Amazon Product Review - Data Cleaning Documentation

1 Introduction

This document outlines the comprehensive data cleaning and preparation processes applied to the Amazon Product Review dataset. These steps ensured analytical accuracy, enhanced interpretability, and supported the successful completion of all project objectives, including the creation of an interactive dashboard and professional reporting.

2 Duplicate Removal

To maintain data integrity and pre-empt misleading insights, duplicates were removed from the **Product ID** column. This guaranteed that each product was represented only once in the dataset, ensuring accurate product counts, calculations, and reliable analysis results.

3 Product Name & Category Shortening

- **Product Name**: Long product names were shortened to display only the most relevant details (e.g., first 3 words). This improves dashboard readability, especially within Pivot Tables, slicers, and visual charts.
- **Product Category**: The hierarchical category string was split, and only the **Main Category** was retained, providing a clearer, more concise view of product segmentation during analysis.

4 Additional Columns Created for Analysis

Five new columns were engineered to simplify analysis and effectively answer the project's task requirements:

🔷 4.1 Conditional Columns

- **Price Range Bucket**:
  Products were grouped based on their discounted price into three distinct ranges:
  ➤ <$200
  ➤ ₹$200–$500
  ➤ >$500
- **Discount Percentage Band**:
  Discount percentages were segmented into 10% intervals for simplified reporting:
  ➤ 0%-10%, 11%-20%, 21%-30%, up to the final band 91%-100%.

- **Low Review**:
  ➤ Products with fewer than 1,000 reviews were tagged "Yes"
  ➤ Products with 1,000 reviews or more were tagged "No"
  This helped isolate products with limited customer engagement.
- **High Discount**:
  ➤ Products with a discount of 50% or more were tagged "Yes"
  ➤ Products with less than 50% discount were tagged "No"
  Enabling quick identification of significant discount promotions.

🔷 4.2 Custom Columns

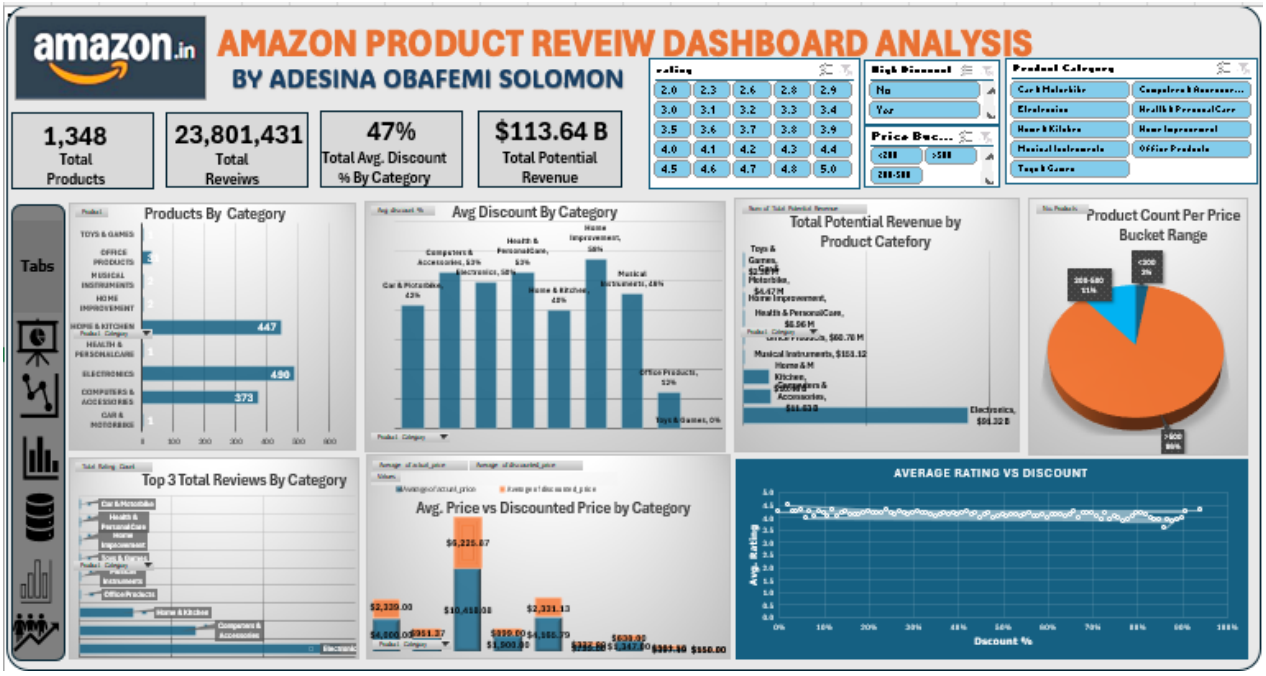
- **Total Potential Revenue**:
  Calculated by multiplying Actual Price by Rating Count, this metric estimates the potential revenue a product could generate based on current price and engagement.
- **Combined Score**:
  Computed by multiplying Rating by Rating Count, the Combined Score provides a balanced metric for identifying top-performing products, considering both popularity and quality.

