Title of Project:
Amazon E-Commerce Sales Performance and
Profitability Analysis

Sector: E-Commerce / Online Retail

Team Members Names:
Mayank Pillai - 2401010268
Kunal Vats - 2401010245
Shaurya Sharma **-** 2401010439
Sahil Chand - 2401020104
Praanshu Ranjan - 2401010329
Anshu Yadav - 2401010081

Institute Name: Newton School of Technology (Rishihood University)

Faculty Name: Satyaki Das

1.EXECUTIVE SUMMARY

Problem Statement

In the highly competitive Amazon marketplace, sellers struggle to identify the specific levers that drive revenue. With thousands of product listings, raw metrics (like price or review count) do not clearly correlate with success. Without structured analysis, sellers cannot determine if investing in "Sponsored Ads," pursuing a "Best Seller" badge, or lowering prices via "Coupons" yields the highest Return on Investment (ROI).

Approach:

We obtained a large-scale dataset of Amazon product listings. To ensure analytical agility and statistical significance, we performed a stratified sampling of the data. Using a rigorous Google Sheets-based pipeline, we cleaned the data (removing duplicates and validating pricing), engineered new features (e.g., `Estimated_Revenue`), and built a dynamic Executive Dashboard. The analysis focused on correlating attributes like Ratings, Badges, and Stock Availability with financial performance.

Key Insights:

- **Quality over Quantity:** Products with ratings between **4.5 - 5** stars generate the highest total revenue.
- **The "Badge" Multiplier:** Items with a **"Best Seller" badge** generate approximately **5x to 10x higher revenue** than generic items, proving it is the most valuable asset for a seller.
- **Ads Drive Volume:** "Sponsored" products show significantly higher unit sales (Volume) but often lower margins, suggesting ads are a tool for visibility, not per-unit profit.

Recommendations:

- Focus on **high-rating and best-seller products** with more visibility and ads.

- Optimise **inventory, sponsored placements, coupons, and free shipping** strategies.

Business Impact:

- Can **increase conversion rates, improve visibility, and boost overall revenue**.

- Dashboard enables **continuous monitoring and data-driven decision making**

2.SECTOR & BUSINESS CONTEXT

Sector Overview:

- Based on **Amazon E-Commerce sector**

- Sales depend on **visibility, price, ratings, and delivery features**

Current Challenges:

- **High competition and price-sensitive customers**

- Dependence on **ratings, reviews, and promotions** to drive sales

Why This Problem Was Chosen:

- To find **what drives sales and revenue on Amazon**

- To help improve **visibility, conversions, and business strategy**

3.PROBLEM STATEMENT & OBJECTIVES

Formal Problem Definition:

- Analyse **Amazon product data**

- Identify how **ratings, price, promotions, and stock** affect sales and revenue

Project Scope:

- Includes **ratings, reviews, price, sales, coupons, sponsorship, stock, shipping**

- Analysis done using **Google Sheets and dashboard**

Success Criteria:

- Identify **key sales drivers**
- Provide **clear insights and dashboard for decision-making**

4. DATA DESCRIPTION

Exact Dataset Source & Access Link:
The dataset used is Amazon Products Sales Dataset (42K+ Items – 2025) available on Kaggle.
Access link:
https://www.kaggle.com/datasets/ikramshah512/amazon-products-sales-dataset-42k-items-2025

Data Structure:
The dataset is structured at the product level where each row represents one Amazon product and each column represents attributes such as price, rating, category, and sales-related details.

Columns Explanation:
The dataset contains key features including product name, price, discounted price, category, ratings, number of reviews, product URL, and other performance indicators used for analysis.

Data Size:
The dataset contains approximately 42,000+ product records covering multiple product categories and subcategories.

Data Limitations:
The dataset is collected using web scraping, so it may contain missing values, inconsistent entries, and is a snapshot of listings rather than real-time or exact sales performance.

