The dataset from the Amazon Product Review Analysis case study has been successfully loaded. Here's a quick overview of the structure:

📊 Key Columns in the Dataset:

product\_id: Unique identifier for each product

product\_name: Name of the product

category: Product category (multi-level hierarchy, pipe-separated)

discounted\_price: Price after discount

actual\_price: Original price

discount\_percentage: Discount as a decimal (e.g., 0.64 = 64%)

rating: Average rating

rating\_count: Number of ratings

user\_id, user\_name, review\_id, review\_title, review\_content: Aggregated review-related data (comma-separated values)

img\_link, product\_link: Links to images and Amazon listings

✅ Here are the insights from your Amazon Product Review Analysis based on the dataset provided:

---

✅ Q1: Average Discount Percentage by Product Category

Category    Avg. Discount (%)

HomeImprovement   57.5%

Computers&Accessories   54.0%

Health&PersonalCare     53.0%

Electronics 50.8%

MusicalInstruments      46.0%

Car&Motorbike     42.0%

Home&Kitchen      40.1%

OfficeProducts    12.3%

Toys&Games  0.0%

---

✅ Q2: Number of Products Listed Per Category

Category    Count

Electronics 526

Computers&Accessories   453

Home&Kitchen      448

Others (Combined) 38

---

✅ Q3: Total Number of Reviews Per Category

| Category               | Total Reviews | |

The Amazon product review dataset has been successfully loaded. Here's a summary of what it contains:

📊 Dataset Overview:

Total Rows: 1,465

Total Columns: 16

Key Columns:

product\_id, product\_name, category

discounted\_price, actual\_price (stored as object—needs conversion)

discount\_percentage, rating (also object—needs conversion), rating\_count

Reviewer data: stored as comma-separated strings (user\_id, user\_name, etc.)

🛠 Data Cleanup Needed:

actual\_price and rating are incorrectly typed as object; we'll need to:

Remove any non-numeric characters (like ₹, commas)

Convert them to numeric types

rating\_count has 2 missing values

1. Cleaning and converting these fields.

2. Performing the 10 specific analysis tasks listed in the case study instructions

✅ The data has been cleaned and is ready for analysis.

Summary After Cleaning:

actual\_price and discounted\_price are now floats.

rating is numeric (some missing values coerced to NaN).

discount\_percentage ranges from 0% to 94%.

rating\_count ranges from 2 to 426,973.

---

Q11: How does the rating relate to the level of discount?

We'll calculate the correlation and plot a scatter plot.

Q12: How many products have fewer than 1,000 reviews?

Q13: Which categories have products with the highest discounts?

We'll group by category and sort by average discount.

Q14: Top 5 products by combined rating and number of reviews.

We'll create a score: rating \* log10(rating\_count + 1) to balance volume and quality.

Proceeding with this analysis now.  
Here are the insights from the additional analysis:

---

✅ Q11: How does the rating relate to the level of discount?

Correlation: -0.16 (weak negative correlation

Interpretation: Products with higher discounts tend to have slightly lower ratings, but the relationship is weak.

📈 A scatter plot has been saved here: rating\_vs\_discount.png

---

✅ Q12: How many products have fewer than 1,000 reviews?

Answer: 326 products

---

✅ Q13: Which categories have the highest average discounts?

Top 5 categories by average discount:

1. Phone Charms – 90%

2. Cable Connection Protectors – 90%

3. Earpads – 90%

4. Adapters (Headphones) – 88%

5. Dust Covers (Keyboard/Mice Accessories) – 87.5%

---

✅ Q14: Top 5 Products by Combined Score (Rating × Log10(Reviews))

Product Name      Rating      Reviews     Score

Amazon Basics High-Speed HDMI Cable (6 Ft)      4.4   426,973     24.77

AmazonBasics Flexible Premium HDMI Cable  4.4   426,973     24.77

Amazon Basics HDMI Cable (6 Ft, 2-Pack)   4.4   426,973     24.77

AmazonBasics HDMI Cable (Alt Variant)     4.4   426,972     24.77

SanDisk Extreme SD UHS I 64GB Card  4.5   205,052     23.90

---

Generate pivot tables for the Excel dashboard?

Package everything (cleaned dataset + plots + pivot data) into an Excel file for dashboard creation?

Let me know how you'd like to proceed.

Your analysis results and cleaned dataset have been saved to an Excel workbook with multiple sheets for your dashboard creation:

📁 Download here: Amazon\_Case\_Study\_Analysis.xlsx

Included Sheets:

Cleaned Data: Full dataset (cleaned and ready)

Avg Discount by Category

Product Count by Category

Total Reviews by Category

Price Comparison (Actual vs Discounted)

Rating Distribution

Revenue by Category

Products by Price Range

Top 5 Products