

Amazon E-Commerce Sales Performance and Profitability Analysis

Sector: Online Retail /E-Commerce

Group :2 | Members: Mayank Pillai, Kunal Vats, Shaurya Sharma, Sahil Chand, Praanshu Ranjan, Anshu Yadav.

Mentor: Satyaki Das | Newton School of Technology.



Business Context & Problem Statement

Sector Context

- Amazon operates in a high-competition e-commerce environment where product visibility, pricing, and ratings directly influence sales outcomes.
- Sellers and category managers need data-backed strategies to optimise product performance.

Problem Statement

What product-level factors — ratings, pricing, promotions, and availability — are most associated with higher estimated sales and revenue on Amazon?

Project Objective

- Analyse ~10,000+ product records to identify observable patterns driving sales performance.
- Deliver an interactive dashboard to support ongoing, data-driven decision-making.



Data Engineering

Data Source

- Dataset: Amazon Products Sales Dataset (42K+ Items – 2025), sourced from Kaggle
- Structure: Product-level data — one row per product listing
- Size: ~42,000+ product records across multiple categories

Key Cleaning Steps

- Missing Values: Removed or imputed missing ratings, reviews, and price fields
- Outliers: Filtered/capped extreme price, rating, and sales values
- Transformations: Converted fields to clean numeric formats

Feature Engineering (New Fields Created)

- Sales_Clean | Estimated_Revenue | is_sponsored | has_coupon | in_stock | Grouped_Rating

Tools Used

- Google Sheets (formulas, filters, pivot tables)

Assumption

- Sales and Estimated Revenue are computed estimates — not verified transaction records



KPI & Metrics Framework

KPI	Definition	Business Relevance
Estimated Revenue	$\text{Sales_Clean} \times \text{Final Price}$	Measures product-level earning potential
Sales_Clean	Cleaned unit sales volume	Indicates demand and conversion
Average Rating	Mean product rating score	Proxy for customer satisfaction
Sponsored %	$\text{Sponsored listings} \div \text{Total listings}$	Measures paid visibility impact
Stock Availability	In-stock vs. out-of-stock share	Links inventory to revenue risk
Best Seller Badge	Binary flag for badge presence	Identifies platform-endorsed products

