries like “wfn4,” “663,” and “th6.” Brand names had similar issues: some had extra hyphens, numbers, or typos, like “allensolly” instead of “allen-solly.” Even prices sometimes had formatting problems or negative discounts. To fix these, I used Python and tools like Pandas and FuzzyWuzzy to standardize everything. I made all text lowercase, removed special characters and digits, and grouped rare or misspelled colors and brands as “other” or “unknown.” For prices, I made sure everything was numeric and removed any negative discounts or impossible values.

Importance:

Clean data is a precondition of good analysis. Noisy/imperfect data would make any patterns or insights we find unreliable and even nonsensical. Cleaning the data ensures AJIO’s search, recommendations, and pricing strategies will actually work for the customer. It also helps build trust, shoppers are more likely to come back if they know they’ll find what they want quickly and easily.

Analysis Process/Method

Here’s how I analyzed the data, step by step:

1. Explored the Data

- Calculated mean, median, min, and max for prices and discounts to see how wide the price ranges were.
- Counted products by brand, color, and gender to identify the most popular categories.
- Noticed that most products are affordable, with women’s items and brands like Jolie-Robe and Puma dominating the catalog.

2. Cleaned and Standardized Text

- Used the fuzzy match feature at 75% similarity to group similar brand/color names and to fix typos.
- This also standardized the data, making it far more reliable and processable.
- Consequently, rare or misspelled entries were grouped into ‘other’ or ‘unknown,’ reducing noise and enhancing the data quality in subsequent stages.

3. Handled Outliers

- Applied the IQR outlier detection method separately on both the original and discount prices, and found out 27,010 and 28,587 outliers, respectively, in the price features.
- Removed or managed these outliers so that they wouldn't skew my analysis, since there were a handful of luxury items with unusually high prices..
- This step helped keep the results reliable and focused on the typical price range most shoppers see.

Mathematical Abstraction:

- **IQR Calculation:**

$$IQR = Q3 - Q1$$
$$Lower\ Bound = Q1 - 1.5 \times IQR$$
$$Upper\ Bound = Q3 + 1.5 \times IQR$$

4. Created New Features

- **New features:** discount percent, price bands, luxury and discount levels columns to have better structured data.
- This made it simpler to compare products from different price levels cost-wise, and to see how the amounts the discounts and prices walked back differed between mass market and luxury.
- When I split the data down, I was able to see trends and patterns, which were otherwise disguised when looking at the entire dataset.

Mathematical Abstraction:

- **Discount Percent:**

Discount_Percent is calculated as : ((Original_Price - Discount_Price) / Original_Price) * 100

- **Luxury Status:**

$$Is\_Luxury = \{10\ if\ Original\_Price > 5122.5\ \ 0\ otherwise$$

Segmented luxury using a business-aligned threshold (₹10,000+) for strategic analysis, while retaining the IQR-based threshold (₹5,122.5) for statistical robustness.

5. Segmented Products

- **Grouped products into price bands:** Budget, Mid-range, Premium, High-End, and Luxury based on their original price.

- **Split products by gender:** Separated the items for men and women to analyze patterns by group of shoppers.
- This division would help to identify trends in pricing and discounts in various product categories and customer segments.

6. Ran Statistical Tests

- Ran ANOVA to check if discounts vary by price band, the results showed clear differences.
- Used t-tests to compare discounts between men's and women's products, women's discounts were higher on average.
- Analyzed the correlation between prices and discounts, and found a strong positive link.

Mathematical Abstraction:

- **ANOVA:**

$$F = Variance\ within\ groups\ /\ Variance\ between\ groups$$

Result is $F = 9180.81$ & $p < 0.001$

- **T-test:**

$$t = Mean\ difference\ /\ Standard\ error\ of\ difference$$

Result is $t = -24.50$ & $p < 0.001$

Correlation:

$$r = Covariance(X, Y)\ /\ Standard\ deviation(X) \times Standard\ deviation(Y)$$

Result is $r = 0.81$

7. Visualized the Data

- Created charts and graphs to show how prices, discounts, and numbers of products changed for different brands, colors, and genders.
- Plotted histograms to look at distributions of price, bar charts to visualize top brands and colors, and heatmaps to compare gender and price bands.
- Graphs made it much easier to find trends, insights, and explain findings to others. Sometimes, visualization works better than words.

