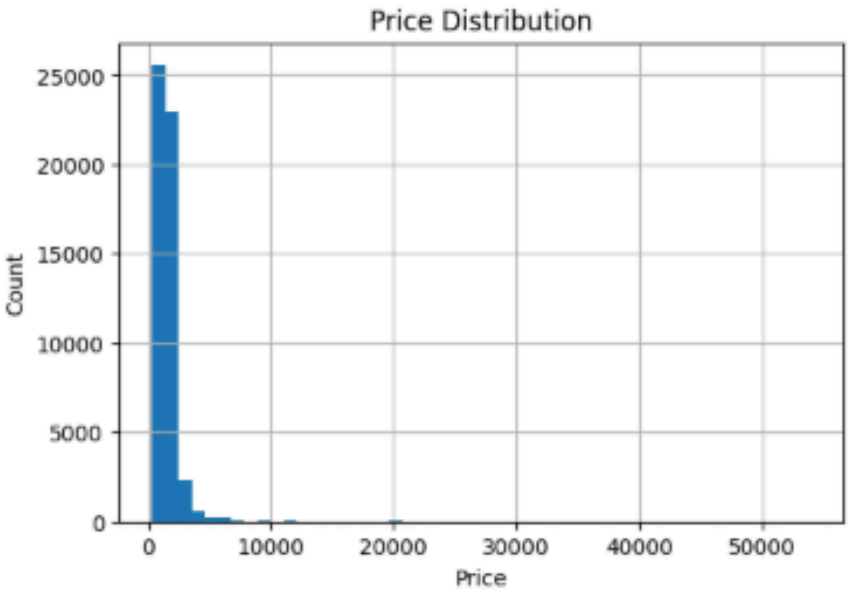
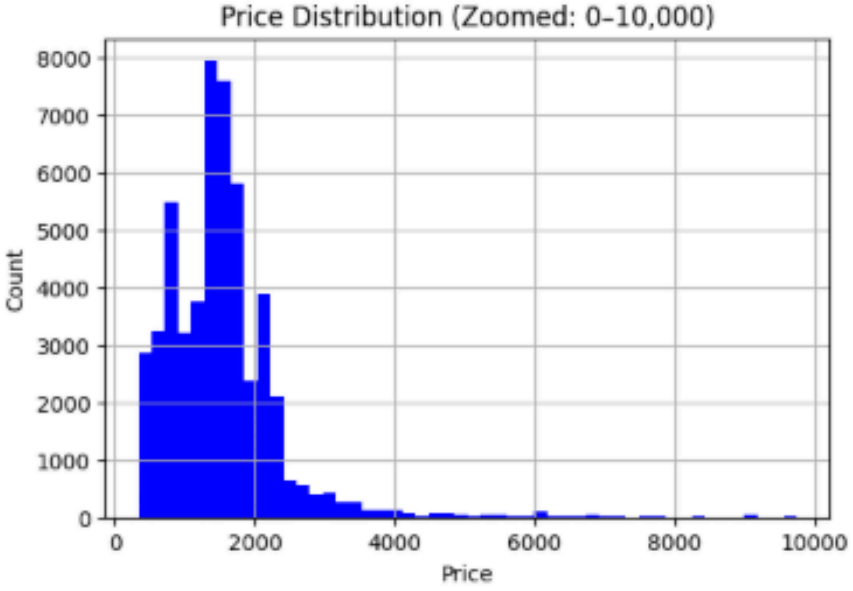
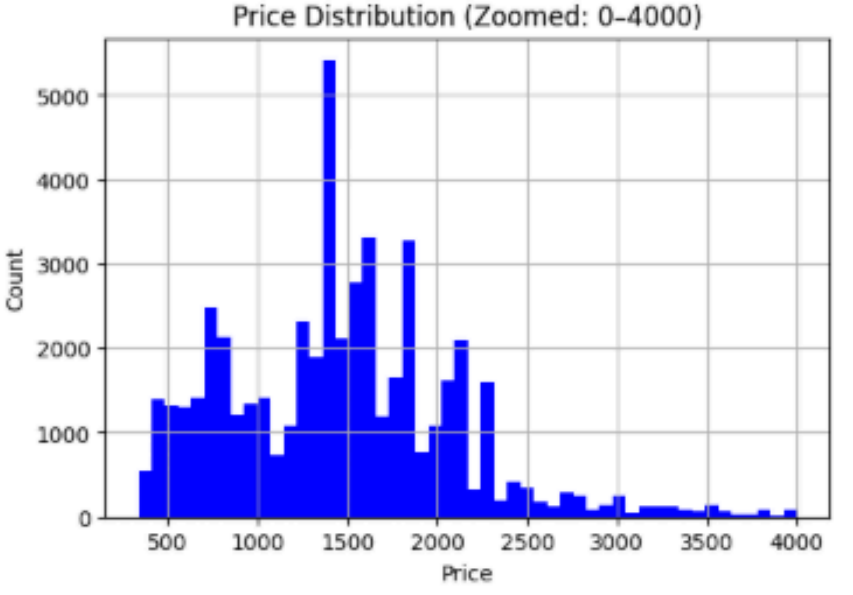
Results

