

Amazon Product Review Analysis Project

Project Overview

This Data Analysis project explores Amazon product and customer review data to uncover actionable insights. The goal is to support better decision-making in areas like product improvement, marketing strategy, and customer engagement through structured Excel-based analysis.

Data Sources

The dataset was sourced from a Digital Learning Management System (LMS) platform as part of a case study project.

Project Problem

In a highly competitive e-commerce environment like Amazon, businesses often struggle to interpret massive volumes of product and customer review data effectively.

Without proper analysis:

- High-performing products might go unnoticed
- Poor ratings may persist unaddressed
- Marketing efforts may fail to target the right product segments.

This project solves that problem by analyzing structured product data to highlight:

- Which products drive the most engagement
- How discount levels affect ratings
- Where review volume can signal potential improvements.

Project Objectives

- Identify opportunities for product improvement
- Provide insight to guide marketing strategies
- Improve customer engagement based on reviews and ratings

Data Cleaning & Preparation

The dataset was cleaned and transformed using Microsoft Excel. Key cleaning steps included:

- Removing duplicates and handling missing values
- Splitting multi-level categories

- Creating helper columns for:
 - Price range buckets
 - Discount percentage range
 - Review counts (by summing-up comma-separated review IDs)
 - Potential revenue calculations
- Formatting the dataset as a Table to enable efficient PivotTable creation

Tools Used

Microsoft Excel used for:

- Data Import and Structuring
- Data Cleaning
- PivotTable Calculations
- Visual Dashboards with Charts

Exploratory Data Analysis (EDA)

Key business questions answered:

- What is the total number of reviews per category?
- Which products have the highest number of reviews?
- How does the rating relate to the level of discount?
- Identify the top 5 products in terms of rating and number of reviews combined
- What is the total potential revenue by category?
- How many products have a 50% or more discount?

Key Insights

- High-Discount Products Are Common: Over 660 products have a discount of 50% or more, showing aggressive pricing strategies across categories.
- Review Distribution Is Uneven: Most products have fewer than 8 individual reviews, indicating that review counts are often underrepresented despite high `rating_count` values.
- Top Products Have Both High Ratings and Reviews: A handful of products (e.g., B09ZHCJDP1, B0BQ3K23Y1, etc.) stand out by combining 4.7+ average ratings with the highest review counts, ideal for promotional focus.
- Electronics Dominate in Revenue Potential: Categories like Electronics show the highest total potential revenue, especially where high prices meet high review activity.
- Ratings Slightly Drop with Higher Discounts: Surprisingly, lower-discounted products have slightly higher average ratings than heavily discounted ones, hinting at a possible trade-off between pricing and perceived quality.

- Review Quality Differs from Rating Counts: The review_count column derived from actual review_ids gives a more accurate picture of real review volume than the raw rating_count column.

Data Analysis Techniques

This section include examples of Excel formulas used during the project, such as:

=N2 * S2

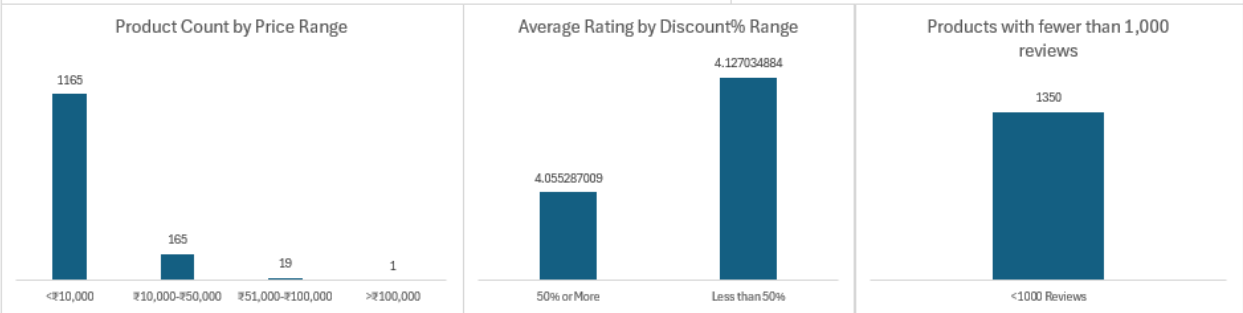
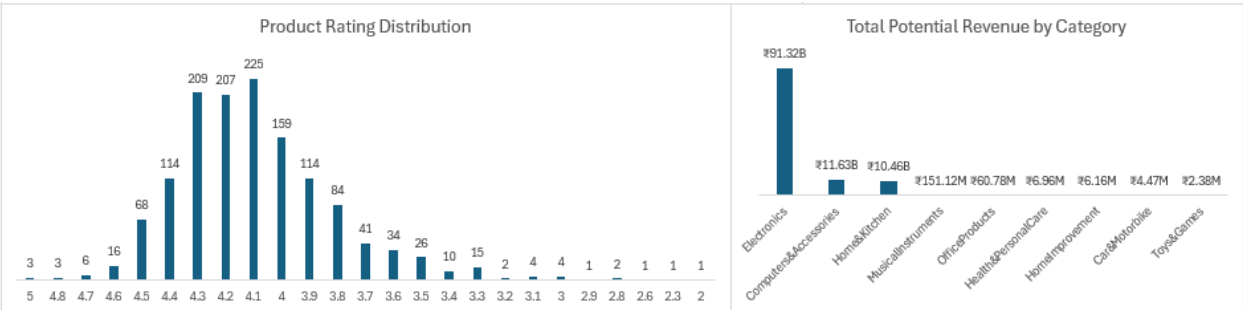
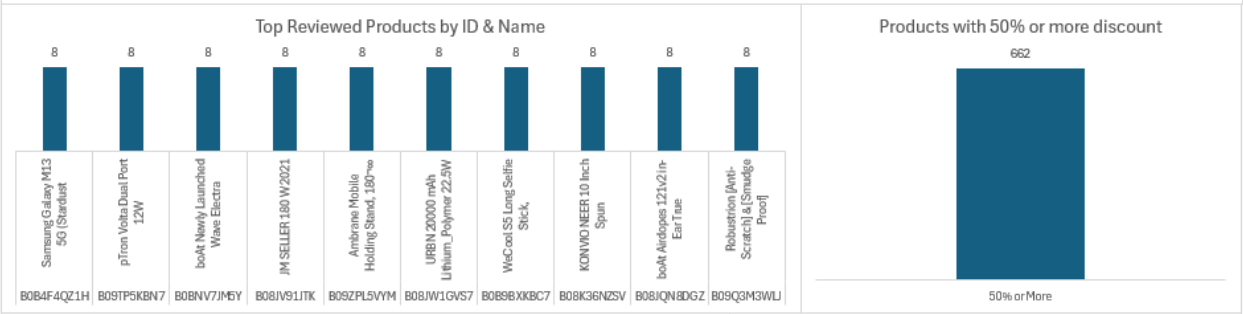
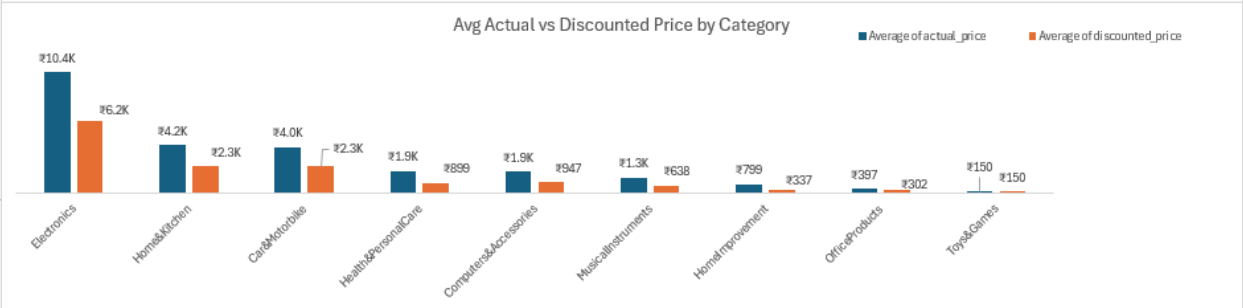
=LEN(U2)-LEN(SUBSTITUTE(U2," ",""))+1