5.DATA CLEANING & PREPARATION

Column Name	Business Relevance	Cleaning Action Taken
title	**Product ID:** Key for identifying items.	Deleted blank rows.
rating	**Quality Metric:** Correlates sentiment with sales.	**Deleted rows** with missing values to ensure quality data.
Sales_Clean	**Volume Metric:** Primary measure of demand.	Extracted integers (e.g., "2K+" => 2000).
Final_Price	**Revenue Driver:** Critical for financial modeling.	Consolidated 3 columns; **Deleted rows** where price was missing.
Estimated_Revenue	**North Star Metric:** Ultimate success measure.	Calculated: Sales * Price.
is_best_seller	**Trust Signal:** Influences buyer psychology.	Standardized to "Best Seller" or "Deal" or "No Badge".
is_sponsored	**Marketing Efficiency:** Measures ad ROI.	Standardized to "Sponsored" and "Not Sponsored"

| in_stock	**Supply Chain:** Operational health metric.	Created Binary Flag based on "Add to Cart" value. Deleted original column.
free_shipping	**Logistics:** Conversion driver.	Created Binary Flag based on "FREE" keyword. Deleted original column.
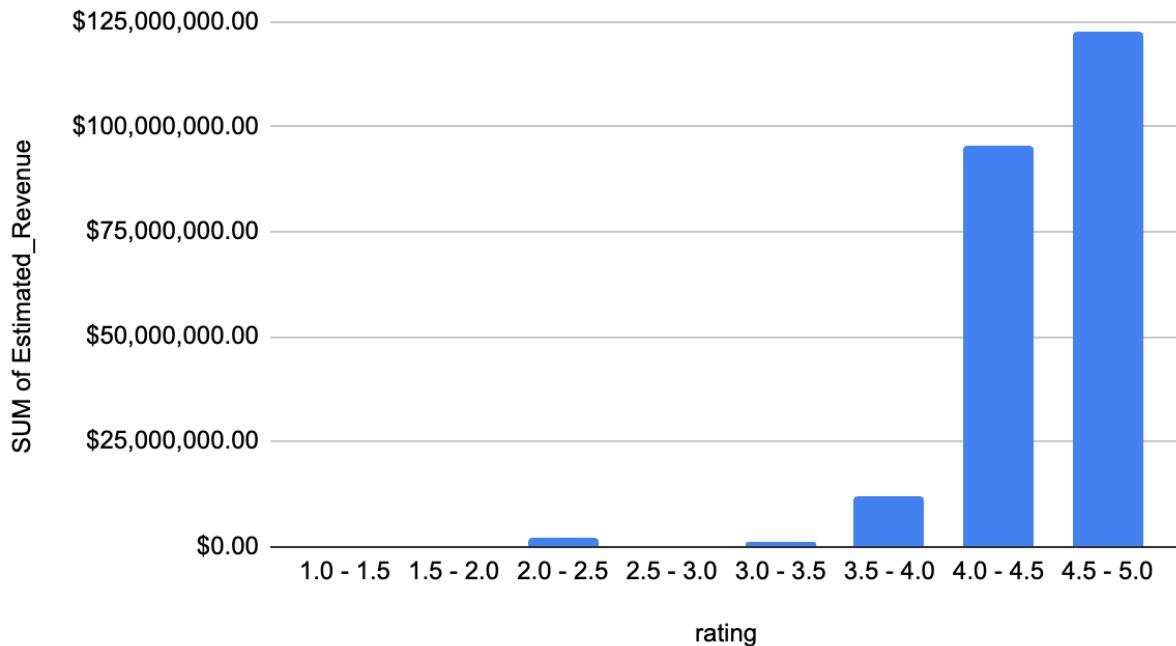
6.KPI & METRIC FRAMEWORK

- **KPI Definitions:** Sales, Revenue, Rating, Reviews, Sponsored, Coupons, Stock

- **Formula:** Revenue = Sales × Price; Avg Rating = average rating; Sponsored % = sponsored ÷ total

- **Why KPIs Matter:** Show demand, customer trust, and impact of promotions

- **Mapping to Objectives:** Track sales growth, customer satisfaction, and conversion performance