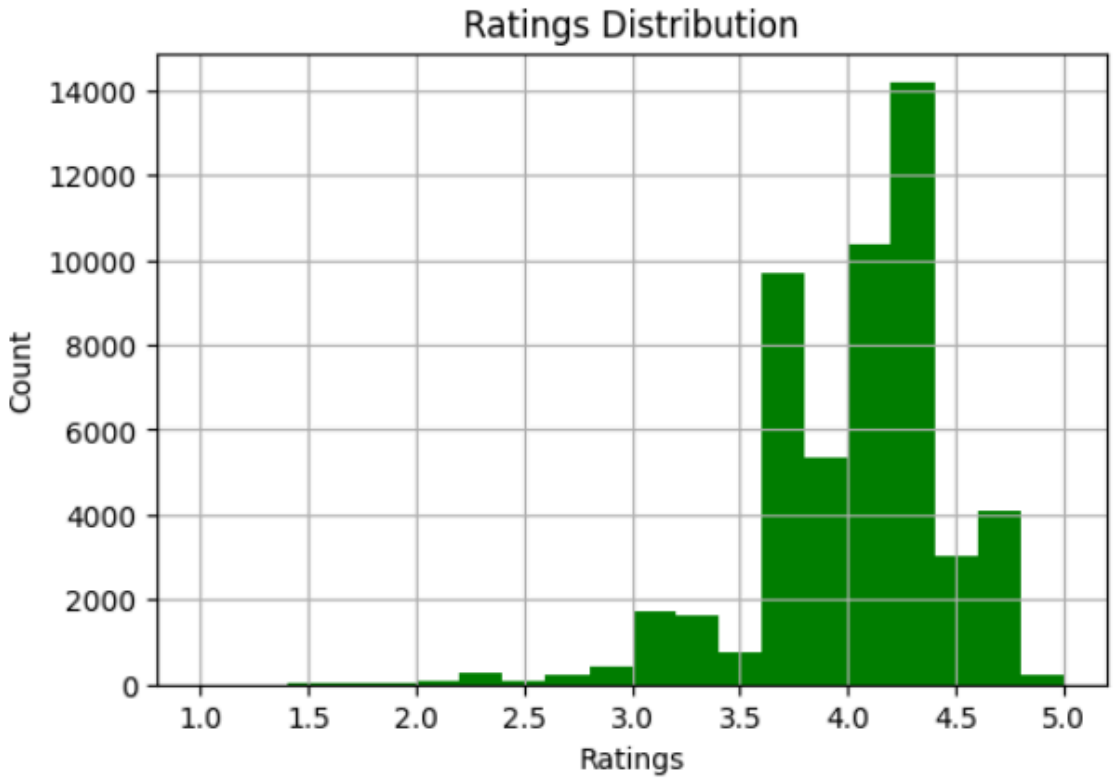
1. Price & Discount Distribution


Price Distribution
Count
Price
0
5000
10000
15000
20000
25000
0
10000
20000
30000
40000
50000


Price Distribution (Zoomed: 0–10,000)
Count
Price
0
1000
2000
3000
4000
5000
6000
7000
8000
0
2000
4000
6000
8000
10000


Price Distribution (Zoomed: 0–4000)
Count
Price
0
1000
2000
3000
4000
5000
500
1000
1500
2000
2500
3000
3500
4000

