Amazon Sales Analysis Report

---

1. Introduction

The purpose of this project is to analyze Amazon sales data using **Power BI** and **Python** to uncover insights about product categories, pricing, discounts, and customer ratings. This analysis helps understand which products perform well and how pricing and discount strategies impact ratings.

Business Questions Addressed:

- What are the **top-selling product categories**?
- How does **price affect customer ratings**?
- What is the **distribution of discount percentages** across products?
- Which are the **best and worst-rated products**?

---

2. Dataset Overview

- **Dataset Source:** Amazon Sales Dataset (from Kaggle)
- **Total Rows:** 1,465
- **Total Columns:** 16

Key Columns:

- `Product_name` – Name of the product
- `category` – Product category
- `Discounted_price` – Price after discount
- `actual_price` – Original price
- `Discount_percentage` – Discount applied (%)
- `rating` – Customer rating (0-5 scale)
- `rating_count` – Number of reviews

---

3. Data Cleaning & Processing (Python)

Before loading the dataset into **Power BI**, we performed data cleaning using **Python (Pandas, NumPy, Matplotlib, Seaborn)**:

Converted numerical columns (prices, ratings, discount %) to the correct data types.
Handled missing values (e.g., filling NaNs with median values where necessary).
Removed special characters (e.g., currency symbols) to ensure proper numerical analysis.
Filtered out irrelevant data (e.g., products with zero ratings or invalid pricing).

📂 **Python Script Used:** Amazon Data Cleaning & Analysis

---

4. Power BI Dashboard & Key Insights

The cleaned data was used to create an **interactive Power BI dashboard**, visualizing key business insights.

4.1 Top-Selling Categories (Bar Chart)

The **most-reviewed category** is *Electronics*, followed by Home & Kitchen products. Certain niche categories have very few reviews, indicating low customer engagement.

4.2 Price vs. Rating Analysis (Scatter Plot)

Higher-priced products tend to have higher ratings, but some low-priced products also receive high ratings, indicating affordability doesn't always mean poor quality. No strong correlation between price and rating, meaning **other factors like brand reputation, features, and durability** play a role.

4.3 Discount Distribution (Histogram)

The most common discount range is **50-60%**. Some products offer extremely high discounts (90%+), which may be marketing tactics to attract buyers.

4.4 Best-Rated & Worst-Rated Products (Table View)

Best-Rated: High-end products with **4.5+ ratings** and thousands of reviews. **Worst-Rated:** Products with **ratings below 3.0** are often cheap, have quality issues, or misleading descriptions.

---

5. Conclusion & Recommendations

Key Takeaways:

- **Electronics & Home Appliances** dominate the market in terms of customer reviews.
- **Expensive products generally have higher ratings**, but price alone does not determine customer satisfaction.
- **50-60% discounts are the most common** and could be a pricing strategy to drive sales.
- Some **low-rated products have high discounts**, indicating quality concerns.

Business Recommendations:

✔️ Optimize pricing strategy: Avoid excessive discounts on low-rated products to maintain brand trust. ✔️ Focus on customer feedback: Improve product quality and descriptions for categories with low ratings. ✔️ Targeted promotions: Offer strategic discounts on high-rated, high-selling items to increase revenue.

---

6. Project Files & Links

Download Full Report (PDF)

Power BI Dashboard (.pbix)

Python Data Cleaning Code