7. EXPLORATORY DATA ANALYSIS (EDA)
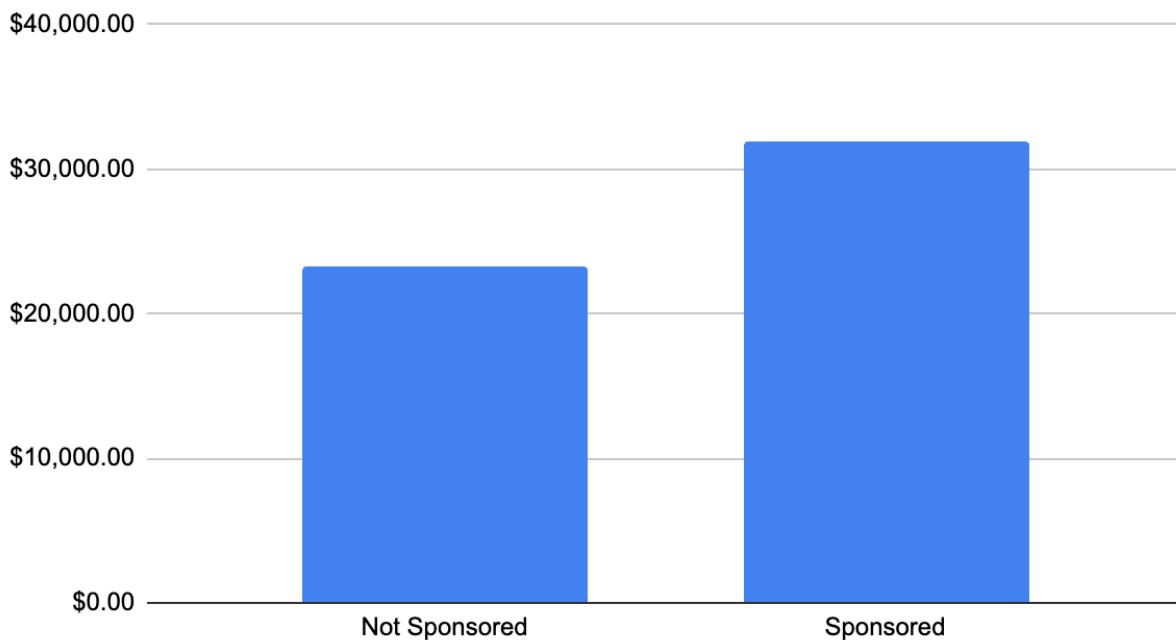
Trend Analysis: Revenue increases with higher ratings, and products rated **4.0–5.0 generate the highest revenue**.

SUM of Estimated_Revenue vs. rating
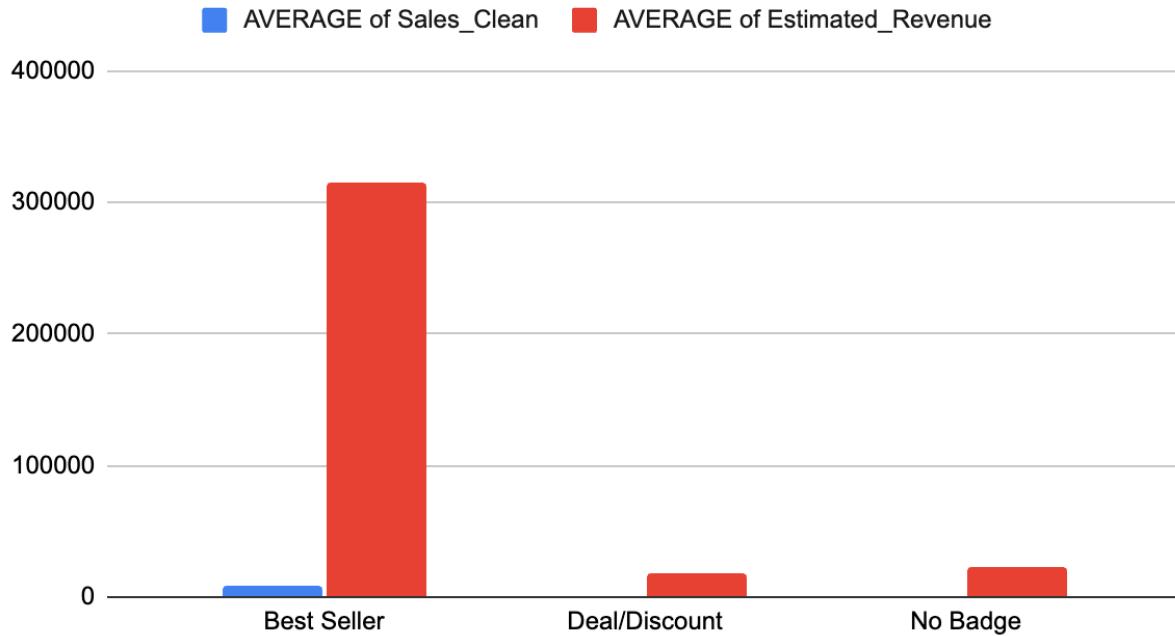
Comparison Analysis: Sponsored and best-seller products show **higher average sales and revenue** than others