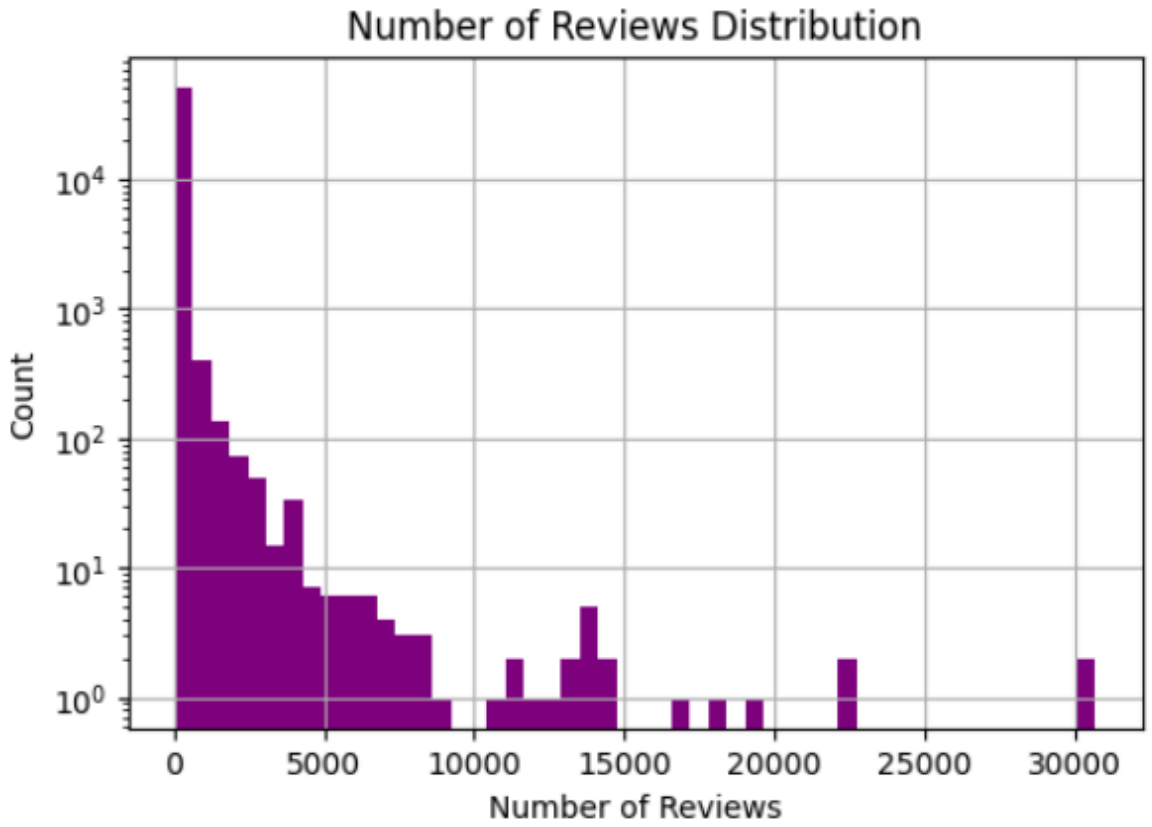
2. Ratings & Reviews Analysis



📊 1. Ratings Distribution

- Most products have ratings in the **3.5–4.5 range**.
- Very few products are below 3.0 (customers rarely buy poorly rated items).
- There's a peak around **4.2–4.5**, which is typical for e-commerce (positive bias).