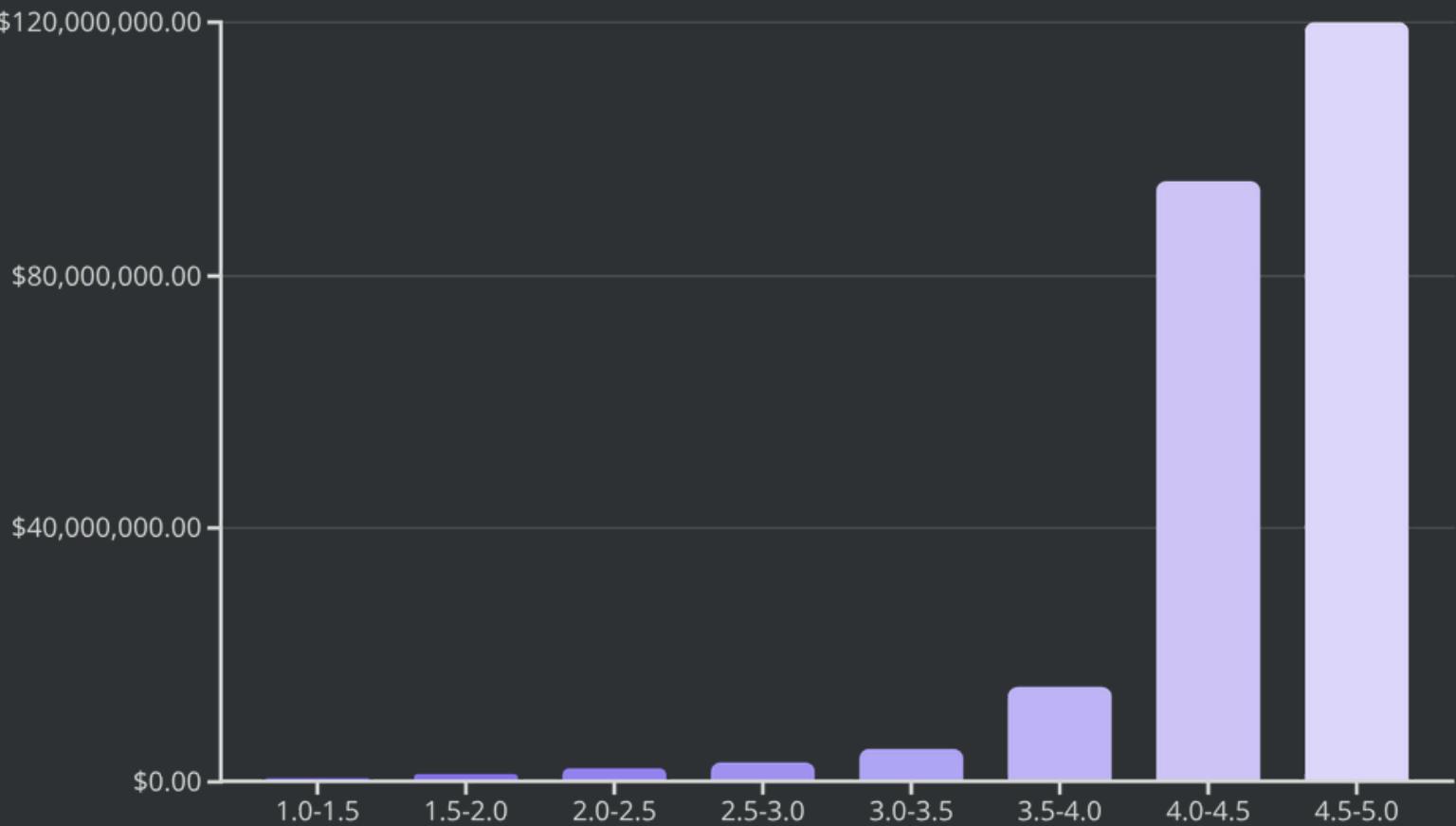
Key Formula

Estimated Revenue = Sales_Clean × Final Price

Key Insight #1 – Ratings Are the Strongest Revenue Differentiator

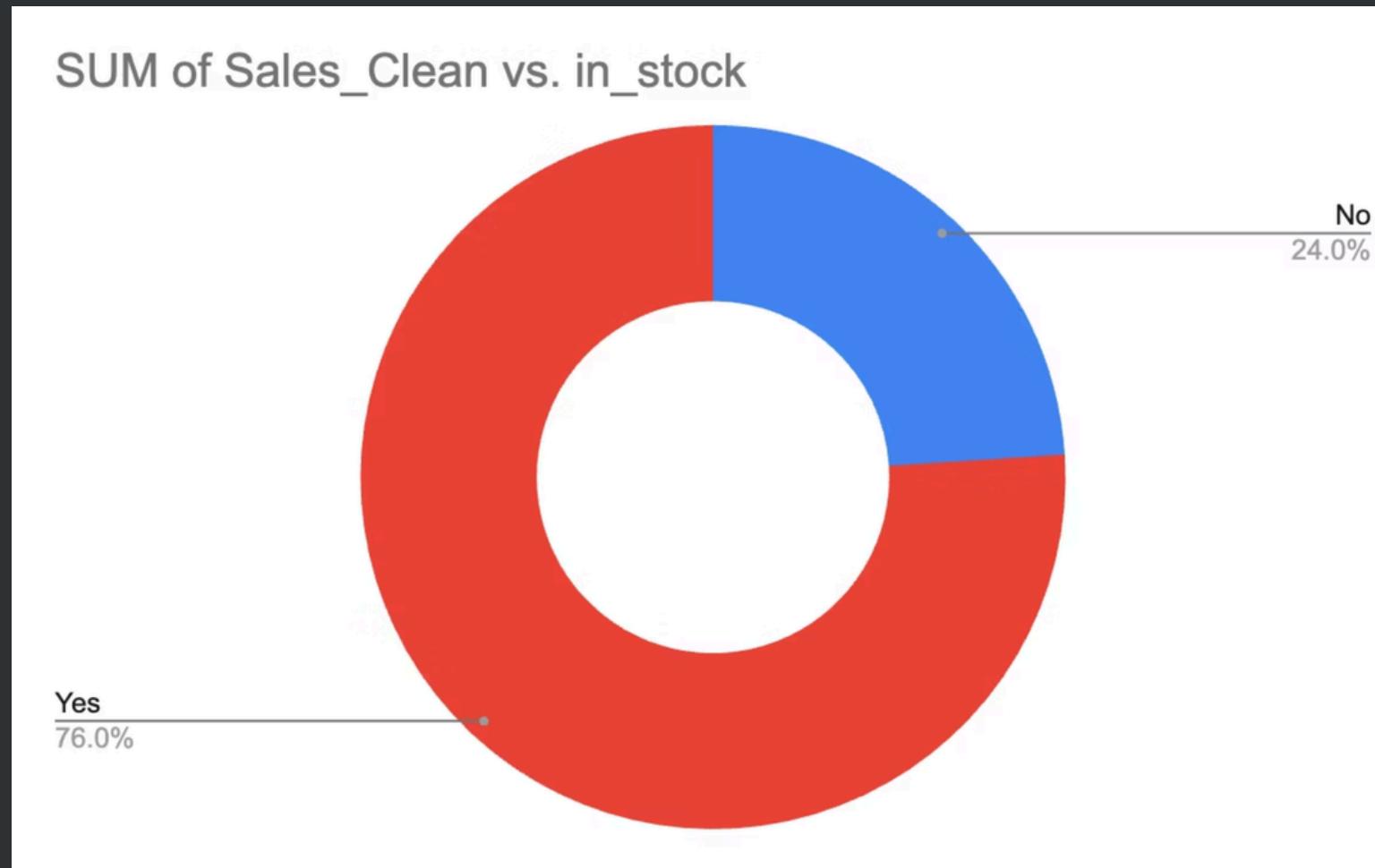
Observation

- Products rated 4.5–5.0 show the highest sum of Estimated Revenue in the dataset
- Products rated 4.0–4.5 rank second in total Estimated Revenue
- Products rated below 3.5 show minimal Estimated Revenue contribution
- The revenue gap between low-rated (below 3.5) and high-rated (above 4.0) products is visually dramatic