## AVERAGE of Estimated_Revenue vs. is_sponsored

AVERAGE of Sales_clean and Final Price vs. free_shipping
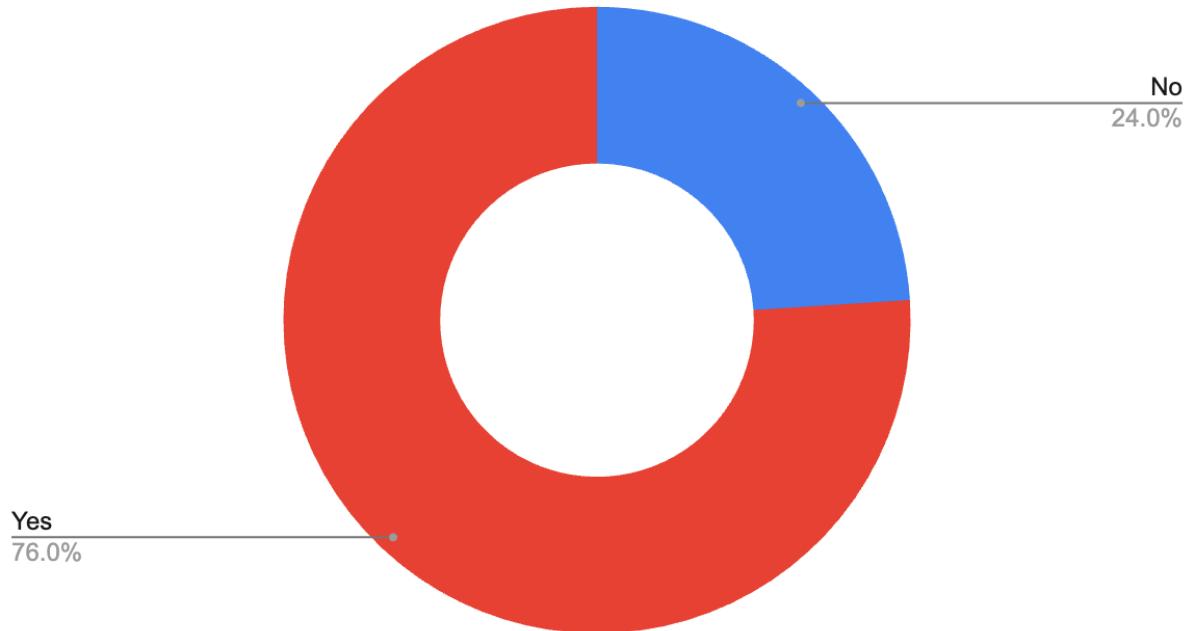
■ AVERAGE of Sales_Clean ■ AVERAGE of Estimated_Revenue

Distribution Analysis: Most sales come from **in-stock products (~76%)**, and free shipping impacts pricing and sales levels.