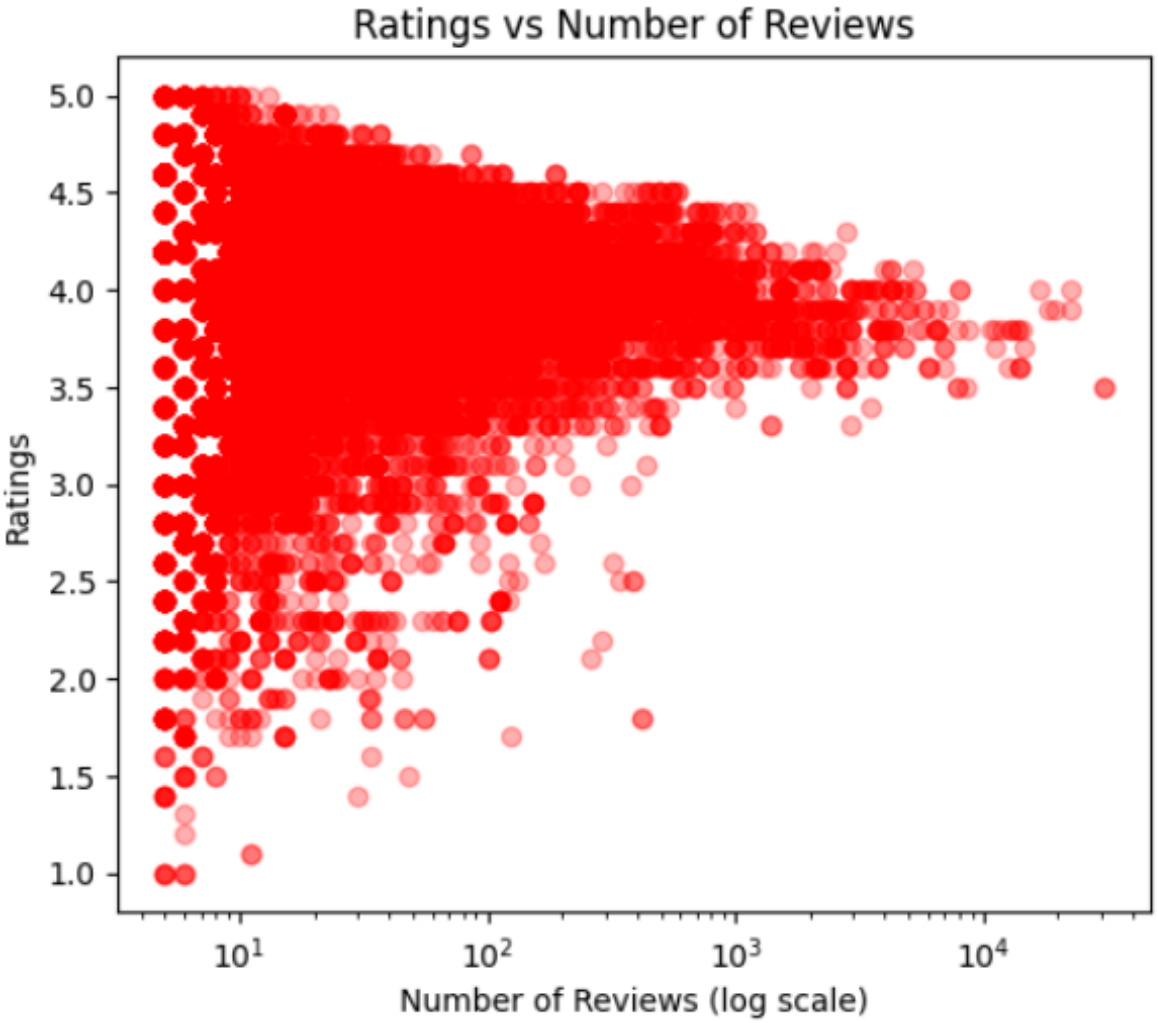
Insight: Ratings are generally favorable, so they don't differentiate products much unless combined with reviews.



📊 2. Number of Reviews Distribution

- The distribution is **heavily skewed**:
  - Majority of products have **very few reviews (0–100)**.
  - A small minority have **thousands of reviews**.
- Using log scale exposed the long tail more clearly.

