1. Data Visualization and Analytics

Amazon E-Commerce Analytics
E-Commerce

Team Members:

- Agrima Ojha (2401010047)
- Suryansh Chattree(2401010466)
- Khyati Batra (2401010229)
- Meghna Nair (2401010274)
- Raghvendra Singh(2401010367)
- Deepanshu Chaudhary(2401010145)

Institute:

Newton School of Technology

Faculty: Aayushi Vashishth

2. Problem Statement(Executive Summary)

The dataset contains **9,999 electronics products** with an average rating of **4.45**. The average monthly demand is **1,652 units**. Customer engagement is uneven, with an average of **5,769 reviews** but a median of only **860**, indicating that a small number of products dominate customer attention. The average product price is **$163**, showing the presence of premium-priced products.

Approach

We analyzed a real-world Amazon electronics dataset (42,000+ products) using Google Sheets. The project focused on:

- Data cleaning and preprocessing
- Handling missing values and inconsistencies
- Creating structured variables for analysis
- Identifying relationships between pricing, discounts, ratings, reviews, and promotional tags

Key Insights

- **Customer Trust Drives Demand**
  Products with higher **ratings** and a larger **number_of_reviews** tend to show higher values in **bought_in_last_month**, indicating strong customer preference for well-reviewed products.

- **Price Competitiveness Influences Purchases**
  Products where the **current/discounted_price** is significantly lower than the

listed_price (effective discount) show higher purchase activity, highlighting the importance of competitive pricing.

- **Buy Box Availability is a Major Conversion Factor**
  Listings with **buy_box_availability = True** consistently demonstrate higher recent purchase volumes, indicating its strong impact on conversion likelihood.

- **Visibility Tags Improve Performance**
  Products marked with **best_seller_flag** or **sponsored_flag** tend to receive higher demand, suggesting that platform visibility plays a critical role in sales performance.

- **Promotional Incentives Support Demand**
  Products with **is_couponed = TRUE** show improved purchase activity, indicating that additional promotional incentives positively influence customer decisions.

Key Recommendations

- **Optimize pricing strategy** by maintaining competitive price levels and using targeted discounts where necessary.
- **Improve product quality and customer experience** to achieve and maintain ratings above 4.0.
- **Increase product visibility** through Sponsored listings and Best Seller optimization for high-potential products.
- **Improve Buy Box eligibility and inventory availability** to prevent lost sales and maximize conversions.

- **Focus marketing efforts on high-rating products with low visibility** to unlock additional demand.

---

3. Sector & Business Context

a. Sector Overview

The e-commerce electronics sector is highly competitive, with multiple sellers offering similar products on online platforms like Amazon. Customer decisions are strongly influenced by product ratings, reviews, price competitiveness, and platform visibility.

b. Current Challenges
- High competition among sellers
- Difficulty in improving product visibility
- Managing pricing and promotional strategies
- Maintaining customer trust through ratings and reviews
- Ensuring product availability to avoid lost sales

c. Why This Problem Was Chosen

This project was selected to understand the key factors that influence product demand and customer behavior in an online marketplace. The analysis helps simulate real-world business decision-making for pricing, promotion, and inventory planning.

---

4. Problem Statement & Objectives

a. Problem Statement: To identify the key factors that influence product demand and performance in the Amazon electronics category.

b. Objectives
   - Evaluate the impact of sponsored listings and Best Seller status
   - Study the effect of Buy Box availability on sales.
   - Develop a dashboard to support data-driven business decisions

c. Project Scope
   - Product-level analysis using Amazon electronics dataset
   - Data cleaning, transformation, and analysis performed in Google Sheets
   - Demand measured using *bought_in_last_month* as a proxy

d. Success Criteria
   - Clean and structured dataset
   - Interactive dashboard with key KPIs
   - Actionable insights for business decisions

---

5. Data Description

a. Dataset Source

- Platform: Kaggle
- Domain: Amazon Electronics
- Approximate size: **9K products**

b. Data Structure

Each row represents one product listing.

c. Key Columns (Cleaned Dataset)
- title – Product name
- rating – Average customer rating
- number_of_reviews – Total reviews
- price-Estimated price of products
- bought_in_last_month – Estimated demand
- is_best_seller – Best Seller indicator
- is_sponsored – Sponsored listing indicator
- is_couponed – Coupon availability
- buy_box_availability – Add to Cart / Unavailable
- badges / sustainability_badges – Product badges

d. Data Limitations
- Snapshot data (no historical trends)
- Demand is estimated, not actual sales
- No customer demographic information