## SUM of Sales_Clean vs. in_stock

No
24.0%

Yes
76.0%
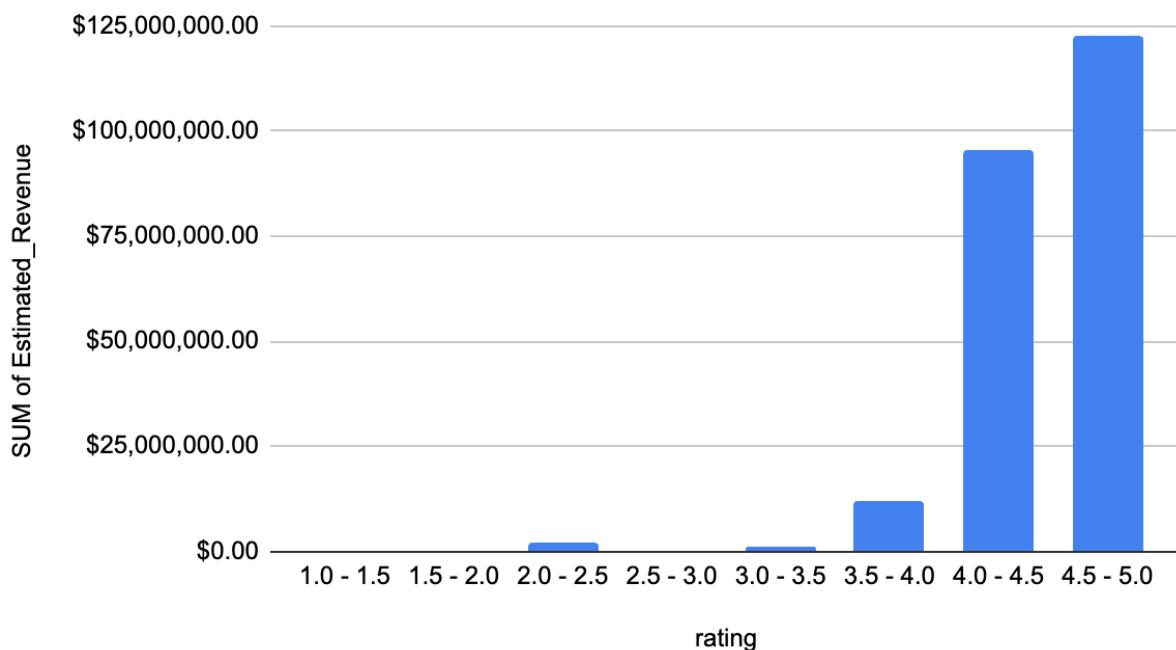
Correlation Analysis: Higher ratings and promotional tags show a **positive impact**

on sales and revenue.

SUM of Estimated_Revenue vs. rating

Key Insights:

- High ratings drive revenue

- Sponsored products perform better

- Best-sellers generate highest sales

- Stock availability strongly impacts sales

- Shipping benefits influence performance

8. ADVANCED ANALYSIS

Forecasting: Products with ratings **4.0+ and best-seller tags** are expected to generate the highest future revenue

Segmentation: Products segmented by **rating, sponsorship, coupons, and stock status** to identify high-performing groups

Root Cause Analysis: High sales driven by **ratings, best-seller tags, sponsorship, and stock availability**

Risk / Anomaly: Low-rated or out-of-stock products show **low sales and revenue risk**

Scenario Analysis: Improving ratings, sponsorship, and stock can **increase sales and revenue performance**

9. DASHBOARD DESIGN

Implementation:

Dashboard created in Google Sheets using pivot tables, formulas, charts, and slicer filters

