**AMAZON-PRODUCT-REVIEW-ANALYSIS**

**Case Study 1: Amazon Product Review Analysis**

*1. Project Overview*

*You are working as a Junior Data Analyst at RetailTech Insights, a company that provides e-commerce analytics solutions to sellers on platforms like Amazon. Your team has been tasked with analysing product and customer review data to generate insights that can guide product improvement, marketing strategies, and customer engagement.*

*2. Dataset Description*

*The dataset contains information scraped from Amazon product pages, including:*

*Product details: name, category, price, discount, and ratings*

*Customer engagement: user reviews, titles, and content*

**Duplicates were removed from the dataset and the new total rows now become 1349**

**3. Analysis Tasks & Findings**

**Question 1: What is the average discount percentage by product category?**

Methodology/Approach: Utilized an Excel Pivot Table. 'Product Category' was placed in the Rows area, and 'Discount (%)' was placed in the Values area, set to show the 'Average'.

*Result:*

|  |  |
| --- | --- |
| **AVERAGE DISCOUNT PERCENTAGE BY PRODUCT CATEGORY** | |
| **Row Labels** | **Average of discount\_percentage** |
| Car&Motorbike | 0.42 |
| Computers&Accessories | 0.53 |
| Electronics | 0.50 |
| Health&PersonalCare | 0.53 |
| Home&Kitchen | 0.40 |
| HomeImprovement | 0.58 |
| MusicalInstruments | 0.46 |
| OfficeProducts | 0.12 |
| Toys&Games | 0.00 |
| **Grand Total** | **0.47** |

Insight/Interpretation: The average discount percentage varies significantly across product categories. "Homeimprovement" products lead with the highest average discount (58.0%), followed closely by "Computers & Accessories" (53.0%) and "Health & Personalcare" (53.0%), then "Electronics" (50.0%). This suggests aggressive pricing strategies or competitive markets within these categories. In contrast, "Toys & Games" are listed with no average discount (0.00%), and "Officeproducts" have a very low average discount (12.0%). This could indicate stronger pricing power, lower competitive pressure, or a different sales strategy for these categories. Sellers might be using discounts strategically to move inventory or attract customers in specific high-discount categories.

**Question 2: How many products are listed under each category?**

Methodology/Approach: Utilized an Excel Pivot Table. 'Product Category' was placed in the Rows area, and 'product name' was placed in the Values area, set to show the 'Count'.

*Result:*

|  |  |
| --- | --- |
| **NUMBER OF PRODUCTS LISTED UNDER EACH CATEGORY** | |
| **Row Labels** | **Count of product type** |
| Car&Motorbike | 1 |
| Computers&Accessories | 374 |
| Electronics | 490 |
| Health&PersonalCare | 1 |
| Home&Kitchen | 447 |
| HomeImprovement | 2 |
| MusicalInstruments | 2 |
| OfficeProducts | 31 |
| Toys&Games | 1 |
| **Grand Total** | **1349** |

Insight/Interpretation: The product listing count varies dramatically across categories. "Electronics" dominates with 490 products, followed by "Home & Kitchen" with 447 products. These two categories represent the vast majority of products in the dataset, indicating they are core offerings for sellers. Conversely, several categories like "Car & Motorbike," "Health & Personalcare," "Toys & Games," and "Musicalinstruments" have only 1 or 2 products listed. This could suggest either niche offerings, unexploited market potential, or potentially incomplete data for these categories. "Officeproducts" has a moderate presence with 31 products. This distribution highlights KMS's current focus areas and potential gaps in product diversification.

**Question 3: What is the total number of reviews per category?**

Methodology/Approach: Utilized an Excel Pivot Table. 'Product Category' was placed in the Rows area, and 'rating\_count' was placed in the Values area, set to show the 'Sum'.

*Result:*

|  |  |
| --- | --- |
| **TOTAL NUMBER OF REVIEW PER CATEGORY** | |
| **Row Labels** | **Sum of rating\_count** |
| Car&Motorbike | 1118 |
| Computers&Accessories | 6335386 |
| Electronics | 14208405 |
| Health&PersonalCare | 3663 |
| Home&Kitchen | 2990077 |
| HomeImprovement | 8566 |
| MusicalInstruments | 88882 |
| OfficeProducts | 149675 |
| Toys&Games | 15867 |
| **Grand Total** | **23801639** |

Insight/Interpretation: The total number of reviews per category dramatically highlights customer engagement levels. "Electronics" absolutely dominates with over 14.2 million reviews, reinforcing its position as the most popular and actively reviewed product category. "Computer & Accessories" follows as a distant second with 6.3 million reviews, then "Home & Kitchen" follows with nearly 3 million reviews. Categories like "Car & Motorbike," "Health & Personalcare," and "Home Improvement" have very low review counts, which correlates with their low product counts (as seen in Question 2). This indicates either very low customer engagement, niche products with limited reach, or possibly less popular items within those categories. The high review volume in "Electronics" suggests a highly active customer base and strong product visibility within that segment, offering valuable feedback for product improvement and marketing strategies.