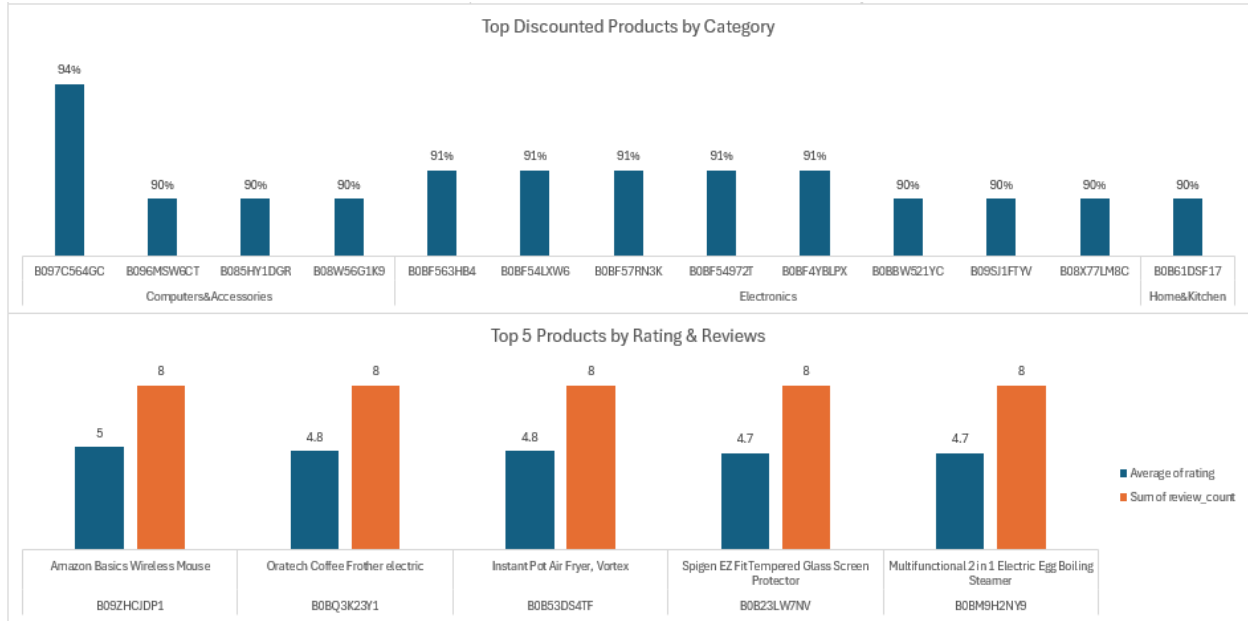
=IF(N2<10000,"<₹10,000", IF(N2<=50000,"₹10,000-₹50,000", IF(N2<=100000,"₹51,000-₹100,000", ">₹100,000")))

Final Analysis & Dashboard

A clean Excel dashboard was developed to visualize the insights using bar charts for categorical comparisons







Note

You can find the cleaned dataset, PivotTable outputs, and the final dashboard here:

[x Amazon case study 1.xlsx](#)