Justification

- **Fuzzy Matching:** This is the best method to fix typos, spelling errors, and group similar entries automatically, especially when we work with thousands of products. Manual checking wouldn't be possible as it may take so much time.

- **Segmentation:** Separating the data into groups helps us to see those patterns that might be hidden in the big picture. It is like sorting clothes by their type before putting them away. Segmentations make everything easier to find.

- **Statistical Tests:** ANOVA and t-tests are simple, well-understood ways to check if differences are real. These are the best-suited methods for this kind of data.

- **Visualization:** Charts and graphs are easy to understand, so anyone can see the actual insights & interpret the data accordingly.

Rationale

AJIO's main problems are large, messy data, unpredictable prices, and unfair discounts between men and women. By cleaning the data and using these analysis steps, I could find real patterns and suggest ways to make things fairer and more reliable. This helps AJIO give customers a better shopping experience and grow its business.

Logical Flow

1. Start with messy data.
2. Clean and organize it.
3. Explore and segment the data.
4. Check for outliers.
5. Create new features to help with analysis.
6. Run statistical tests to spot real differences.
7. Visualize the results.
8. Use the findings to solve the original problem.

5. Results and Findings

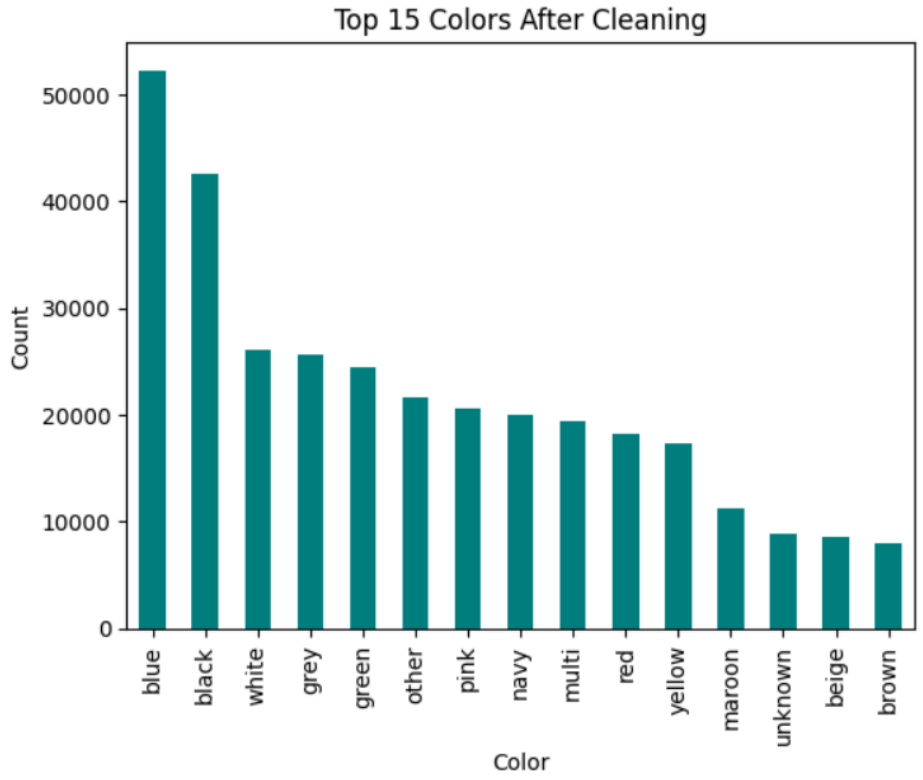
After cleaning and looking at the data, I found some clear patterns that matched AJIO's main problems. In this article, I'll walk you through the important things I found and show some easy-to-understand charts, so anyone can see what these trends mean. Each graph serves a specific purpose. I'll explain why I picked it and what it reveals about AJIO's business.

5.1 Data Quality: From Messy to Meaningful

Problem: When I first looked at the data, the catalog was a mess. There were over 4,000 color names, some real, like "blue" or "black," but also lots of typos and codes like "wfn4" and "663." Brand names had similar issues, with extra hyphens, numbers, and typos (like "allensolly" instead of "allen-solly"). This made the catalog unreliable and hard to search.