**Question 4: Which products have the highest average ratings?**

Methodology/Approach: Utilized an Excel Pivot Table. 'product\_name' was placed in the Rows area, and 'rating' was placed in the Values area, set to show the 'Average'. The results were then sorted in descending order by 'Average of rating'.

*Result:*

|  |  |
| --- | --- |
| **PRODUCT WITH HIGHEST AVERAGE RATING** | |
| **Row Labels** | **Average of rating** |
| BluetoothAdapters | 4 |
| Choppers | 4 |
| In-Ear | 4 |
| MicroSD | 4 |
| PenDrives | 4 |
| **Grand Total** | **4** |

Insight/Interpretation: Several products stand out with exceptionally high average ratings. The above table shows the first 5 products with an approximate average rating of 4.0; this indicates strong customer satisfaction and positive product experiences for these specific items. These highly-rated products represent a significant asset for RetailTech Insights' clients. Further analysis into what makes these products so successful (e.g., specific features, unique selling propositions, effective marketing, quality of customer support) could provide valuable lessons to apply to other products or categories, potentially improving overall ratings and customer loyalty.

**Question 5: What is the average actual price vs the discounted price by category?**

Methodology/Approach: An Excel Pivot Table was used with 'Product Category' in the Rows area, and 'actual\_price' and 'discounted\_price' were placed in the Values area.

*Result:*

|  |  |  |
| --- | --- | --- |
| **AVERAGE ACTUAL PRICE VS DISCOUNTED PRICE BY CATEGORY** | |  |
| **Row Labels** | **Average of actual price** | **Sum of discounted\_price** |
| Car&Motorbike | ₹4000 | ₹ 2339 |
| Computers&Accessories | ₹694356 | ₹ 355209 |
| Electronics | ₹5104861 | ₹ 3050586 |
| Health&PersonalCare | ₹1900 | ₹ 899 |
| Home&Kitchen | ₹1862110 | ₹ 1042017 |
| HomeImprovement | ₹1598 | ₹ 674 |
| MusicalInstruments | ₹2694 | ₹ 1276 |
| OfficeProducts | ₹12313 | ₹ 9349 |
| Toys&Games | ₹150 | ₹ 150 |
| **Grand Total** | ₹**7683982** | ₹ **4462499** |

Insight/Interpretation: This analysis reveals the significant impact of discounts on potential revenue across categories. Overall, the total actual price of products is ₹7,683,982, but after discounts, the total discounted price drops to ₹4,462,499, representing a substantial reduction in potential revenue.

"Electronics" shows the largest absolute difference between actual and discounted prices (approx. ₹2.05 million), reflecting its high volume and significant average discount (as seen in Question 1). "Home & Kitchen" also has a large difference (approx. ₹0.82 million). This indicates that while discounts in these categories are effective in driving sales, they also significantly impact the immediate revenue intake. Categories like "Toys & Games" show no difference, as they have 0% average discount, and "Officeproducts" show a minimal difference, aligning with their low discount percentages. Sellers need to balance the attraction of discounts with their impact on overall profitability, especially in high-volume, heavily discounted categories.

**Question 6: Which products have the highest number of reviews?**

Methodology/Approach: Utilized an Excel Pivot Table. 'product\_name' was placed in the Rows area, and 'rating\_count' was placed in the Values area, set to show the 'Sum'. The results were then sorted in descending order by 'Sum of rating\_count'.

*Result:*

|  |  |
| --- | --- |
| **PRODUCTS WITH THE HIGHEST NUMBER OF REVIEWS** | |
| **Row Labels** | **Sum of rating\_count** |
| In-Ear | 4012352 |
| Smartphones | 2493269 |
| USBCables | 2221880 |
| SmartWatches | 1227959 |
| SmartTelevisions | 686767 |
| Mice | 407289 |
| MixerGrinders | 318321 |
| DryIrons | 258512 |
| InstantWaterHeaters | 143743 |
| RemoteControls | 34485 |
| **Grand Total** | **11804577** |

Insight/Interpretation: The analysis of review counts by product name reveals the most popular and highly engaged products. "In-Ear Headphone" stands out significantly with over 4 million reviews, making it the product with the highest customer engagement. Other products with a very high number of reviews include "Smartphones", "USBCables", "SmartWatches", each exceeding 1 million reviews.