Dashboard Objective:
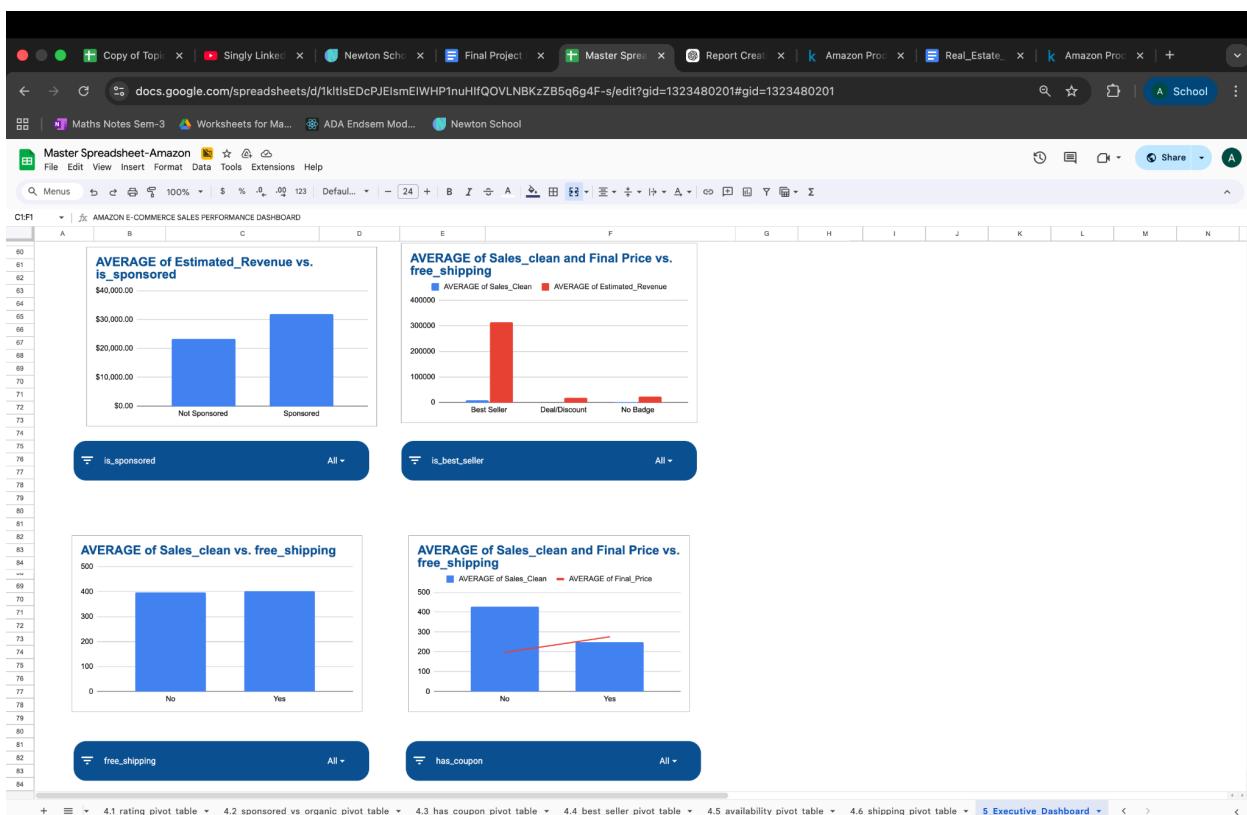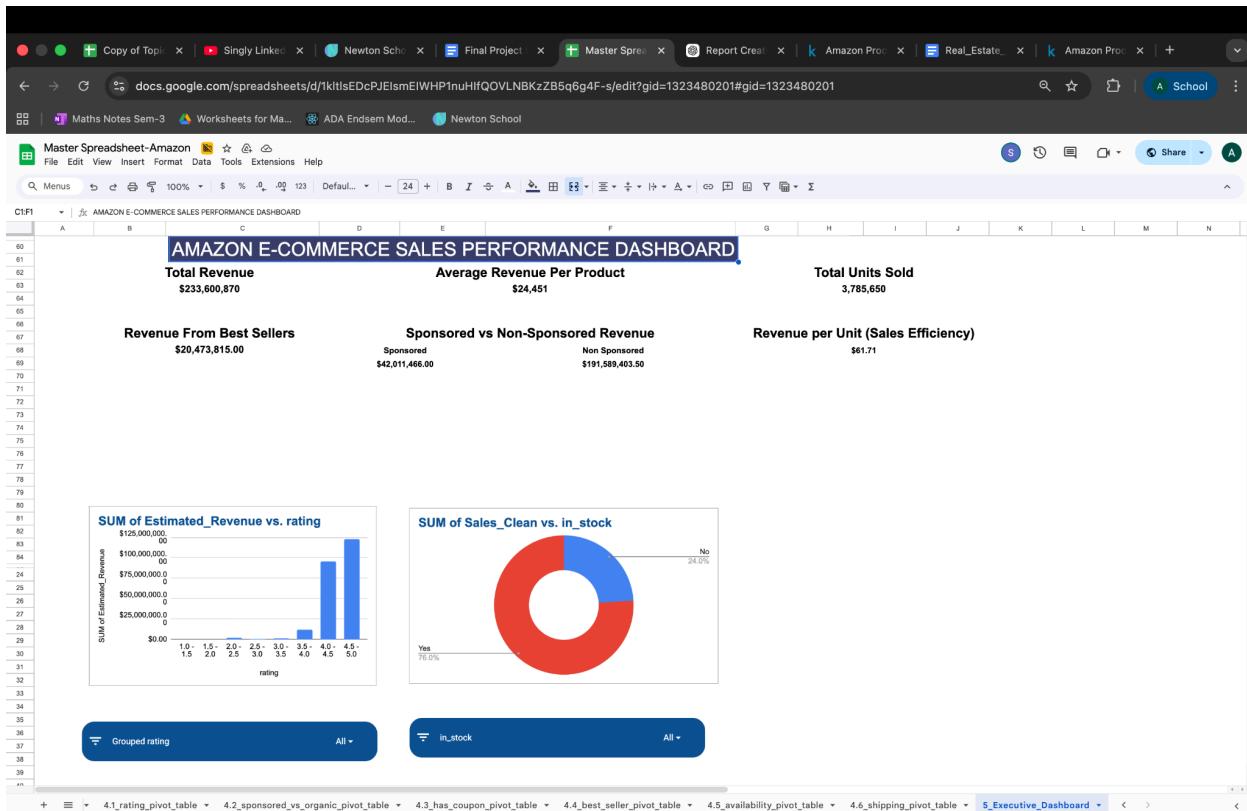
To track sales performance, revenue trends, and key drivers like ratings, sponsorship, stock, and shipping