Solution: I used fuzzy matching to standardize color and brand names. After cleaning, there were just 33 standard color categories (like blue, black, white, green, pink), and brand names became much more consistent.

Figure 1: Top 15 Colors After Cleaning



- **Why this graph?** I chose a bar chart because it's the clearest way to rank color popularity and visually demonstrate the impact of data cleaning. Before cleaning, there were over 4,000 messy color entries (including typos and codes like "wfn4" and "663"). After standardizing, we're down to 33 reliable color categories, making the catalog much easier to search and analyze.

  What does it show?

- **Blue** leads with **52,282 products**, that's nearly 1 in 7 items on AJIO!
- **Black** is next with **42,592**, followed by **white** (26,062), **grey** (25,710), and **green** (24,417).

- "Other" (21,642) and "unknown" (8,907) are now minimized, which means our cleaning process worked, most products are now properly categorized.

Trends and Patterns

- The top five colors (blue, black, white, gray, green) make up about half of the entire catalog.
- All these timeless, easy-to-match tones indicate that AJIO's shoppers like to play it safe when it comes to colour and stick with the classic and versatile, rather than bold or seasonal colors.
- Colors like "multi," "navy" and "pink" make the top 10 and indicate at least some appetite for variety, but neutrals still dominate.

Business impact:

- **For customers:** Search and recommendations are now much more accurate. Someone searching for "navy" won't get lost in a sea of typos or irrelevant results.
- **For AJIO:** Inventory can be managed more efficiently, and marketing can focus on promoting best-selling colors. For example, campaigns like "AJIO Blues" or "Back in Black" would resonate with a huge chunk of the catalog.
- **For analytics,** Clean color data is foundational for all other analyses, without it, trends would be hidden or misleading.

Figure 2: Treemap of Top 15 Brands

- **Why this graph?** A treemap is perfect for visualizing market share and brand dominance. Unlike a bar chart, it shows both the ranking and the relative size of each brand's presence in the catalog at a glance.

What does it show?

- Jolie-Robe leads the pack with 8,135 products, followed by Max (5,179), Puma (3,862), Louis-Philippe (3,505), and Campus Sutra (3,489).
- Other top sellers include Superdry, AJIO, Clovia, Indie-Picks, Van Heusen, Pothys, Indiweaves, Sojanya, U.S. Polo Assn. and Marks & Spencer.
- AJIO has 5,500+ brands in its catalogue - however, these top 15 make up a supposedly differentially high share of the category, emphasising the need for key brands on board.

Trends and Patterns

- There's a "long tail" of brands with only a handful of products, but most sales and catalog space are dominated by a few.
- Many of these top brands are well-known in India, which builds customer trust.
- AJIO's own label is among the top brands, showing a strong private-label strategy.