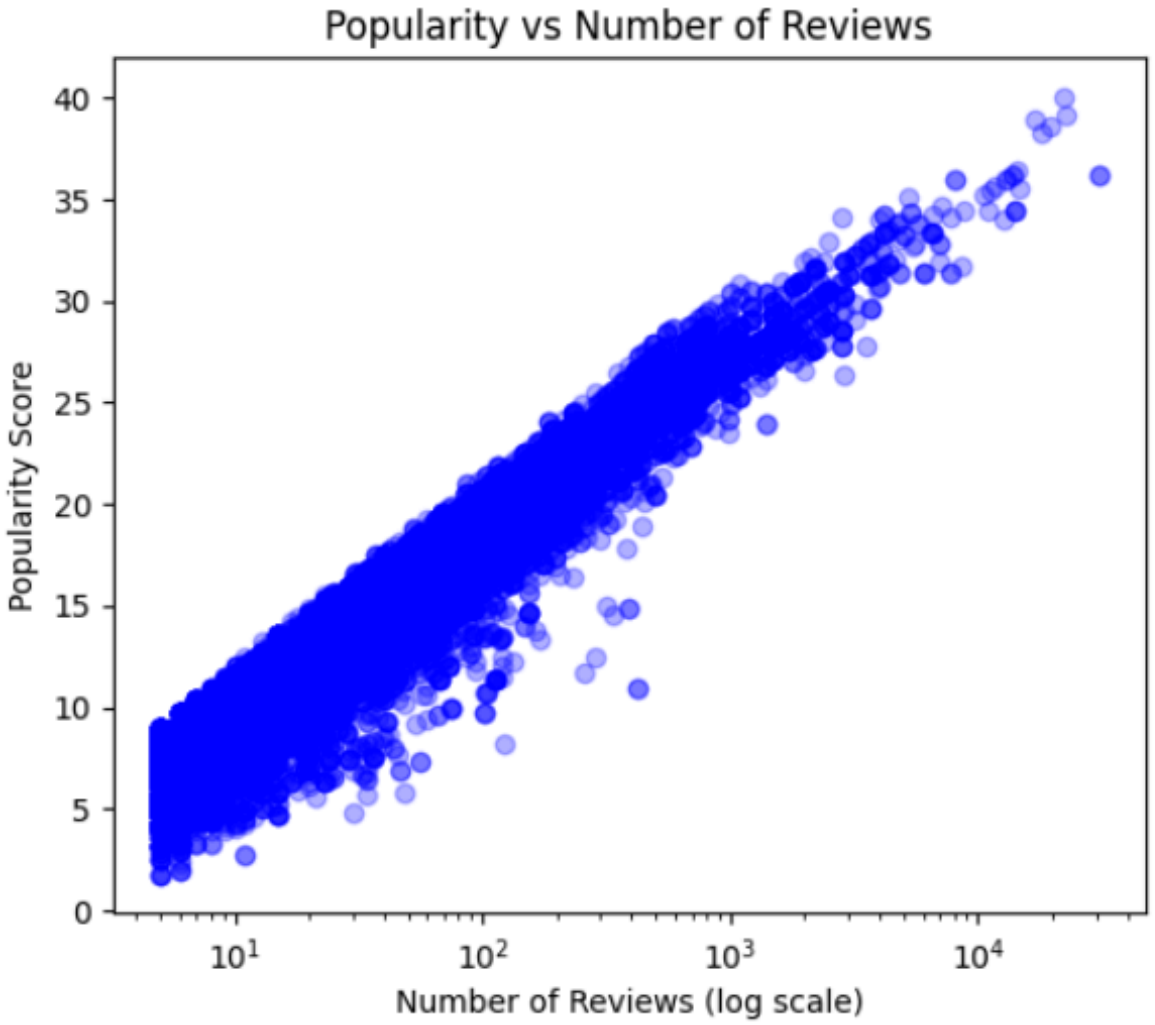
Insight: Most products struggle to gain reviews → social proof is scarce.



📊 3. Ratings vs Number of Reviews

- Products with **very few reviews** can have extreme ratings (perfect 5★ or very low).
- As review count increases, ratings **stabilize around 4.0–4.3**.

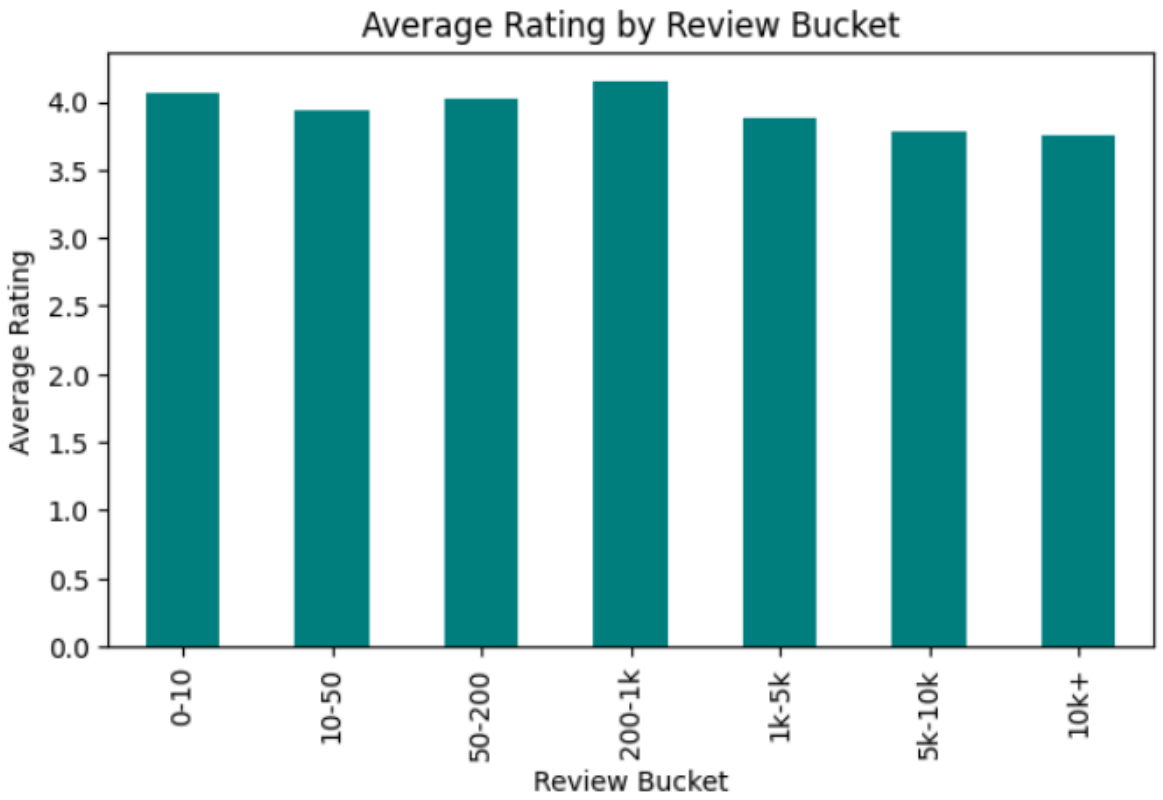
Insight: Ratings are unreliable for low-review products → review **volume** matters more than raw rating.