6. Data Cleaning & Preparation

a. Missing Values Handling
- Blank review counts and purchase values were converted to numeric where possible
- Missing badges were replaced with **"No Badge"**

- Inconsistent availability values were standardized

b. Data Type Standardization
- Ratings and review counts converted to numeric

c. Transformations
- Extracted numeric values from text fields (e.g., bought_in_last_month)
- Standardized category names for availability and badges

- Trimmed extra spaces and corrected text inconsistencies

d. Feature Engineering
- Refined Best Seller and Sponsored column values for consistent labeling and clearer grouping in pivot tables
- Reordered columns for easier analysis and pivot creation
- Dropped columns delivery_details and collected_at

e. Outlier Treatment
- Reviewed price, rating, and number_of_reviews columns for extreme or unrealistic values.
- Verified unusually high prices to ensure they were valid premium electronic products and not data errors.
- Ensured ratings remained within the valid range (1–5) and removed any invalid entries if found.

d. Assumptions

- Missing purchase values indicate low or no recent demand
- Buy Box availability directly impacts purchasing ability
- All prices are in the same currency

7. KPI & Metric Framework

KPI Definition: Key Performance Indicators(KPIs) are quantifiable, outcome-based measurements used to evaluate how effectively an organization is achieving its most critical strategic objectives over a specific period

1. **KPI 1**:
   a. Formula: ***AVERAGE(PRICE)***
   b. Why it matters:Helps understand the overall pricing level of products in the electronics category.Supports analysis of how price levels influence demand and customer behavior
   c. Objective Mapping: To analyze the impact of pricing on product demand

2. **KPI 2**:
   a. Formula: ***AVERAGE(rating)***
   b. Why it matters: Indicates overall customer satisfaction and product quality perception.
   c. Objective Mapping: Helps evaluate the relationship between product quality and demand.

3. **KPI 3**:
   a. Formula: ***AVERAGE(number_of_reviews)***
   b. Why it matters: Review count reflects product trust and market maturity
   c. Objective Mapping: Supports analysis of customer trust impact on sales

4. **KPI 4**:
   a. Formula: ***MEDIAN(number_of_reviews)***
   b. Why it matters: Represents typical engagement and highlights review concentration
   c. Objective Mapping:To identify engagement inequality and detect products that dominate customer attention.

5. **KPI 5**:

    a. Formula: ***AVERAGE(bought_in_last_month)***

    b. Why it matters: Acts as a proxy for product demand and sales performance.

    c. Objective Mapping:Directly measures product demand.

---

8. Exploratory Data Analysis (EDA)

a. Trend Analysis

- The large difference between average reviews (5,769) and median reviews (860) indicates a highly skewed distribution, where a small number of products receive very high customer attention. Similarly, the average price ($163) is much higher than the median price ($69.95), confirming the presence of high-priced premium products.

b. Observations

- The average product rating across the dataset is **4.45**, indicating generally high customer satisfaction.
- The average number of reviews is **5,769**, while the median is only **860**, indicating that customer engagement is highly concentrated among a small number of products.
- The average monthly demand is approximately **1,652 units**, showing active purchasing behavior in the electronics category.
- The rating distribution chart shows a clear concentration between **4.0 and 4.8**, with very few low-rated products.

c. Comparison Analysis

Comparison analysis was conducted across key business indicators.

a. Sponsored vs Organic
- Sponsored products show higher average monthly purchases, while organic products have higher average review counts (6,410 vs 3,396), indicating stronger long-term customer trust.

b. Best Seller vs Non–Best Seller
- Products with the **Best Seller badge** have the highest average purchases.
- Products without badges or promotional recognition show significantly lower demand.

c. Buy Box Availability
- Around 68% of products are available for Add to Cart, while 31.7% are unavailable.

d. Distribution Analysis

Distribution analysis helps understand the spread of key variables.

a. Rating Distribution
- Most products fall within the **4.0–4.8 range**

b. Price Distribution
- Price distribution is highly skewed, with an average price of $163 and a median price of $69.95, indicating the presence of premium-priced products in the category.

e. Correlation Analysis

Correlation relationships were analyzed between key variables.

a. **Rating vs Bought in Last Month**

Higher-rated products generally show higher purchase volumes.

c. **Number of Reviews vs Demand**

Products with more reviews tend to perform better, indicating trust-driven purchasing.

f. Written Insights with Charts

Based on dashboard visualizations:

- Sponsored products sell nearly 2× more than organic listings.
- Best Seller products significantly outperform all other categories.
- Approximately **26% of products are unavailable**, indicating lost revenue potential.representing lost revenue opportunities.
- Products with sustainability or feature badges show higher average review counts, indicating improved customer engagement.
- High ratings (>4.0) are strongly associated with higher demand.