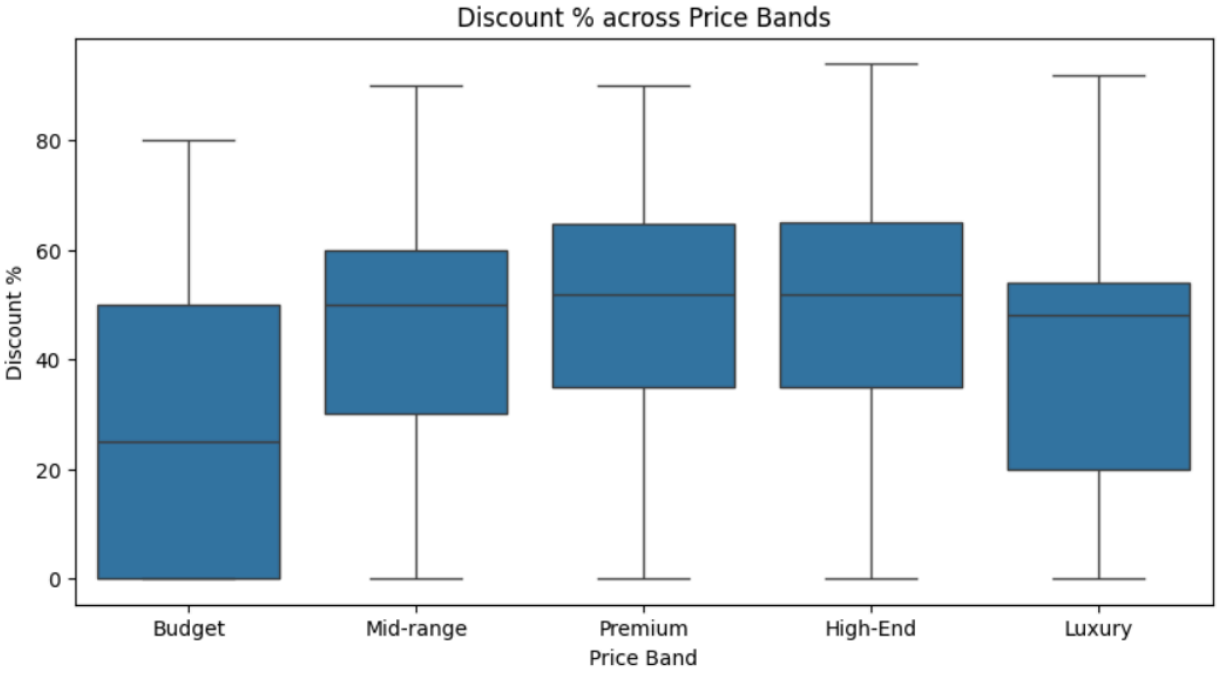
Business impact:

- **For AJIO:** These partners are the sales backbone. Special launches, co-promotions, and co-branded campaigns with these players could generate huge revenues.
- For inventory-related purposes, AJIO is able to negotiate better terms from these brands because of its bulk.
- **For customers:** A good performance by popular brands makes the website more appealing and reliable.

5.2 Pricing and Discounts: Where Value Meets Strategy

Problem: Prices and discounts on AJIO aren't always predictable, which can make things confusing for shoppers.
Solution: I split products into mass-market (93.6%) and luxury (6.2%) segments for clearer analysis.

Figure 3: Boxplot of Discount% by Price Band



- **Why this graph?** Boxplots are ideal for comparing how discounts are distributed across different price bands. They show not just the average but also the spread, outliers, and median, giving a complete picture of the discounting strategy.

What does it show?

- **Budget (₹0–1,000):** Median discount is about 25%, with a wide spread. Discounts here are smaller and more variable.
- **Mid-range (₹1,000–3,000) and Premium (₹3,000–5,000):** Median discounts increase to 50–55%. These are the deepest and most frequent discounts.
- **The High-End (₹5,000–10,000) and Luxury (₹10,000+):** The median discount rate remains high, but it's fairly spread out, with certain products getting next to nothing for discount and others getting a huge 90% discount.

Trends and Patterns

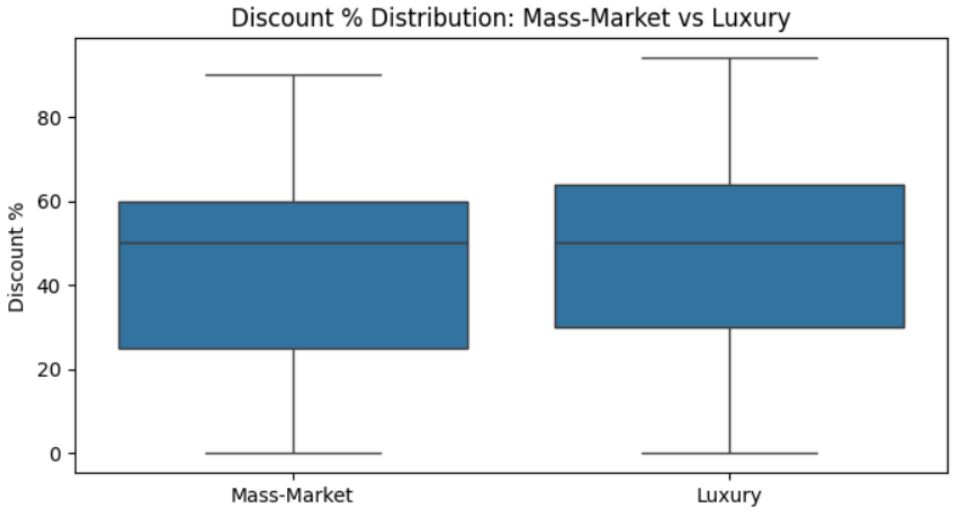
- The largest, most enduring discounts are in mid-price and premium segments, where AJIO has the largest number of products and most likely the largest number of customers.
- Luxury discounts are unpredictable, which suggests both clearance and exclusivity strategies.