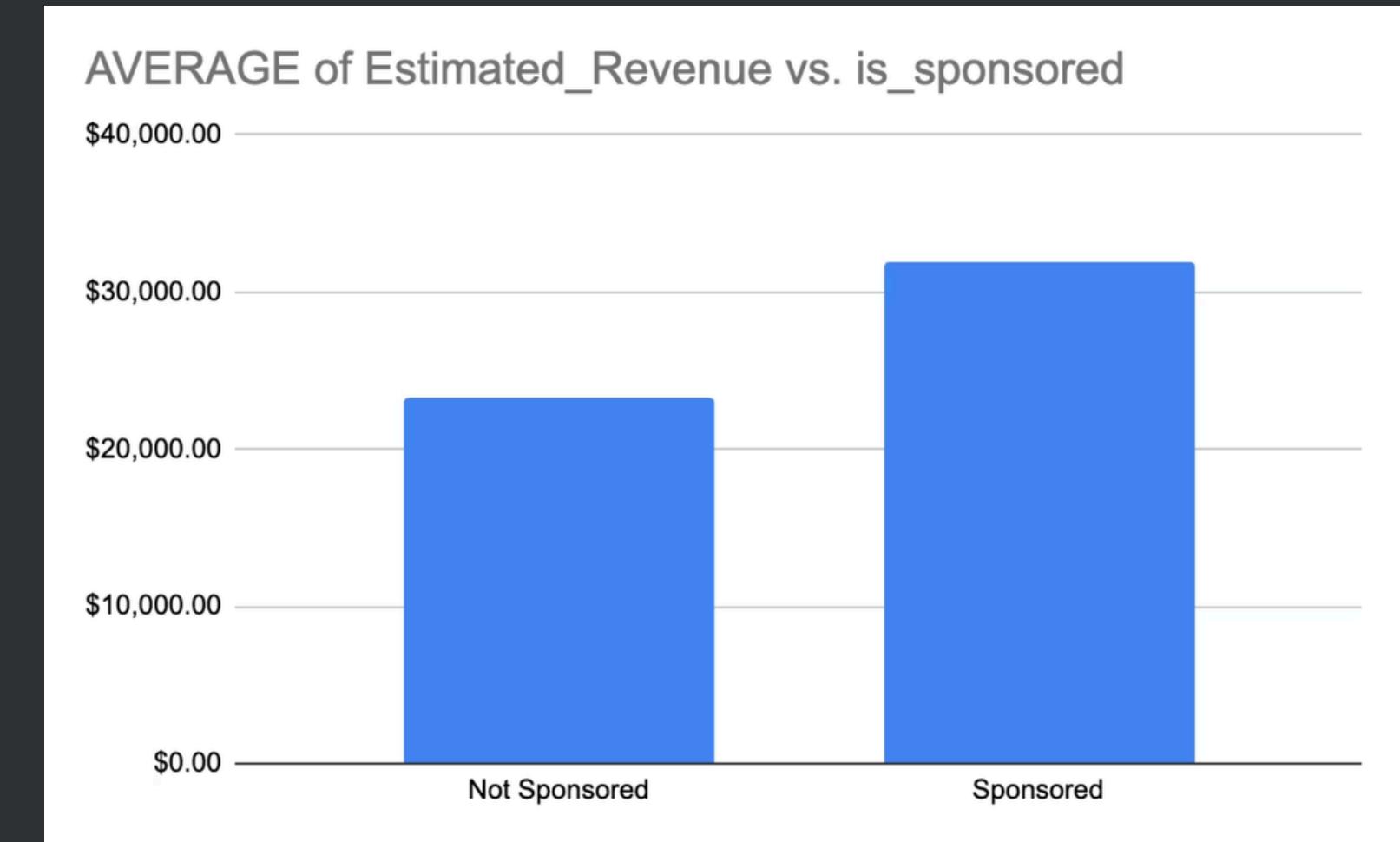


Higher product ratings are strongly associated with higher estimated revenue — though causation cannot be confirmed from this dataset alone

Key Insight #2 – Stock Availability & Sponsored Listings Drive Performance



Finding A — Stock Availability: ~76% of total Sales_Clean volume comes from in-stock products; ~24% from out-of-stock listings. Interpretation: Stock unavailability appears associated with significantly lower sales volume