9. **Advanced Analysis**

a. Forecasting (Applicability)

Since the dataset is a **snapshot** and not time-series, long-term forecasting was not performed.
However, **bought_in_last_month** was used as a proxy indicator to estimate short-term demand trends.

b. Segmentation

Products were segmented based on performance drivers:

Segment 1: High Rating + High Demand

- Core revenue drivers
- Ideal candidates for sponsored promotion

Segment 2: High Rating + Low Demand

- Quality products lacking visibility
- Opportunity for advertising or Buy Box optimization

Segment 3: Sponsored + Low Reviews

- Newly promoted products

Segment 4: No Buy Box / Unavailable

- Immediate revenue loss risk

c. Root Cause

Low-performing products were analyzed for common issues:

Primary causes of low demand:

- No Buy Box availability
- Low rating (<4.0)
- Low number of reviews
- No visibility (not sponsored, no Best Seller badge)

d. Risk / Anomaly Analysis

Identified Risks

- **31.7% product unavailability** → direct revenue risk
- Products with very low ratings → brand reputation risk
- High-priced products with low reviews → low conversion probability

Anomalies

- Some sponsored products with low reviews but high demand, indicating promotion-driven sales.

e. Scenario Analysis

Scenario 1: Improve Buy Box Availability

If unavailable products become available:
→ Potential increase in total demand by up to **30%**

Scenario 2: Sponsored Promotion for High-Rating Products

High-rating but low-visibility products:

→ Expected demand increase of **20–40%**

Scenario 3: Review Improvement Strategy

Increasing review count improves customer trust and conversion.

10. Dashboard Design

The dashboard is designed to help sellers and marketplace managers analyze key factors affecting product demand, pricing strategy, customer engagement, and product visibility in the Amazon electronics category.

a. Filters

Interactive filters are provided for:

- Sponsored Status (is_sponsored)
- Buy Box Availability
- Badges
- Tags

b. KPI Summary

- Average Price: $163.07
- Average Rating: 4.45
- Average Number of Reviews: 5,769
- Median Number of Reviews: 860
- Average Sales (Bought in Last Month): 1,652

- Availability: **74.1% available / 25.9% unavailable**
- Organic reviews: **6,410**
- Sponsored reviews: **3,396**

The difference between average and median values indicates skewed distribution in both pricing and customer engagement.

Visual Components

1. Bar Chart: Past Month Organic vs Sponsored Sales
 Shows the impact of paid visibility on product demand.

2. Area Chart: Composition by Badge Type
 Compares sales performance across categories such as Amazon's Choice, Best Seller, Limited-time Deal, and No Badge.

3. Bubble Chart: Marketplace Positioning by Badge
 Displays the relationship between average rating, average price, and customer engagement across badge types.

4. Donut Chart: Product Availability
 Shows Add-to-Cart vs Unavailable products (74.1% available, 25.9% unavailable).

5. Line Chart: Rating Distribution
 Illustrates the concentration of products in the high-rating range.

6. Bar Chart: Price Distribution by Tags
 Compares average and median prices across product tags to identify premium segments.

11. Insights Summary

1. **Sponsored listings drive higher demand**
   Sponsored products show significantly higher average purchases compared to organic listings.

2. **Organic products have stronger customer trust**
   Organic listings have higher average review counts, indicating long-term credibility.

3. **Best Seller products dominate sales**
   Products with Best Seller status generate the highest monthly demand.

4. **Buy Box availability is critical for sales**
   Approximately **26% of products are unavailable**, leading to direct revenue loss.

5. The large gap between **average and median reviews** indicates that product engagement is highly concentrated among a small number of popular products.

6. **Price distribution is highly skewed**, with a few high-priced premium products increasing the average price.

7. **High ratings strongly influence demand**
   Most high-performing products have ratings above **4.0**.

8. **Review count impacts purchase decisions**
   Products with more reviews show higher purchase volumes due to increased trust.

9. **Demand is concentrated among a few products**
   A small percentage of listings contribute to a large share of purchases (long-tail pattern)

10. **Product badges improve engagement**
    Products with sustainability or feature badges have higher average reviews.

