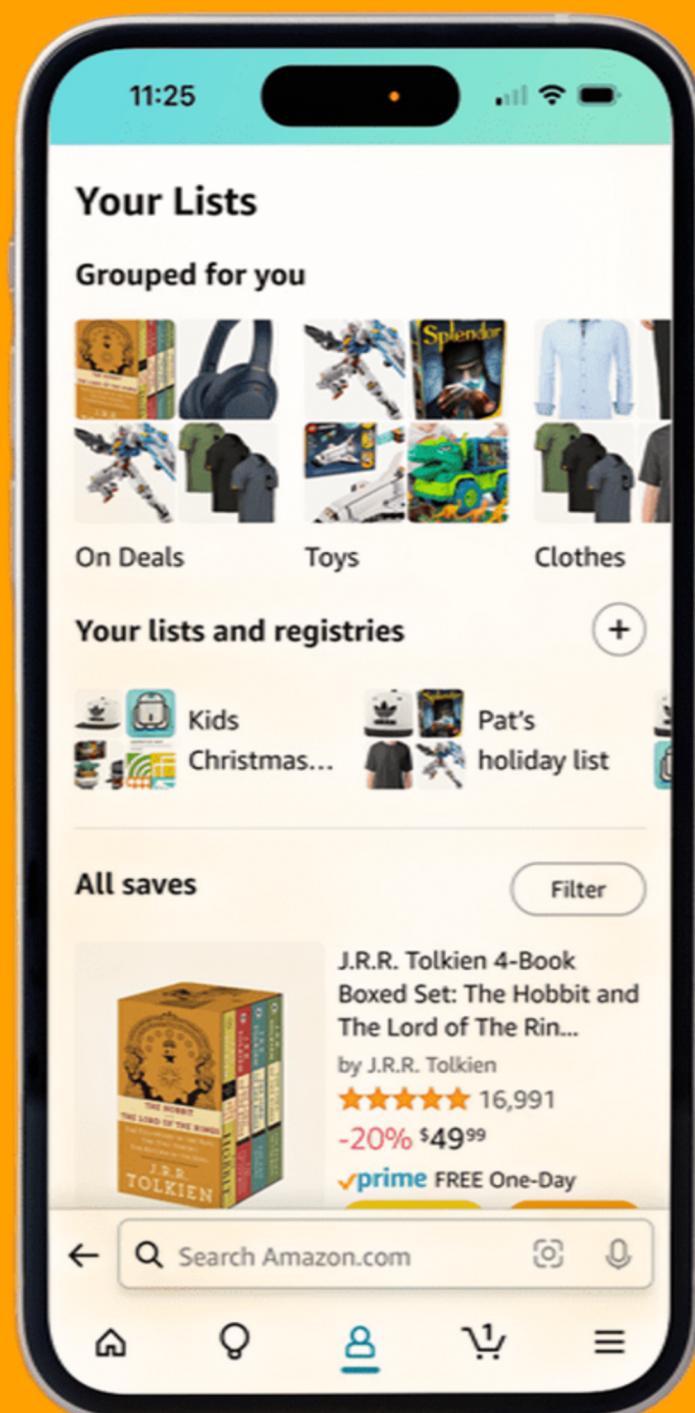


Amazon Sales & Product Reviews Analysis

BY Sarah Hani Farid



Agenda



01

The Problem

02

Data Quality & Cleaning

03

Business Questions & Analysis

04

Exploratory Data Analysis (EDA)

05

Insights & Recommendations



The Problem Statement:

Amazon faces the challenge of understanding the relationship between pricing, discounts, and customer ratings to optimize product performance. The goal is to leverage sales and review data to generate actionable insights that support pricing, marketing, and customer satisfaction

02



Data Quality & Cleaning

Schema and Data Types Check

- Checked schema and confirmed data types and logical consistency for prices, discounts, and ratings.