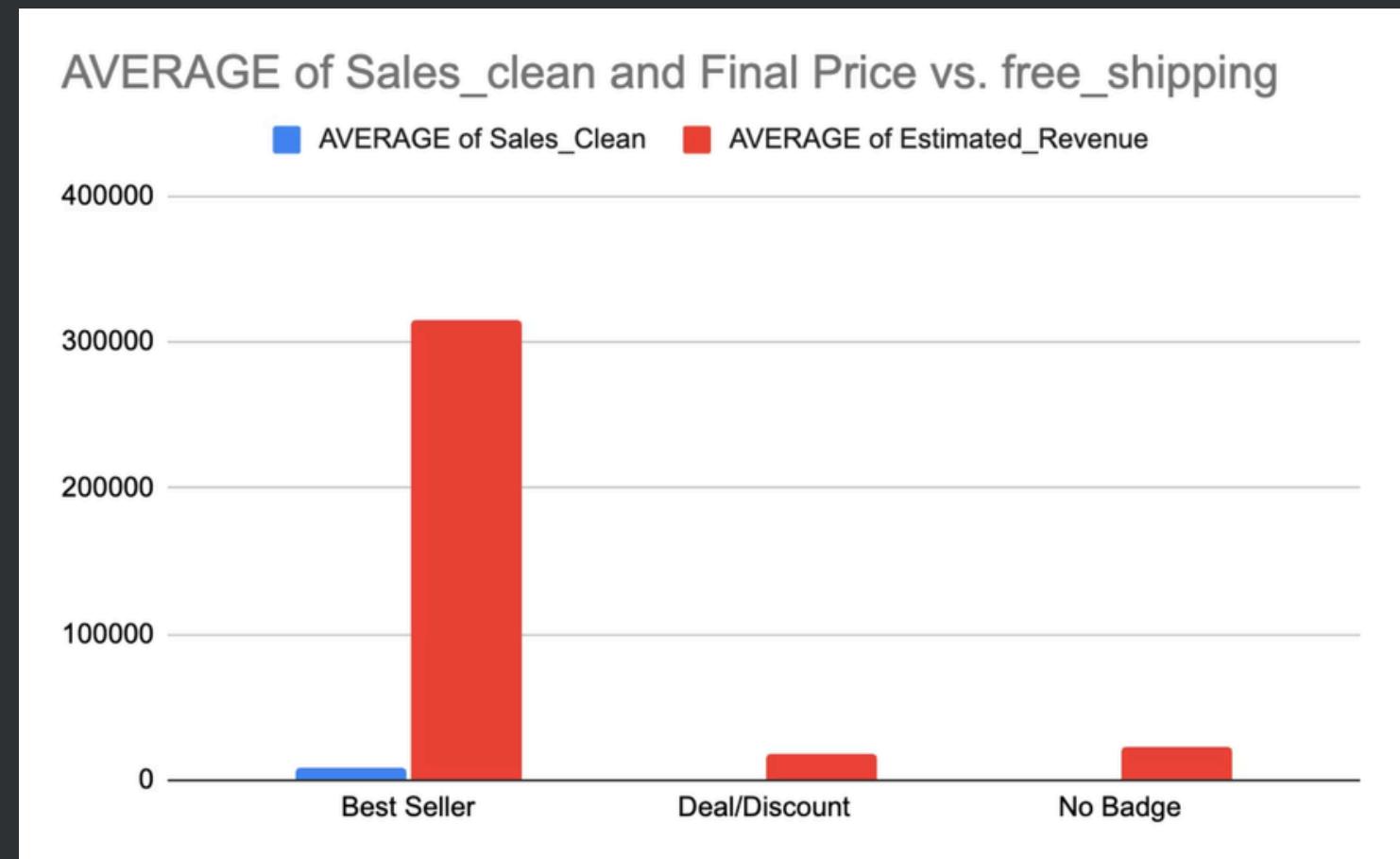
Finding B — Sponsorship: Sponsored products show higher average Estimated Revenue (\$32,000) compared to non-sponsored products (\$23,000) — both estimated figures. Interpretation: Sponsored placement is associated with improved average revenue per product

Key Insight #3 – Best Seller Badge Signals Outsized Revenue Concentration

Observation

- Best Seller products show substantially higher average Estimated Revenue (~\$315,000) than Deal/Discount (~\$15,000) or No Badge (~\$20,000) products
- Average Sales_Clean (unit volume) is comparatively modest even for Best Sellers, suggesting Best Sellers carry higher price points
- Deal/Discount and No Badge products perform at a similar, much lower revenue level

Best Seller status is associated with a significant concentration of estimated revenue — suggesting platform badge signals matter more than discount strategy for revenue performance