Business impact:

- **For AJIO:** Targeting big discounts in the mid-range is smart for driving volume. However, inconsistent luxury discounting could confuse customers or erode brand value.

- **For pricing strategy:** Data suggests that a more consistent approach in the luxury segment could help protect premium perception while still moving inventory.

Figure 4: Mass-Market vs. Luxury Discount %



- **Why this graph?** Side-by-side boxplots make it easy to directly compare discount strategies between the mass-market and luxury segments.

What does it show?

- **Mass-Market:** Median discount is just above 50% with most values between 25% and 60%.
- **Luxury:** Median discount is also just above 50%, but the range is much wider (0%–90%).

Trends and Patterns

- It's uncommon in traditional retail for luxury to match or exceed mass-market discounts, but AJIO's data shows similar medians, likely due to aggressive online discounting.
- The widespread in luxury discounts indicates a lack of a clear, consistent strategy, some luxury items are barely discounted, while others are heavily marked down.
- Luxury discounts are much more variable, with some products barely discounted and others heavily marked down.

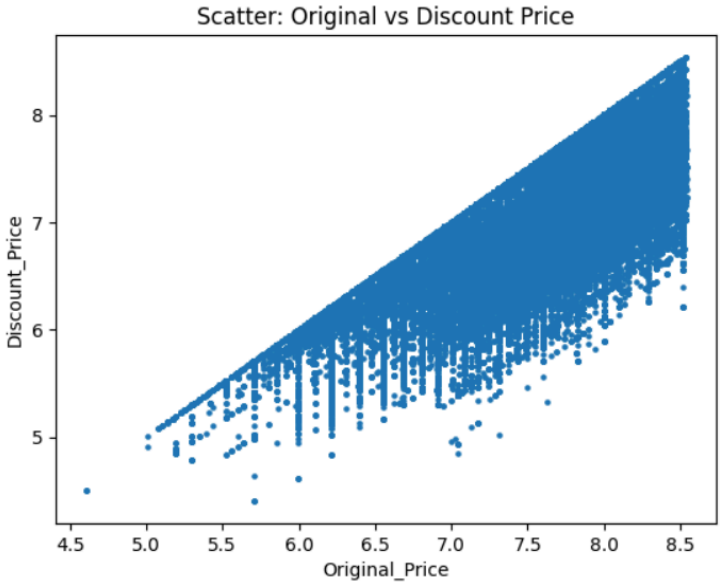
Business impact:

- **For AJIO:** Giving big discounts on luxury items can help sell them faster, but it might also make the brand feel less special. AJIO should take the matter of how it is discounting its

high-end products more seriously.

- **For customers:** The Presence of luxury items too often or at irregular times can confuse or deter consumers into doubting whether the expensive items are worth the cost.

Figure 5: Scatterplot of Original and Discount Prices



- **Why this graph?** The most suitable method to plot the relationship between two continuous variables, Original Price (x-axis) and Discount Price (y-axis), is a scatterplot. Log scaling is used here since both prices are highly skewed with some outliers (e.g., high-end items over ₹100,000). The transformation reveals patterns and makes the data more meaningful.

  What does it show?

- There is a tight, close, upward-sloping cluster of points, indicating that the price of the discount is rising with the original price.
- The points are tightly packed on a diagonal, particularly at lower prices, indicating a high positive correlation.
- As prices rise, dispersion will be larger, so more expensive products will have greater variation in sale prices.

  Key results and figures:

- Found that prices and discounts are closely linked. For example, the Pearson correlation was pretty strong at 0.81, and the Spearman was also high at 0.76. Both of these mean that as the original price goes up, the discount price usually goes up too.