Handling Missing Values, Duplicates, and Outliers

- Detected data quality issues; dropped 2 null values to keep the dataset clean and ready for analysis and Verified no duplicate records

Our Business Questions & Analysis

1

- What is the overall distribution of product ratings?

2

- What discount ranges are most common across products?

3

- Is there a relationship between product price and rating?



4

- Which categories have the most products listed?



5

- Are actual prices strongly correlated with discounted prices?

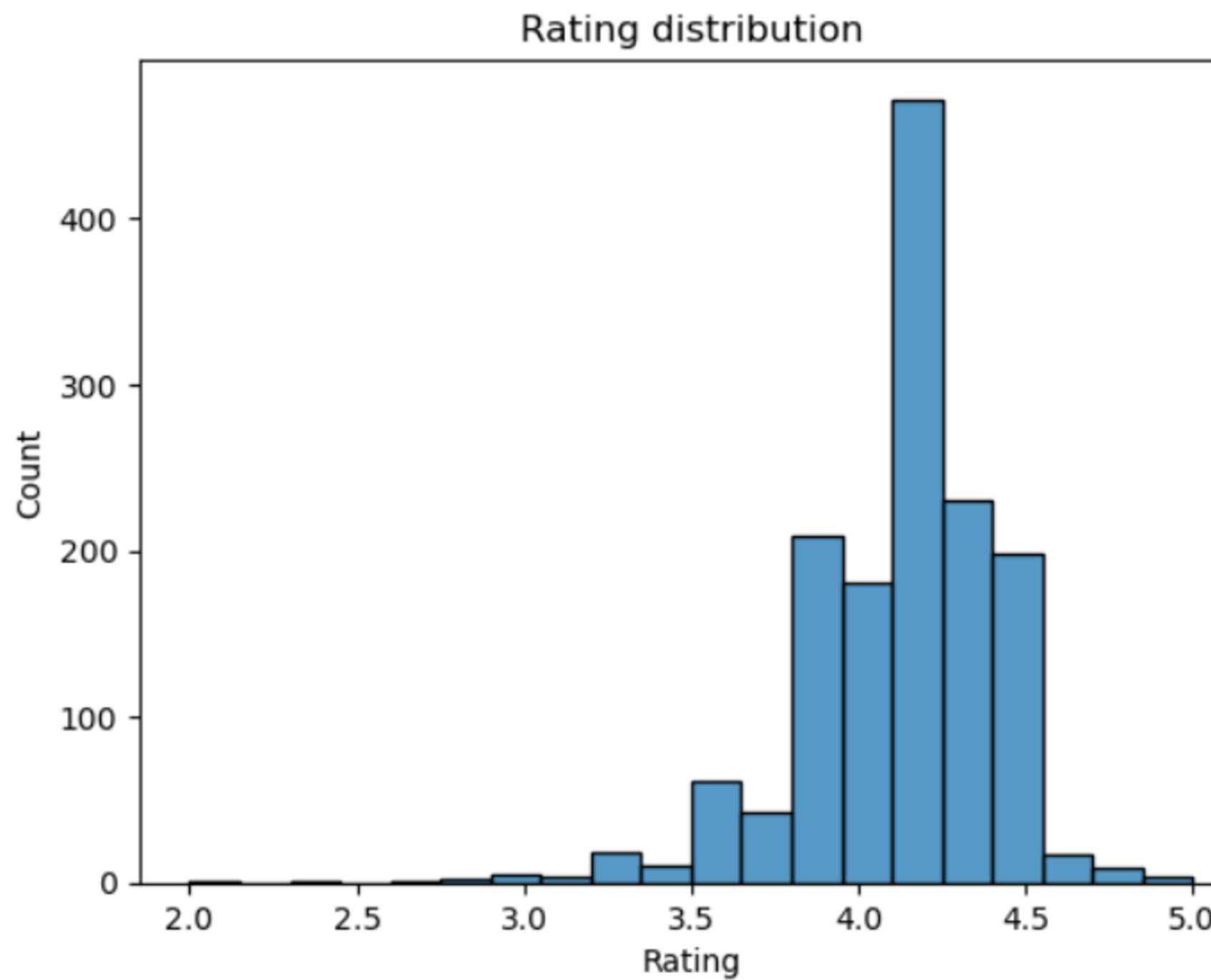
6

- Are there significant outliers in product prices, discounts, and ratings that may affect analysis?