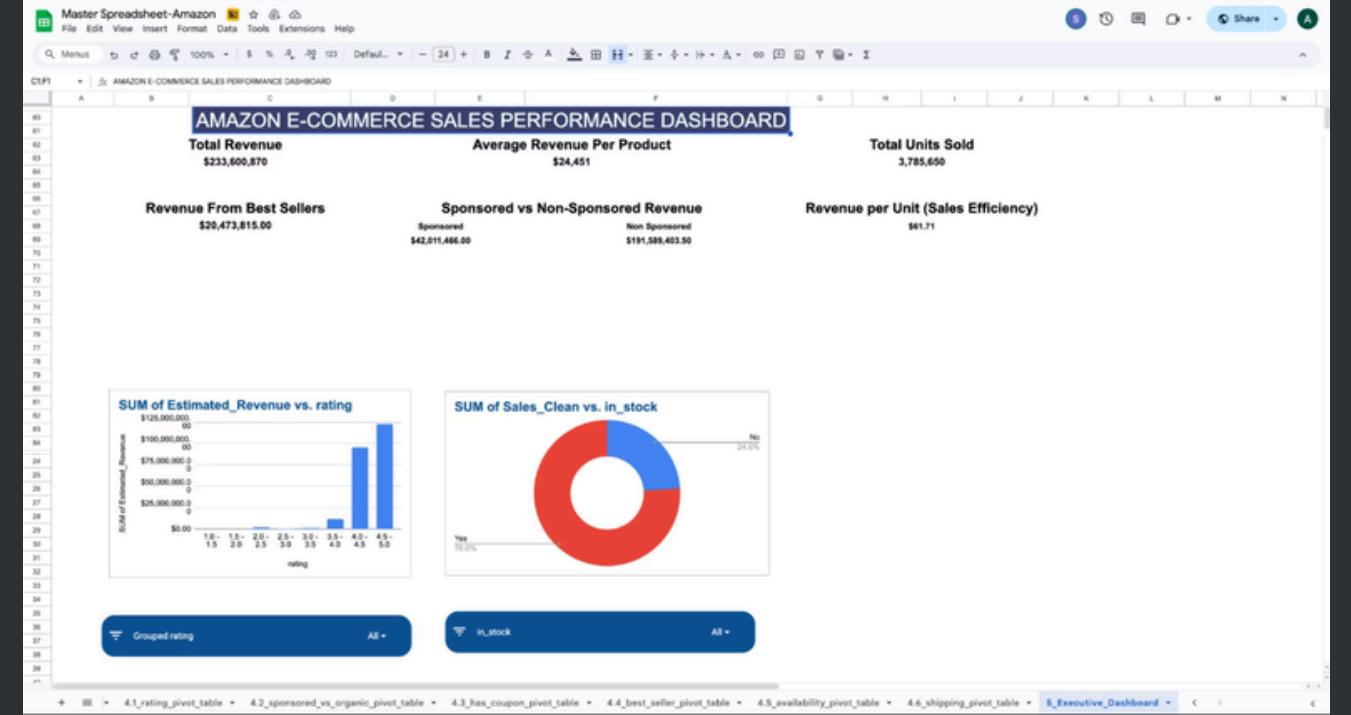


ⓘ Caution: Best Seller badge assignment methodology is not confirmed in this dataset — results should be treated as observational