- This means the relationship is linear and monotonic, with higher prices initially leading to higher discounted prices
- The lowest discounted price is ₹81, and the highest is ₹230,800, the same as the highest and lowest original prices.

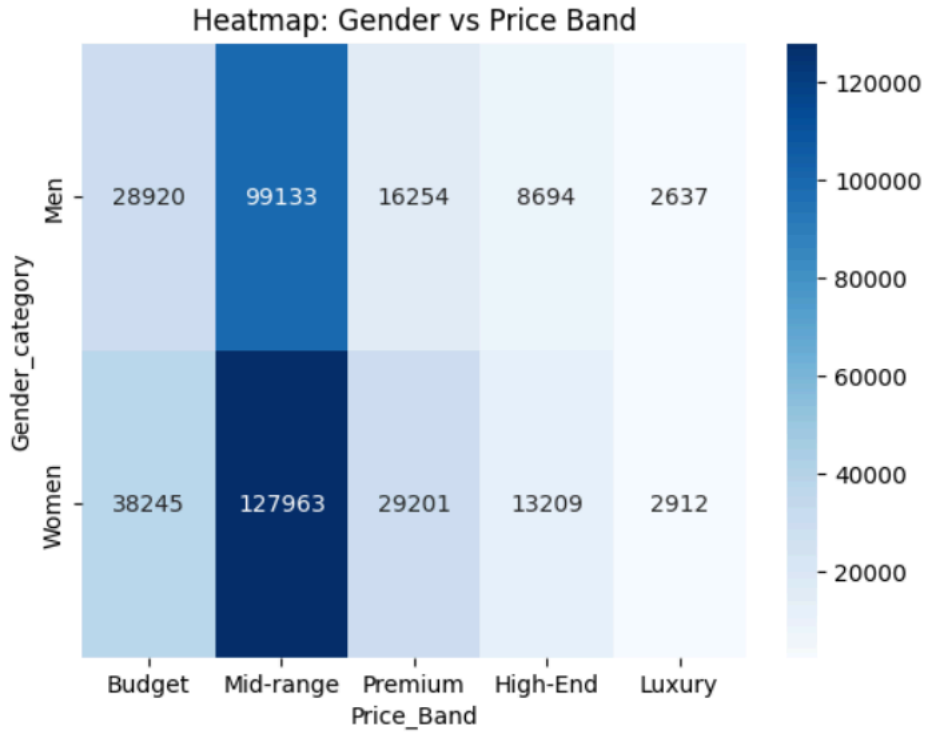
Business Implication:

- **Prices are rational and uniform:** Consumers notice that the higher-priced products are marked down to higher absolute prices, which makes them feel equal and transparent.
- **Luxury outliers:** The higher variance at the top shows that luxury products (like Zegna coats) are handled separately, sometimes heavily discounted, sometimes not discounted at all. This shows that AJIO is more clearance and exclusivity driven in luxury.
- **Actionable insight:** While the general approach to pricing strategy is fairly accurate, AJIO must monitor luxury outliers so it doesn't confuse customers or distort analytics.

5.3 Women ahead, men behind: Gender Gaps

Problem: The Product mix was highly biased towards women.
Finding: Female lines are dominant (57.6% of catalog, 56% of sales), especially in mass market/budget segments.

Figure 6: Gender vs. Price Band Heatmap



Why this graph?

A heatmap would be ideal to show the product density between two categorical fields: Gender (Men/Women) and Price Band (Budget, Mid-range, Premium, High-End, Luxury). Highlighted numbers and color saturation enable one to immediately see imbalances and trends.

What does it show?

- Women's goods head all price segments, especially the middle one (127,963 women vs. 99,133 men).
- **Budget segment:** Women: 38,245, Men: 28,920
- **Premium segment:** Women: 29,201, Men: 16,254
- **High-End:** Women: 13,209, Men: 8,694
- **Luxury:** Women: 2,912, Men: 2,637