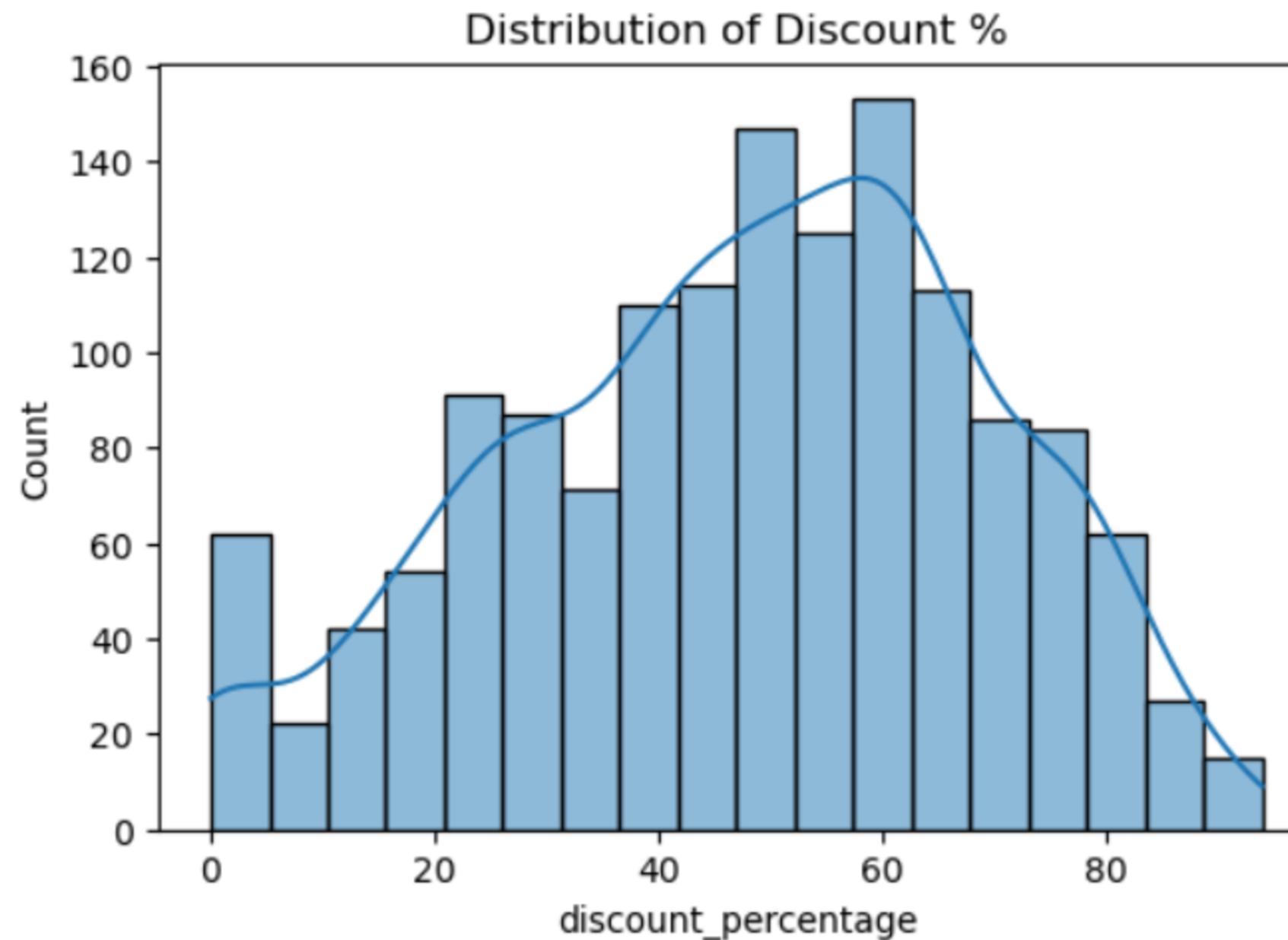
04

Exploratory Data Analysis (EDA)