📊 4. Popularity vs Number of Reviews

- Strong linear relationship (on log scale): more reviews → higher popularity score.
- This validates our **popularity proxy formula (ratings × log(reviews))**.

Insight: Popularity is driven by **review volume** more than price or discount.



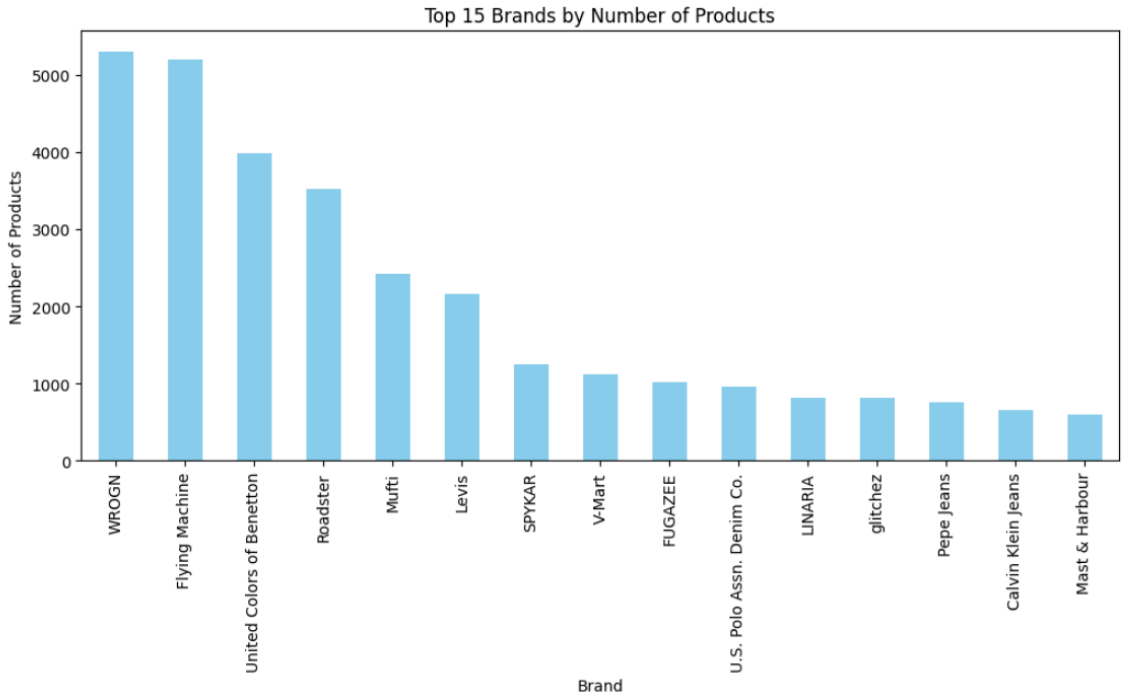
📊 5. Average Rating by Review Bucket

- Products with **0–10 reviews** have inflated average ratings (close to 4.2–4.3).
- As reviews grow, average rating slowly declines and stabilizes around **3.8–4.0**.

Insight: High review volume tends to reveal more balanced (and slightly lower) ratings, reflecting real customer sentiment.