Executive Dashboard Walkthrough

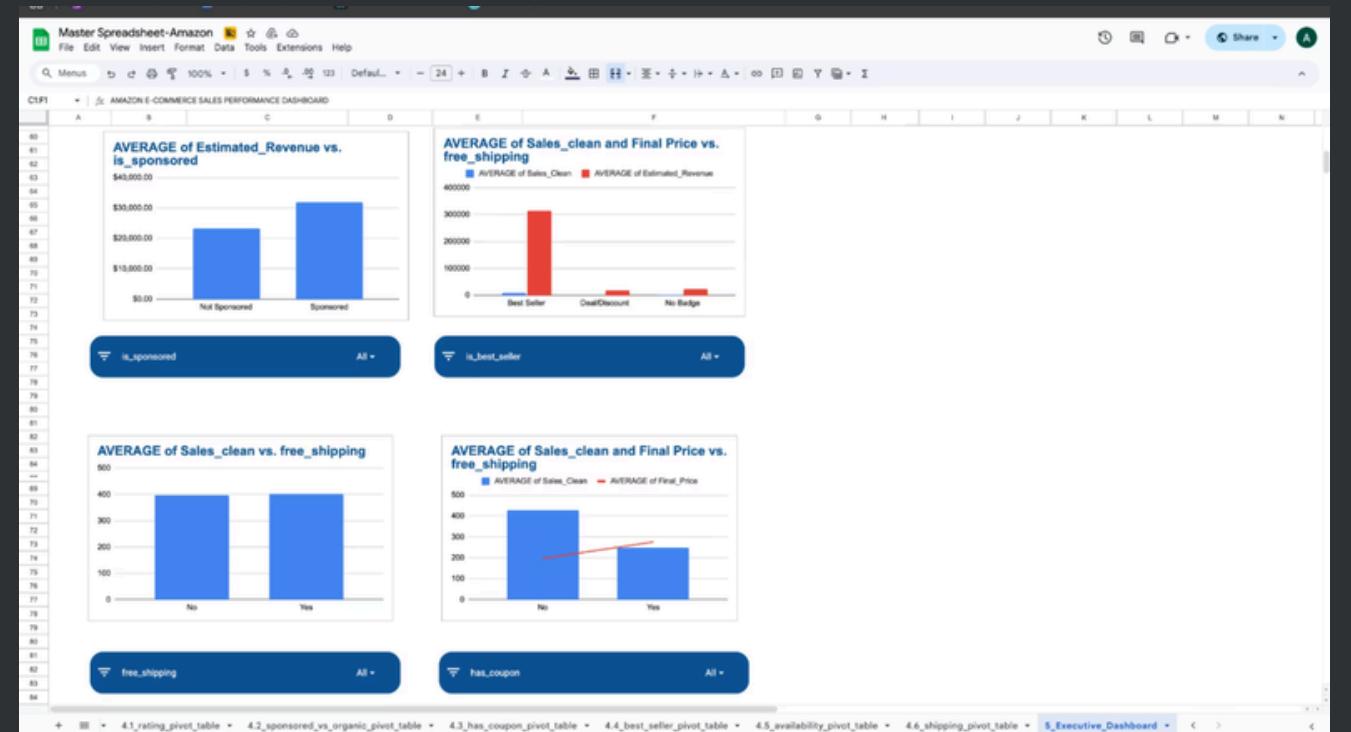
TOP PANEL — KPI Summary Cards

- Total Estimated Revenue: \$233,600,870
- Average Revenue Per Product: \$24,451
- Total Units Sold: 3,785,650
- Revenue per Unit (Sales Efficiency): \$61.71
- Revenue from Best Sellers: \$20,473,815