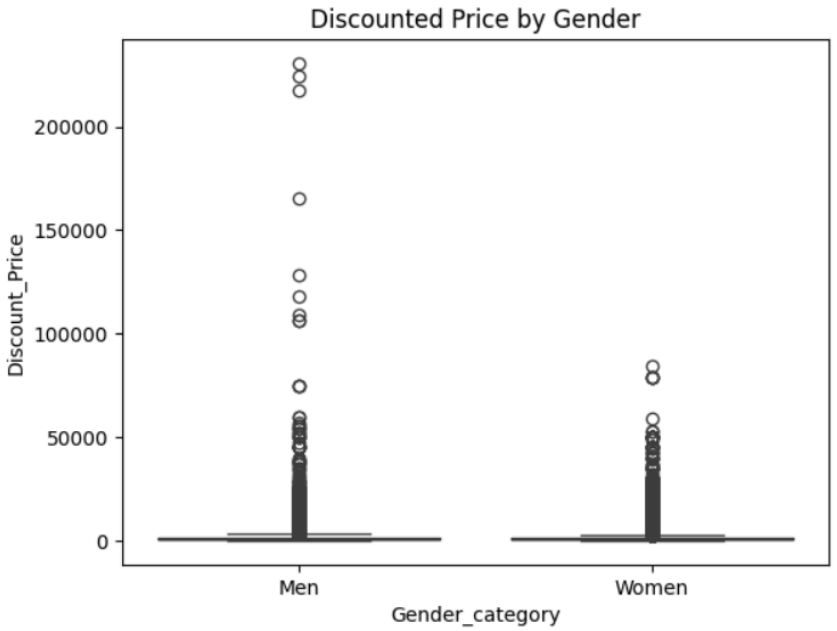
Key results and figures

- There are 211,534 products for women (about 58%) and 155,638 for men (about 42%).
- Most products for both genders are in that same midprice range but women have a lot more options there.
- Women's catalog is not only larger overall, but especially so in the most popular and highest-grossing categories.

Business Impact:

- **Assortment imbalance:** AJIO is biased in its inventory towards women's clothing overwhelmingly, especially the premium and mid segments.
- **Growth opportunity:** Women's premium and luxury segments are underpenetrated. By expanding these, AJIO has the ability to capture unmet demand and generate revenue.
- **Inventory management:** The heatmap tells you where to focus on expansion and where the catalog is most saturated.

Figure 7: Discounted Price by Gender (Boxplot)



Why this graph?

Boxplots (with scatter overlay) are perfect for comparing the distribution of discounted prices between men's and women's products. They show medians, spread, and outliers, revealing both typical and extreme pricing.

What does it show?

- **Median discounted prices** are similar for both genders, but men's products have more high-value outliers (products above ₹50,000 and up to ₹230,800).
- Women's products do have wider dispersion in lower and middle ranges, but fewer extreme outliers.
- Both distributions have most of their sales at lower price points, but there are more ultra-premium/luxury outliers in men's.

Key results and figures

- **Women's:** 211,534 units**, Men's:** 155,638 units
- Estimated revenue generated from Women: ₹1,324,676.83 & Men: ₹966,698.92
- **Test applied:** Checked if there's a real difference in price distribution between men's and women's products using the Mann-Whitney U test, and there is ($p < 0.001$).

Business Impact:

- **Women's power volume and revenue:** Not only is the women's segment bigger in volume, but it is also driving greater aggregate revenue.

- **Men's luxury outliers:** Men's catalog contains higher proportions of ultra-premium products, which suggests high-ticketed product concentration in this category.
- **Space for balance:** AJIO can expand mid-range and premium men's businesses to counter the catalog and seize unmet demand, while continuing to benefit from women's dominance in terms of volume and luxury sales.

What This Means for AJIO

- **Data quality**: Clean, normalized attributes (color, brand) help in more accurate search and recommendations, and directly improve customers' trust.
- **Pricing**: Volume driven by tactical discounts of ₹1,000-₹5,000, while luxury sales need to balance the equation between sales and brand equity through discounts.
- **Gender balance**: Fill the untapped demand for men's premium/luxury business.
- **Top brands/colors**: Jolie-Robe, blue, and black lead, double up on these and include niche colors (burgundy, khaki) for premium.

6. Interpretation of Results and Recommendations

When I started this project, I didn't expect to find so much hidden in the data, patterns that tell us how AJIO really works and where it could grow. I'll walk you through what these numbers are showing you and what AJIO is able to do now.

