Report of Findings: Amazon Products Data Analysis

---

1. Dataset Summary

- The dataset consists of **1465 products** with **16 attributes**.
- Initial data types were mostly textual and required cleaning for analysis.
- Some missing values were observed, especially in rating_count.

---

2. Data Cleaning and Preprocessing

- Removed special symbols (₹, %, commas) from price-related and discount fields.
- Converted important fields like discounted_price, actual_price, discount_percentage, rating, and rating_count into **numeric types**.
- Missing or invalid entries were carefully managed during conversion.

---

3. Feature Scaling

- Applied **Min-Max Scaling** to normalize all numeric attributes between **0 and 1**.
- Scaled Columns:
    - discounted_price
    - actual_price
    - discount_percentage
    - rating
    - rating_count
- Scaling ensured fair and comparable visualizations without skew from large values.

---

4. Visual Analysis and Key Observations

Histograms

- **Discounted Price Distribution:** Most products are priced low, with a few expensive outliers.
- **Rating Distribution:** Majority of ratings lie between **3.5 and 4.5**, indicating overall customer satisfaction.

Boxplot

- Presence of **price outliers**, suggesting a few very expensive products compared to the rest.

Scatterplot (Discounted Price vs Rating)

- **No strong correlation** between price and rating.

- Expensive products are **not necessarily better rated**.

Pairplot

- Shows strong positive relationship between discounted_price and actual_price.

Heatmap (Correlation Matrix)

- **High correlation** between actual_price and discounted_price.
- **Weak or no correlation** between price-related fields and product ratings or rating count.

---

5. Key Insights

- Most products are **affordable** and **well-rated**.
- **Heavily discounted products** are common, but discounts don't correlate with higher ratings.
- **Expensive items** do not guarantee better customer satisfaction.
- **Pricing strategies** appear **independent of customer ratings** — companies often discount products irrespective of ratings.

✅ **Conclusion**

- The Amazon dataset is cleaned, scaled, and ready for modeling or deeper business insights.
- Marketing teams should not assume that increasing product price guarantees better ratings.
- Further detailed analysis (e.g., by product categories or brands) could provide even sharper insights.