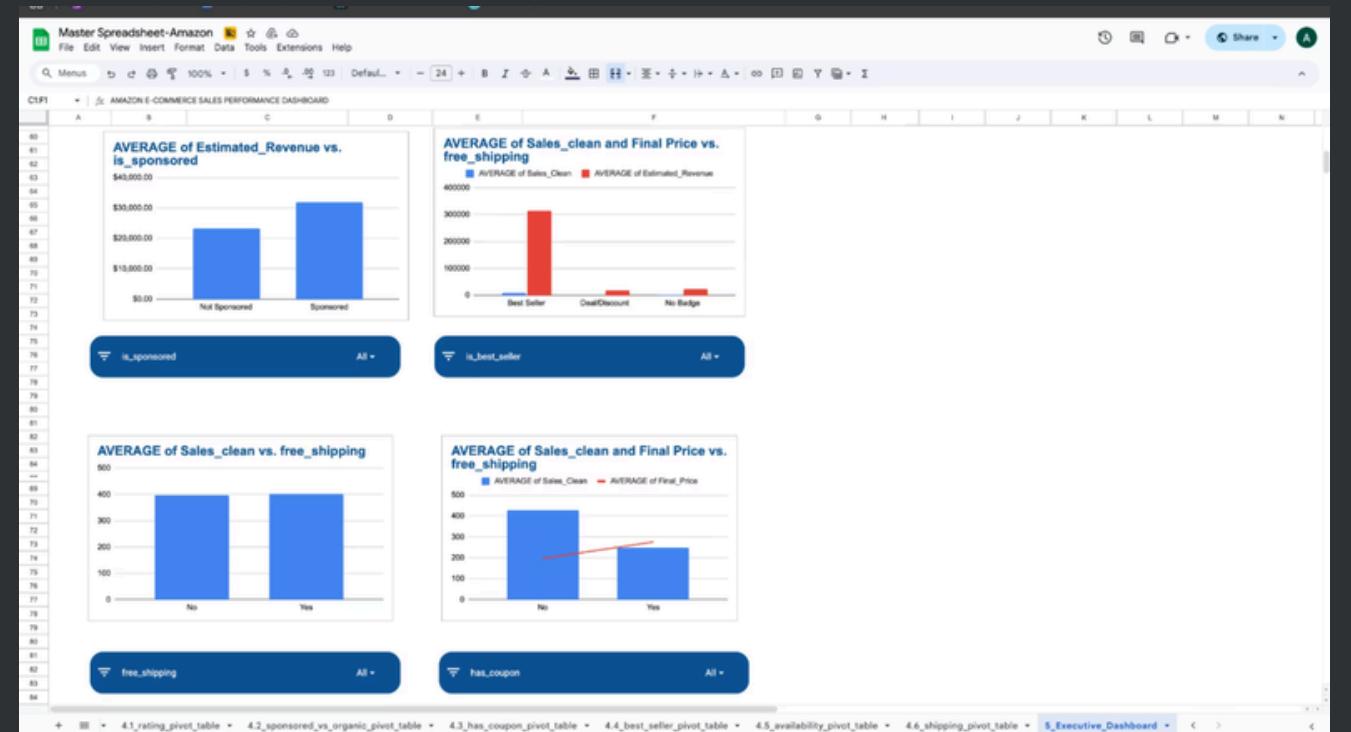
MIDDLE PANEL — Visual Analysis

- Rating vs. Estimated Revenue (bar chart)
- Stock Availability vs. Sales_Clean (donut chart)



BOTTOM PANEL — Comparison Drilldowns

- Sponsorship, Best Seller, Coupon, and Shipping comparisons
- Interactive slicers: Rating group, in_stock, is_sponsored, free_shipping, has_coupon



Recommendations & Estimated Business Impact

1	Prioritise high-rated products (4.0–5.0) in listings and ads	Rating drives estimated revenue	Improved conversion rates
2	Expand sponsored placements for top-performing SKUs	Sponsored products show higher avg. revenue	Up to 15–20% efficiency gain*
3	Maintain robust stock availability across categories	76% of sales volume from in-stock items	Reduced revenue loss risk by 10–15%*
4	Focus inventory and promotion budgets on Best Seller products	Best Sellers show outsized revenue concentration	Faster revenue concentration
5	Apply free shipping selectively based on margin analysis	Free shipping influences pricing and conversion	Improved conversion rates

*Estimated impact figures sourced from Impact Estimation section of Final Project Report

Limitations, Future Scope & Conclusion

Limitations

- Dataset is web-scraped — may contain missing, inconsistent, or duplicated entries
- Sales and Estimated Revenue are computed approximations, not verified transaction records
- Snapshot data only — long-term trends and real-time demand cannot be confirmed
- Customer demographics and purchase intent are not captured in this dataset

Future Scope

- Time-series forecasting for demand prediction by product category
- Customer segmentation analysis using purchase history data
- Integration of marketing spend data to calculate true ROI of sponsorship

Conclusion

 *This project identified observable associations between ratings, sponsorship, stock availability, and best-seller status with estimated sales performance on Amazon. The resulting dashboard provides a reusable framework for ongoing, data-driven product strategy.*



Limitations

Dataset quality, approximations, snapshot data, missing demographics.



Future Scope

Time-series, customer segmentation, marketing ROI.



Conclusion

Identified associations, framework for product strategy.