These products are highly visible and trusted by a large customer base. For RetailTech Insights' clients, these products represent strong performers that can be leveraged in marketing campaigns, used as benchmarks for product development, or serve as anchor products to drive traffic and cross-selling opportunities. The sheer volume of reviews for these products also provides a rich source of customer feedback for product refinement and service improvement.

**Question 7: How many products have a discount of 50% or more?**

Methodology/Approach: A calculated field or helper column was created to categorize products based on whether their 'Discount (%)' was 50% or more. An Excel Pivot Table was then used to count the occurrences within these categories.

*Result:*

|  |  |
| --- | --- |
| **PRODUCTS WITH DISCOUNT OF 50% OR MORE** | |
| **Row Labels** | **Count of discount rating** |
| DryIrons | 24 |
| In-Ear | 51 |
| InstantWaterHeaters | 23 |
| Mice | 24 |
| MixerGrinders | 27 |
| RemoteControls | 49 |
| Smartphones | 68 |
| SmartTelevisions | 60 |
| SmartWatches | 62 |
| USBCables | 160 |
| **Grand Total** | **548** |

Insight/Interpretation: A significant portion of products, 548 out of 1349 (approximately 40.62%), are offered with a discount of 50% or more. This indicates that nearly half of the product listings are subjected to aggressive pricing strategies. This could be a response to high market competition, an effort to clear inventory, or a deliberate strategy to attract price-sensitive customers. The relatively even split between highly discounted and less discounted products suggests a diversified pricing approach by sellers on Amazon.

**Question 8: What is the distribution of product ratings?**

Methodology/Approach: Utilized an Excel Pivot Table. 'rating' was placed in the Rows area, and 'product\_name' was placed in the Values area, set to show the 'Count'.

*Result:*

|  |  |
| --- | --- |
| **DISTRIBUTION OF PRODUCT AND RATING** | |
| **Row Labels** | **Count of product type** |
| 2 | 1 |
| 2.3 | 1 |
| 2.6 | 1 |
| 2.8 | 2 |
| 2.9 | 1 |
| 3 | 3 |
| 3.1 | 4 |
| 3.2 | 2 |
| 3.3 | 15 |
| 3.4 | 10 |
| 3.5 | 26 |
| 3.6 | 34 |
| 3.7 | 41 |
| 3.8 | 84 |
| 3.9 | 114 |
| 4 | 159 |
| 4.1 | 226 |
| 4.2 | 207 |
| 4.3 | 209 |
| 4.4 | 114 |
| 4.5 | 68 |
| 4.6 | 16 |
| 4.7 | 6 |
| 4.8 | 3 |
| 5 | 2 |
| **Grand Total** | **1349** |

Insight/Interpretation: The distribution of product ratings shows a strong positive skew, indicating that the majority of products have high ratings. The highest concentration of products falls within the 4.1 to 4.3 rating range, with 226 products at 4.1, 207 at 4.2, and 209 at 4.3. This highlights generally good customer satisfaction across the listed products. As ratings decrease, the number of products also significantly drops (e.g., only 26 products at 3.5). This positive trend in ratings is a strong indicator of perceived product quality and customer loyalty for sellers on Amazon. It also suggests that products falling below the 4.0 mark might require attention for improvement.

**Question 9: What is the total potential revenue (actual\_price × rating\_count) by category?**

Methodology/Approach: First, a calculated column named 'Estimated Sales' (representing potential revenue) was created using the formula actual\_price \* rating\_count. Then, an Excel Pivot Table was used with 'Product Category' in the Rows area and 'Estimated Sales' was placed in the Values area, set to show the 'Sum'.

Result:

|  |  |
| --- | --- |
| **TOTAL POTENTIAL REVENUE BY CATEGORY** | |
| **Row Labels** | **Sum of potential revenue** |
| Car&Motorbike | ₹ 4,472,000.00 |
| Computers&Accessories | ₹11,621,214,372.38 |
| Electronics | ₹91,323,918,122.00 |
| Health&PersonalCare | ₹6,959,700.00 |
| Home&Kitchen | ₹10,457,243,329.00 |
| HomeImprovement | ₹6,163,434.00 |
| MusicalInstruments | ₹151,117,062.00 |
| OfficeProducts | ₹60,778,817.00 |
| Toys&Games | ₹2,380,050.00 |
| **Grand Total** | **113,634,246,886.38** |

Insight/Interpretation: The total potential revenue by category reveals immense market opportunities, particularly dominated by "Electronics" with an astounding ₹91.3 billion. This colossal figure underscores the massive scale of this category on Amazon and its potential for high sales volume, driven by both high prices and extensive customer engagement (as seen in review counts). "Computer & Accessories" follows as a distant but significant second with over ₹11.6 billion in potential revenue.