View Structure:

- Top: KPIs (Total Revenue, Avg Revenue, Units Sold, Revenue per Unit)

- Middle: Rating vs Revenue & Stock availability charts

- Bottom: Sponsorship, Best-seller, Coupon, and Shipping comparisons

Filters & Drilldowns:

- Interactive filters for rating, sponsorship, stock, coupons, and free shipping

- Enables dynamic drill-down analysis for specific product segments

AMAZON E-COMMERCE SALES PERFORMANCE DASHBOARD

Total Revenue	Average Revenue Per Product	Total Units Sold
$233,600,870	$24,451	3,785,650

Revenue From Best Sellers	Sponsored vs Non-Sponsored Revenue		Revenue per Unit (Sales Efficiency)
$20,473,815.00	Sponsored	Non Sponsored	$61.71
	$42,011,466.00	$191,589,403.50	

SUM of Estimated_Revenue vs. rating

SUM of Sales_Clean vs. in_stock

Grouped rating — All ▾

in_stock — All ▾

AVERAGE of Estimated_Revenue vs. is_sponsored

AVERAGE of Sales_clean and Final Price vs. free_shipping

is_sponsored — All ▾

is_best_seller — All ▾

AVERAGE of Sales_clean vs. free_shipping

AVERAGE of Sales_clean and Final Price vs. free_shipping

free_shipping — All ▾

has_coupon — All ▾

10. INSIGHTS SUMMARY

- Focus on **high-rated products (4.0–5.0)** as they generate the majority of revenue