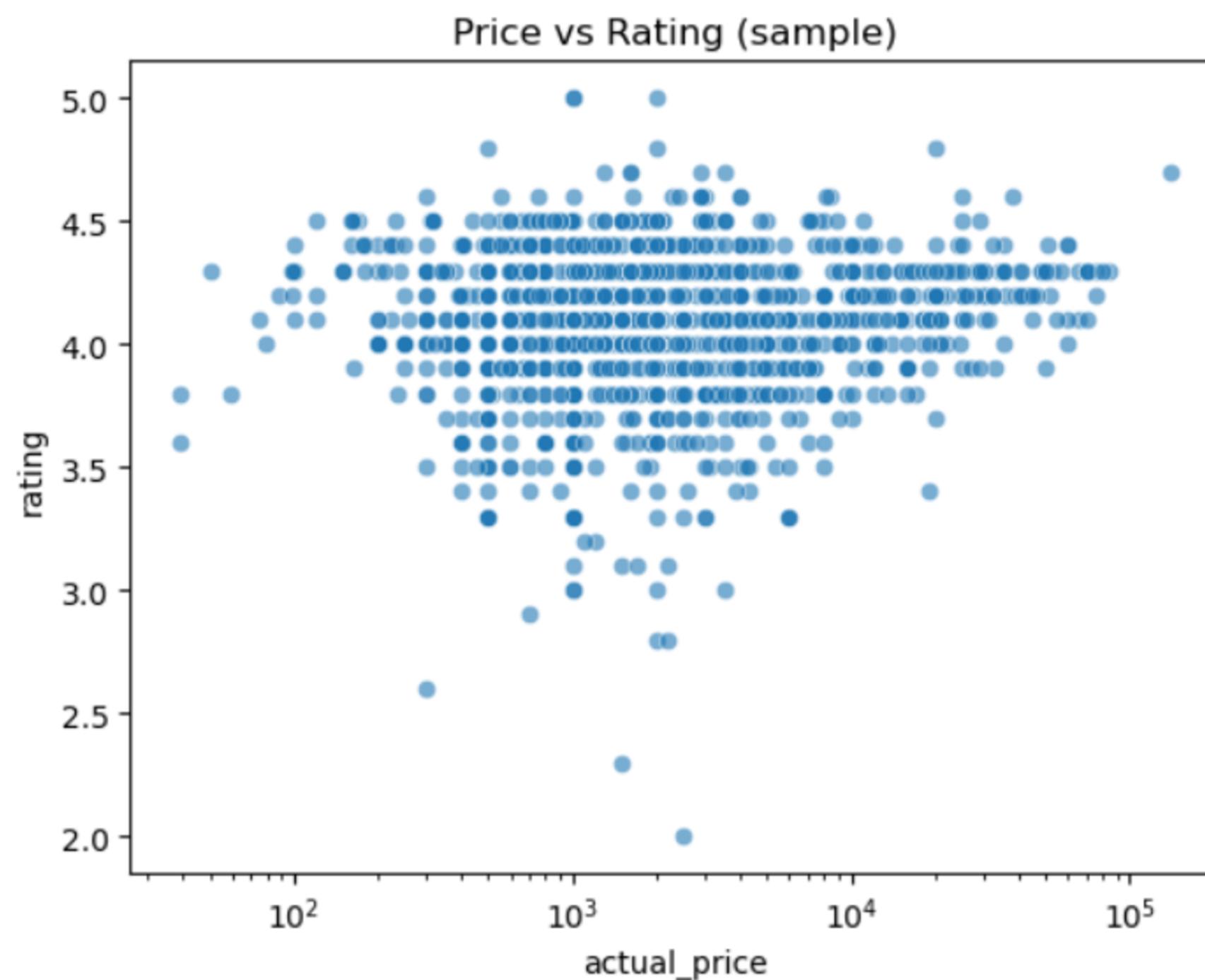
1. Distribution of Product Ratings

- Visualized the overall spread of ratings (0–5).
- Most products fall between 3–5, showing generally positive customer feedback.