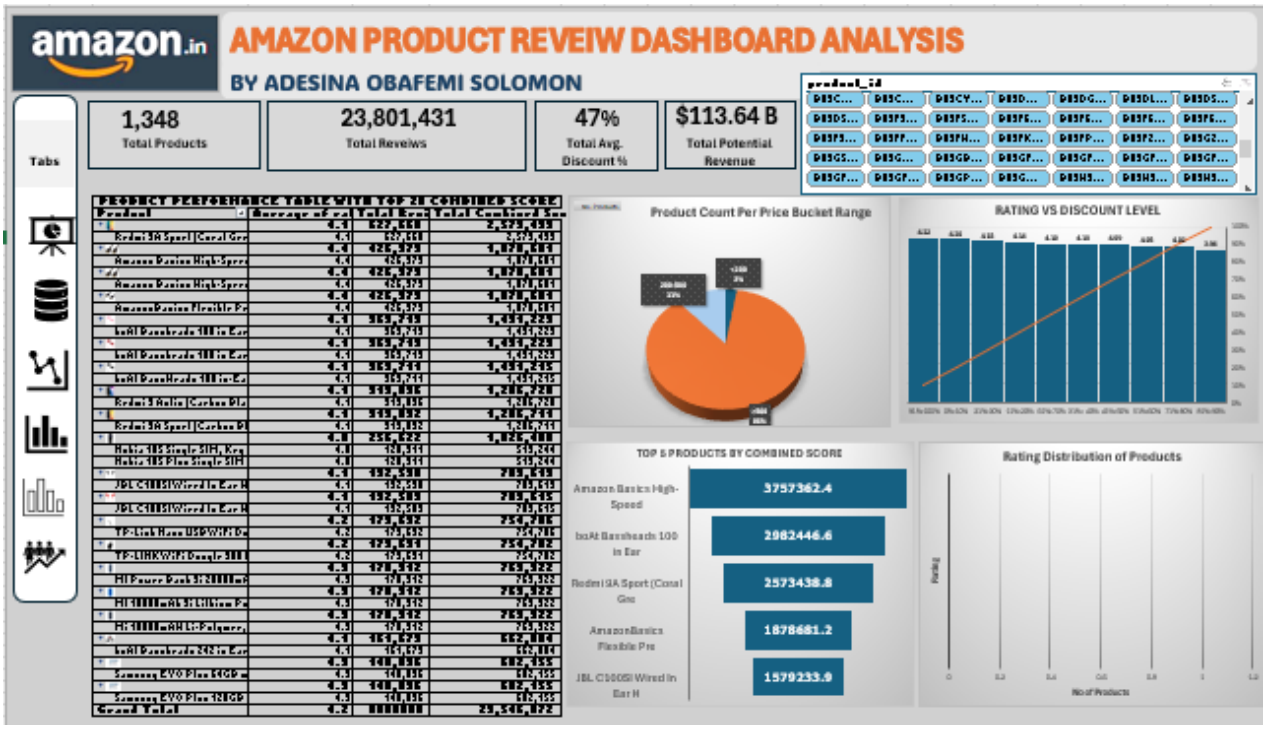
5 Additional Professional Considerations

✓ All erroneous or missing values in critical fields (rating, rating_count, actual_price, etc.) were removed to maintain dataset reliability.
✓ Data types were enforced: numbers as Decimal/Whole Number, text as String, ensuring formula and visualization accuracy.
✓ Power Query transformations were applied to automate and structure the dataset dynamically, feeding directly into the Data Model for advanced analysis.
✓ The cleaned dataset supported an interactive Excel dashboard with KPI cards, charts, and slicers for actionable insights.

6 Conclusion

The data cleaning process transformed the raw Amazon dataset into an analysis-ready, reliable foundation for business insights, pricing strategies, product reviews evaluation, and performance measurement. The cleaning steps, combined with engineered fields, ensured the final deliverables met professional analytical standards and supported clear decision-making for stakeholders and assessors.


amazon.in
AMAZON PRODUCT REVEIW DASHBOARD ANALYSIS
BY ADESINA OBAFEMI SOLOMON
1,348
Total Products
23,801,431
Total Reveiws
47%
Total Avg. Discount % By Category
$113.64 B
Total Potential Revenue
Tabs
Products By Category
Avg Discount By Category
Total Potential Revenue by Product Catefory
Product Count Per Price Bucket Range
Top 3 Total Reviews By Category
Avg. Price vs Discounted Price by Category
AVERAGE RATING VS DISCOUNT
Avg. Rating
Discount %


amazon.in
AMAZON PRODUCT REVEIW DASHBOARD ANALYSIS
BY ADESINA OBAFEMI SOLOMON
Tabs
1,348
Total Products
23,801,431
Total Reveiws
47%
Total Avg. Discount %
$113.64 B
Total Potential Revenue
PRODUCT PERFORMANCE TABLE WITH TOP 20 COMBINED SCORE
Product
Average of rating
Total Reviews
Total Combined Score
Grand Total
4.2
29,545,872
Product Count Per Price Bucket Range
RATING VS DISCOUNT LEVEL
TOP 5 PRODUCTS BY COMBINED SCORE
Amazon Basics High-Speed
3757362.4
boAt Bassheads 100 in Ear
2982446.6
Redmi 9A Sport (Coral Gre
2573438.8
AmazonBasics Flexible Pre
1878681.2
JBL C100SI Wired In Ear H
1579233.9
Rating Distribution of Products
Rating
No of Products