📌 Overall Takeaway from Ratings & Reviews

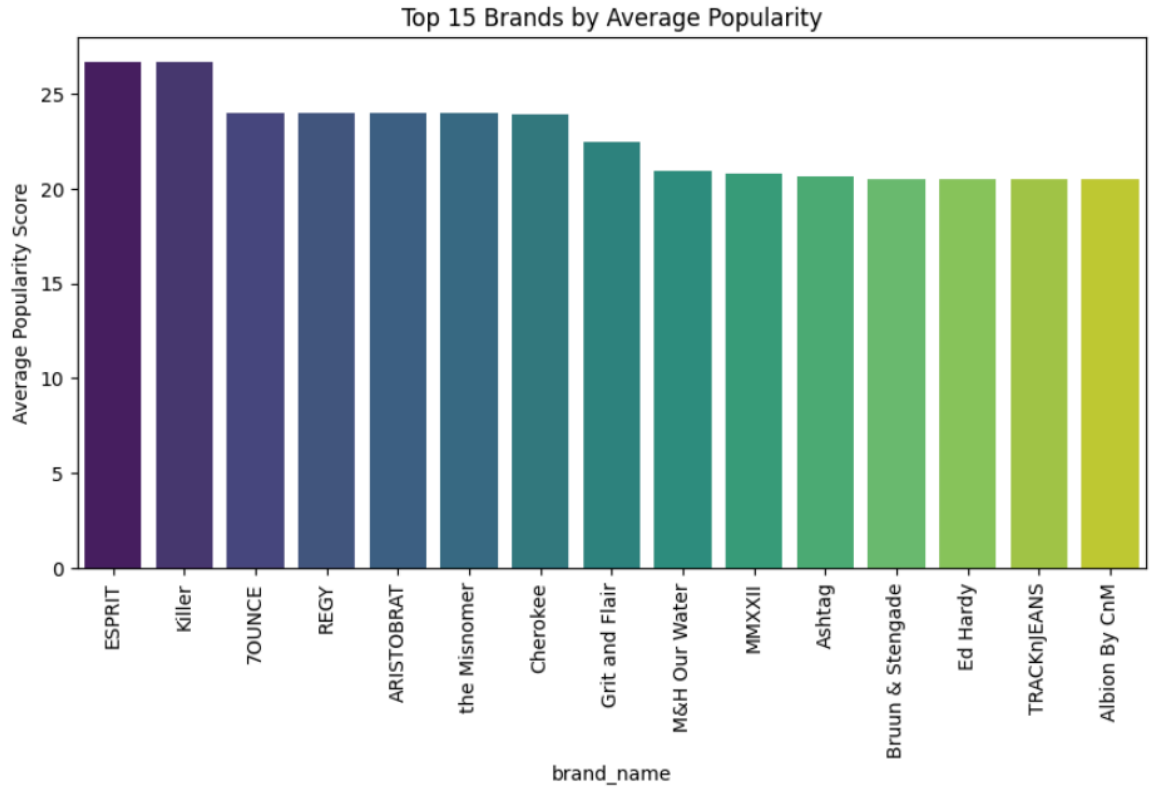
- **Social proof (reviews) is king**: Customers trust products with many reviews more than those with high ratings but few reviews.
- Ratings alone are **misleading**, especially for low-review products.
- To boost trust in new products, Myntra should **incentivize early reviews** (campaigns, badges, offers).



📊 1. Top 15 Brands by Number of Products

- **WROGN, Flying Machine, Roadster, United Colors of Benetton** dominate Myntra's catalog with **3k–5k products each**.
- Smaller brands contribute far fewer products.

Insight: Myntra's jeans collection is **highly concentrated** among a few brands → catalog strategy depends heavily on them.



📊 2. Top 15 Brands by Average Popularity

- **ESPRIT, Killer, 7OUNCE** have the **highest average popularity scores**, even though they have fewer products.
- Big brands (like WROGN, Flying Machine) don't appear here → they dominate in quantity, not necessarily in popularity.

Insight: Smaller but consistent brands often **outperform larger brands in popularity**. This means Myntra could promote these "hidden winners."

📊 3. Brand-Level Discount vs Rating

- Most brands cluster around **low to moderate discounts (<15%)** with ratings in the **3.8–4.3 range**.
- No clear positive correlation between **higher discounts and better ratings**.
- In fact, brands offering **very high discounts (>25%)** don't necessarily achieve higher ratings.