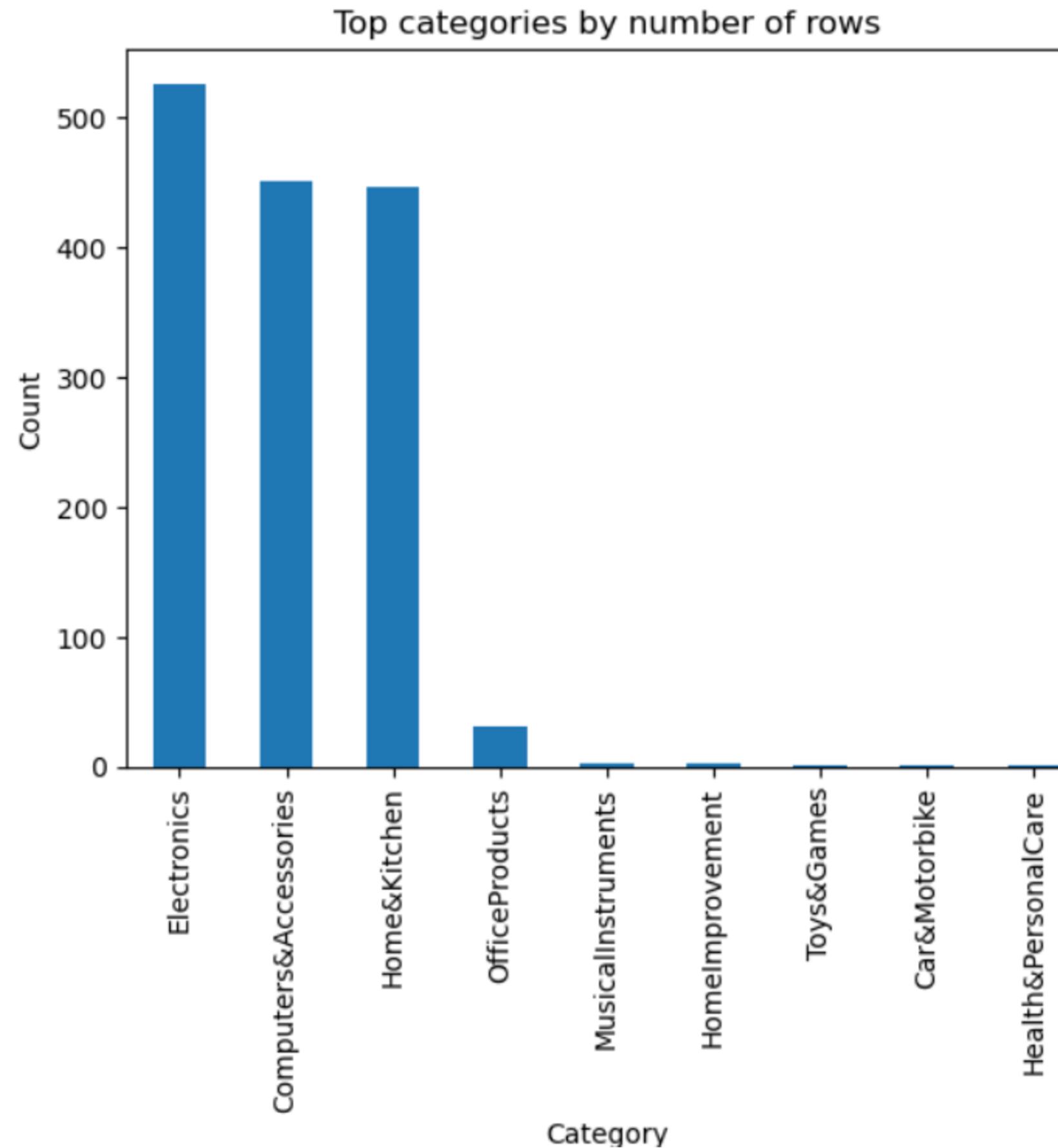
2. Common Discount Ranges

- Majority of discounts fall in the 20-50% range, with peaks around 40-60%.
- Few products have extreme discounts above 70%, which are less common.



3. Relationship Between Product Price and Rating

- Found weak correlation: higher price does not necessarily lead to higher ratings.
- Customer satisfaction seems driven by value/quality, not just price.



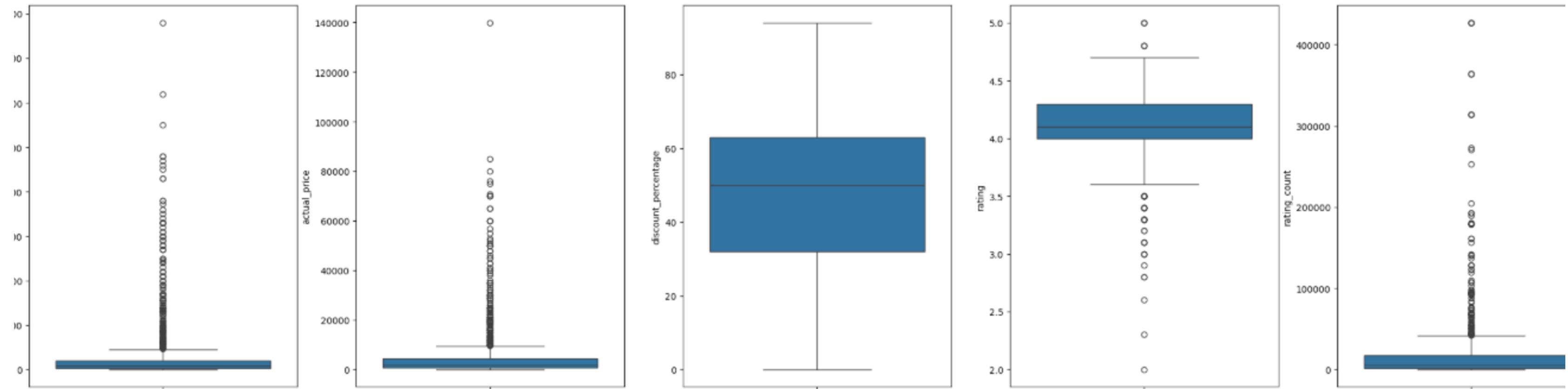
4. Most Represented Categories

- Electronics, Computers & Accessories, and Home & Kitchen are the top categories by data volume, each with nearly 500 rows. The remaining categories have much fewer entries, indicating a strong focus or dominance of those top three in the dataset.



5. Correlation Between Actual and Discounted Prices

- a very strong correlation (0.96) between actual price and discounted price. This means discounts are usually consistent and follow a predictable pattern.



6. Outliers in Prices, Discounts, and Ratings

- Uneven distribution of earnings and hours: Most values are low, but there are a few very high outliers—indicating significant variation in both income and working hours.
- Clear positive skew: The box plots show that the data is skewed upward, meaning rare high values are pulling the average up.

05

Insights & Recommendations



Optimize Discount Strategies

Apply targeted discounts on competitive/high-demand categories instead of uniform discounts to maximize both revenue and acquisition.

Category Expansion

Focus on underrepresented but growing categories to capture new market opportunities.

Boost Ratings & Reviews

Actively encourage genuine reviews and resolve issues on low-rated products to strengthen trust and drive conversion