12. Recommendations

1. Increase Sponsored Promotion for High-Rating Products

- **Insight:** Sponsored products have higher demand
- **Action:** Promote products with rating >4 and good reviews.
- **Impact:** Higher visibility and faster sales growth.
- **Feasibility:** High (platform feature available).

2. Improve Buy Box Eligibility

- **Insight:** 30%+ products unavailable.
- **Action:** Optimize pricing, seller performance, and inventory.
- **Impact:** Immediate increase in conversion.
- **Feasibility:** Medium–High.

3. Focus on Review Generation Strategy

- **Insight:** Higher reviews = higher demand.
- **Action:** Encourage customer feedback and post-purchase engagement.
- **Impact:** Increased customer trust and conversion.
- **Feasibility:** High.

4. Identify and Promote High-Quality Low-Visibility Products

- **Insight:** Some high-rating products underperform.
- **Action:** Provide sponsored ads or coupons for these products.
- **Impact:** Unlock hidden revenue.
- **Feasibility:** High.

5. Optimize Inventory for Best Seller Products

- **Insight:** Best Sellers drive maximum demand.
- **Action:** Maintain higher stock levels for top-performing items.
- **Impact:** Prevent lost sales.
- **Feasibility:** Medium.

6. Highlight Product Badges and Certifications

- **Insight:** Badge products receive higher engagement.
- **Action:** Display sustainability/features clearly in listings.
- **Impact:** Improved customer perception.
- **Feasibility:** High.

7.Price Segmentation Strategy

- **Insight:** Price distribution is skewed with premium outliers.
- **Recommendation:** Segment products into budget, mid-range, and premium categories and apply category-specific pricing and promotion strategies.
- **Impact:** Better targeting and improved conversion.

13. Impact Estimation

Revenue Impact

- Improving Buy Box availability for unavailable products could increase overall demand by **20–30%**.

- Sponsoring high-potential products may increase sales by **15–40%**.

Cost Saving

- Targeted promotion reduces unnecessary advertising on low-potential products.
- Better inventory planning reduces overstock and stock-out costs.

Efficiency Improvement

- Data-driven product prioritization improves marketing ROI.
- Helps focus operational efforts on high-performing products.

Service Improvement

- Better availability improves customer experience.
- Higher ratings and reviews increase customer confidence.

Risk Reduction

- Monitoring low-rated products prevents brand reputation issues.
- Reduces revenue loss due to product unavailability.

14. Limitations

Data Issues

- Dataset is a snapshot.
- Some fields required text-to-numeric conversion.
- Demand is estimated using *bought_in_last_month*.

Assumption Risks

- Assumed missing purchases as low demand.
- Assumed Buy Box availability directly reflects sales ability.

What Cannot Be Concluded

- Exact revenue or profit.
- Customer demographics or behavior trends.
- Seasonal demand patterns.

---

15. Future Scope

- Time-series analysis to study demand trends.
- Category-level performance comparison.
- Price optimization and discount analysis.
- Machine learning model for demand prediction.
- Integration with inventory and sales data for accurate revenue analysis.

---

16. Conclusion

This project analyzed Amazon electronics product data using Google Sheets to identify key drivers of product performance. The analysis shows that customer trust (ratings and reviews), platform visibility (Sponsored and Best Seller), and product availability are the most critical factors influencing demand. The developed dashboard enables data-driven decision-making for pricing, promotion, and inventory management. Implementing the

recommended strategies can improve revenue, operational efficiency, and customer experience.

---

17. Appendix

Data Dictionary (Summary)

Column	Description
title	Product name
rating	Average rating (1–5)
number_of_reviews	Total reviews
Price	Estimated price of the products
bought_in_last_month	Estimated monthly demand
is_sponsored	Sponsored listing flag
is_best_seller	Best Seller indicator
buy_box_availability	Add to Cart / Unavailable

badges   Product

certifications/features

18. Contribution Matrix

Team Member	Dataset & Sourcing	Cleaning	KPI & Analysis	Dashboard	Report	PPT	Overall Role
Khyati Batra	✔	✔			✔		Strategy Lead
Agrima Ojha	✔	✔	✔	✔			Project Lead
Suryansh Chattree	✔	✔	✔	✔			Dashboard Lead
Meghna Nair	✔		✔			✔	Analysis Lead
Deepanshu Chaudhary			✔				Data Lead
Raghvendra Singh			✔				PPT & Quality Lead

Declaration: We confirm that the above contribution details are accurate and verifiable through version history and submitted artifacts.

Team Signature Block:

Meghna

Aquima Ojha

Suryansh

Deepanshu

Raghavendra