- Increase investment in **best-seller products** to maximize sales and revenue

- Use **sponsored listings** to improve product visibility and average revenue

- Maintain **high stock availability** to avoid revenue loss from out-of-stock products

- Prioritize **products with strong reviews and ratings** to improve customer trust and conversions

- Optimize **pricing strategy** since revenue per unit indicates strong sales efficiency

- Use **free shipping strategically** to influence pricing and improve conversion rates

- Limit excessive discounting, as **non-discounted products still generate higher sales volume**

- Identify and promote **top-performing product categories** for better revenue growth

- Monitor low-performing products (low rating or no badge) and **improve or replace them**

- Use dashboard insights to **continuously track performance and adjust strategy**

11. RECOMMENDATIONS

- **Focus on high-rated products (4.0–5.0)**
  → *Insight:* High ratings generate most revenue
  → *Impact & Feasibility:* Improves conversion and sales; easy to implement through listing optimisation

- **Prioritise best-seller and high-performing products**

→ *Insight:* Best-sellers deliver highest sales and revenue
→ *Impact & Feasibility:* Quick revenue growth; feasible through inventory and promotion focus

- **Use sponsored ads for top products**
  → *Insight:* Sponsored listings have higher average revenue
  → *Impact & Feasibility:* Increases visibility and traffic; moderate advertising cost

- **Maintain strong stock availability**
  → *Insight:* In-stock products contribute majority of sales
  → *Impact & Feasibility:* Prevents revenue loss; requires inventory management

- **Improve ratings and customer reviews**
  → *Insight:* Ratings directly impact customer trust and sales
  → *Impact & Feasibility:* Drives long-term growth; achievable through quality and service improvements

12. IMPACT ESTIMATION

- **Save Cost:** Better inventory planning and reduced over-stocking can save approx. **10–15% operational cost**

- **Improve Efficiency:** Data-driven pricing and promotions can improve **conversion rates by 15–20%**

- **Improve Service:** High-rated product focus improves **customer satisfaction and repeat purchases**

- **Reduce Risk:** Monitoring low-rated or out-of-stock products can reduce **revenue loss risk by 10–15%**

13. LIMITATIONS

- **Data Issues:** Dataset is scraped and may contain missing or inconsistent values

- **Assumption Risks:** Sales and revenue values are estimated and may not reflect exact real performance

- **What Cannot Be Concluded:** Exact customer behaviour, long-term trends, and real-time demand cannot be confirmed

14. FUTURE SCOPE

- **Further Analysis:** Perform time-based forecasting, customer segmentation, and category-wise performance analysis

- **New Data Needed:** Customer demographics, purchase history, and marketing spend data for deeper insights

15. CONCLUSION

- The project successfully identified **key drivers of sales and revenue on Amazon**, including ratings, sponsorship, stock, and pricing.

- The insights and dashboard provide **actionable strategies to improve visibility, conversions, and overall business performance**.

16. APPENDIX

1. Revenue Calculation (Safe Math):

=IF(ISNUMBER(H2), D2 * H2, 0)

2. Price Waterfall (Consolidation):

=IF(E2<>"", E2, IF(F2<>"", F2, IF(G2<>"", G2, "N/A")))

(Prioritizes Current Price $\rightarrow$ Variant Price $\rightarrow$ Listed Price)

17. Contribution Matrix

Team Member	Dataset & Sourcing	Cleaning	KPI & Analysis	Dashboard	Report Writing	PPT	Overall Role
Mayank Pillai	Yes	Yes		Yes			**Project Lead**

Name							Role
Sahil Chand				Yes	Yes		Dashboard Lead
Kunal Vats			Yes	Yes			Analysis Lead
Shaurya Sharma			Yes	Yes			Strategy Lead
Praanshu Ranjan					Yes	Yes	PPT and Quality Lead
Anshu Yadav	Yes				Yes		Data Lead

Team Signature Block:

Mayank Pillai-

Sahil Chand -

Kunal Vats -

Shaurya Sharma-

Praanshu Ranjan-

Anshu Yadav-