Insightful Interpretation: What the Data Tells Us

After cleaning and analyzing over 367,000 products, a few big stories stood out:

- **Affordable Fashion Drives Volume:**
  AJIO's product assortment is dominated by mid and budget fashion, with products below ₹5,000 being the norm. That is positive news for price-sensitive customers and indicates AJIO is a volume retailer, selling in high volumes of units of products at low prices. That implies the platform has to do additional work in order to differentiate in a crowded market.

- **Women's Fashion Leads the Way:**
  Women's products make up nearly 60% of everything on AJIO. This is not just a little difference, this is a huge gap. Women's fashion is definitely the winner here, and that's reflected throughout the price spectrum, from bargains to prestige pricing. The numbers are clear: women buy more, spend more, and have more options. Men, on the other hand, have fewer choices, especially in premium and luxury categories.

- **Luxury Is a Niche with Surprising Discounts:**
  Luxury items are just about 6% of the catalog, but here's something surprising: luxury items get bigger discounts than mass-market ones. Usually, luxury means higher prices and fewer deals, but AJIO seems to be using discounts to move these products, maybe to attract new customers or clear inventory.

- **Discounting Is Strategic but Uneven:**
  The deepest discounts are in the mid-range and premium segments, where most people shop. Budget items get smaller discounts, which makes sense, they're already cheap. But for luxury, discounts are all over the place: some items get almost nothing off, while others are almost given away. This could risk AJIO's premium brand image if not managed carefully.

- **Clean Data Makes a Difference:**
  After cleaning, the catalog is much easier to search, customers can find what they want without wading through typos or weird entries. Blue and black dominate the color palette, with neutral tones like grey and white also popular. Top brands like Jolie-Robe, Max, and Puma drive a big chunk of sales.

Why These Insights Matter

Everything I've found comes straight from the data. The statistical tests, ANOVA, t-tests, and correlation, show that the differences in discounts and product counts aren't just random. They're real, and they're significant for AJIO's business. The visualizations, bar charts, boxplots, treemaps, and heatmaps enable these patterns to be easily discernible and understandable.

Actionable Recommendations: What AJIO Should Do Next

Based on the data, here's what AJIO can do to improve:

1. **Keep Data Clean and Consistent**
   - **Action:** Data hygiene is something which be integrated and not be ad-hoc about.
   - **Why:** Clean data translates into better search, looser recommendations, and "happier" customers.

2. **Expand Men's Offerings**
   - **Action:** Add more men's products, especially in premium and luxury segments.
   - **Why:** Men's fashion is underserved, and there's a real opportunity to grow this part of the business.

3. **Use Luxury Discounting Tactics**
    - **Action:** Verify how the discounts on high-end products are being calculated. Based on facts, decide which products are being deeply discounted and which must be kept as exclusive.
    - **Why:** Overuse of deep discounts harms the prestige image of the luxury brand, but good discounting gains new consumers.

4. **Focus on Top Brands and Colors**
    - **Action:** Double down on partnerships with top brands (like Jolie-Robe, Max, and Puma) and keep popular colors in stock.
    - **Why:** These are what customers want most, and they drive sales.

5. **Track and Adjust Price Segment Discounts**
    - **Action:** Apply data to put sensible discounts for each price band, deepest in mid-band, more cautious in luxury.
    - **Why:** It maximizes profits and safeguards sales.

Impact and Benefits: Why This Matters for AJIO

These recommendations can be evaluated and tracked as results, when carried out by AJIO can be measured and quantified by:

- **Better Customer Experience:** With more optimized data and product inventory, users will be able to search for the items they want faster and more confidently within the platform.
- **Better Sales:** Growth in men's fashion and smart discounting can attract new customers and produce sales.
- **Smarter Inventory Management:** It's all about the best brands and colors, selling what's selling, wasting nothing, and making more money.
- **Stronger Brand:** Consistent, fact-based discounting and pricing present AJIO as professional and credible-looking, which is necessary in a competitive market.

The Big Picture

AJIO is already a fashion online leader, but the numbers show that there's still potential to expand, especially in men's fashion, optimizing luxury plays, and clean data. These adjustments aren't so much about the numbers, they're about refining the shopping experience for everyone shopping with AJIO.