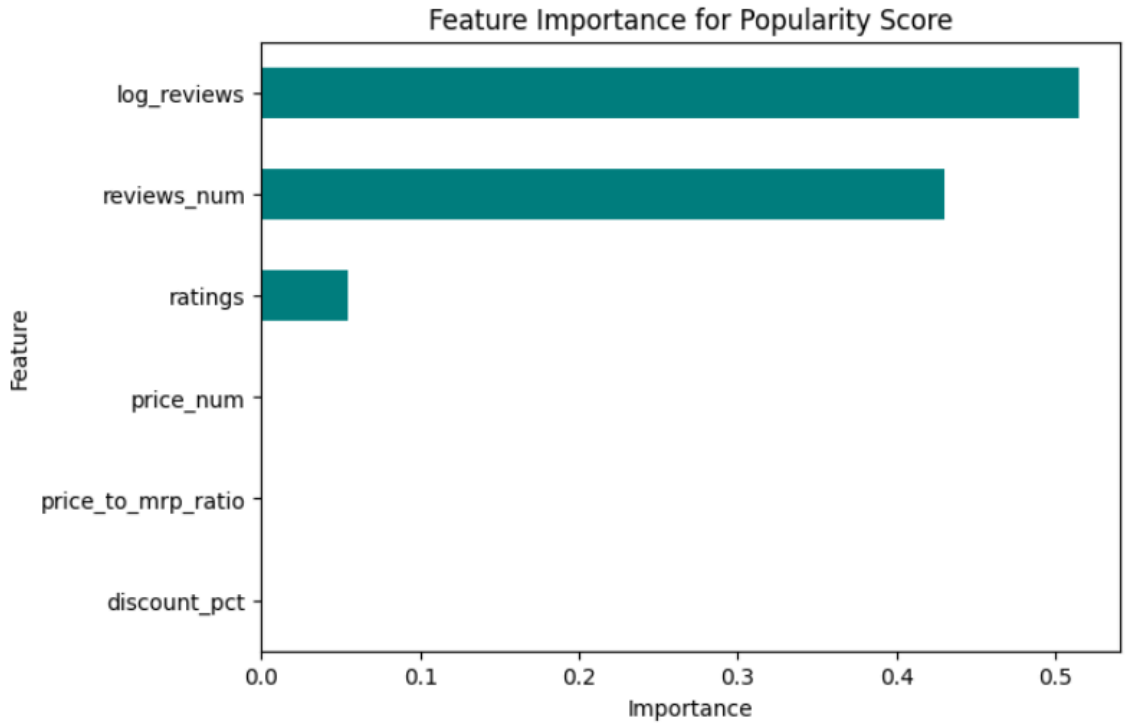
Insight: Deep discounts don't guarantee higher customer satisfaction; **product quality and consistency matter more**.

	brand_name	avg_rating	rating_variance	n_products
1	7OUNCE	4.3	0.0	3
117	FIBONACCI	3.7	0.0	3
110	ESPRIT	3.9	0.0	2
230	MR BUTTON	3.7	0.0	4
20	Albion By CnM	4.2	0.0	2
416	the Misnomer	4.3	0.0	3
193	Killer	3.9	0.0	2
347	The Dance Bible	4.2	0.0	2
14	ARISTOBRAT	4.3	0.0	2
342	TRACKnJEANS	4.2	0.0	2

📊 4. Consistent Brands (Low Rating Variance)

- Brands like **7OUNCE, ESPRIT, MR BUTTON, Albion By CnM** show **low variance in ratings** (ratings stable across all their products).
- Even with few products, their **consistency** stands out.

Insight: Myntra can **promote these brands as "Trusted Sellers"** → reliability attracts customers who care about quality.



Conclusion

1. **log_reviews (~52%)**
   - The most important feature.
   - Taking the log of reviews smooths extreme values, and it captures the fact that **popularity grows rapidly with more reviews, but the effect tapers off**.
   - Meaning: products with many reviews are far more likely to be popular.
2. **reviews_num (~42%)**
   - The second most important feature.
   - Raw number of reviews is also highly predictive.
   - Confirms that **review volume (social proof)** drives popularity.
3. **ratings (~6%)**
   - Ratings help, but they're much less influential than reviews.
   - Example: A 4.2★ product with 5,000 reviews is more trusted than a 4.8★ product with only 10 reviews.
4. **price_num, price_to_mrp_ratio, discount_pct (~0%)**

- Pricing and discounting have almost **no effect** on predicting popularity.
- Customers are influenced more by **what other people say** (reviews/ratings) than by how cheap the product is.