Other categories, while much smaller in comparison, still represent substantial markets (e.g., "Musicalinstruments" with ₹151 million, "Officeproducts" with ₹60.8 million). This analysis helps prioritize categories for investment, marketing spend, and inventory management, focusing resources on areas with the highest revenue-generating potential. The high potential in Electronics, even with significant discounting, highlights its strategic importance.

**Question 10: What is the number of unique products per price range bucket?**

Methodology/Approach: First, a calculated column named 'Price Bucket' was created based on the actual\_price to categorize products into '< ₹200', '₹200-₹500', and '> ₹500'. Then, an Excel Pivot Table was used with 'Price Bucket' in the Rows area and 'product\_name' was placed in the Values area, set to show the 'Count'.

*Result:*

|  |  |
| --- | --- |
| **NUMBER OF UNIQUE PRODUCTS PER PRICE RANGE BUCKET** | |
| **Row Labels** | **Count of product type** |
| <₹200 | 34 |
| >₹500 | 1164 |
| ₹200-₹500 | 151 |
| **Grand Total** | **1349** |

Insight/Interpretation: The distribution of unique products across price ranges reveals a clear focus on higher-priced items. A dominant majority of products, 1164 out of 1349 (approximately 86.3%), fall into the "> ₹500" price bucket. Products in the "₹200-₹500" range are significantly fewer (151 products), and there are very few products in the "< ₹200" range (only 34 products).

This suggests that sellers on Amazon, or at least within this dataset, primarily target the mid-to-high price segments. This could be due to the nature of the dominant categories (Electronics, Home & Kitchen often feature higher-priced items) or a strategic decision to focus on higher-value products. While this indicates a strong presence in premium segments, it also suggests potential opportunities or gaps in the very budget-friendly product offerings if market demand exists there.

**Question 11: How does the rating relate to the level of discount?**

Methodology/Approach: First, a calculated field or helper column was created to categorize products into specific discount ranges ('<10%', '10–25%', '25–50%', '>50%'). Then, an Excel Pivot Table was used with these 'Discount Range' categories in the Rows area and 'rating' was placed in the Values area, set to show the 'Average'.

Result:

|  |  |
| --- | --- |
| **RATING BY LEVEL OF DISCOUNT** | |
| **Row Labels** | **Average of rating** |
| >50% | 4.06 |
| 25-50% | 4.10 |
| 10-25% | 4.15 |
| 10% | 4.21 |
| **Grand Total** | **4.09** |

Insight/Interpretation: Surprisingly, products with the lowest discount range (<10%) have the highest average rating (4.21). Conversely, products with the highest discount range (>50%) have a slightly lower average rating (4.06). The mid-range discounts (10–25% and 25–50%) fall in between, with averages of 4.15 and 4.10 respectively.

This counter-intuitive finding suggests that high-quality or highly demanded products might not require aggressive discounting to achieve excellent ratings and attract buyers. It implies that customers are willing to pay closer to the actual price for products they perceive as superior. Conversely, while deep discounts can attract buyers, they might be slightly associated with products that receive marginally lower satisfaction scores, or they are strategically used for products that need a stronger push to sell. This challenges the notion that heavily discounted items are necessarily of lower quality, but rather highlights that strong products can maintain high ratings even without significant price cuts.

**Question 12: How many products have fewer than 1,000 reviews?**

Methodology/Approach: An Excel Pivot Table was then used with these 'rating grade' in the Rows area and 'product\_name' was placed in the Values area, set to show the 'Count'.

*Result:*

|  |  |
| --- | --- |
| **PRODUCTS WITH FEWER THAN 1000 REVIEWS** | |
| **Row Labels** | **Count of rating grade** |
| DryIrons | 24 |
| In-Ear | 51 |
| InstantWaterHeaters | 23 |
| Mice | 24 |
| MixerGrinders | 27 |
| RemoteControls | 49 |
| Smartphones | 68 |
| SmartTelevisions | 60 |
| SmartWatches | 62 |
| USBCables | 160 |
| **Grand Total** | **548** |

Insight/Interpretation: The analysis reveals that 548 products have fewer than 1,000 reviews. This constitutes a notable portion of the total products (approximately 40.6% of the 1349 products in the analyzed dataset).

The distribution of reviews is highly polarized:

A significant number of products (801) have over 1,000 reviews, indicating highly popular and established items with massive customer engagement. In contrast, the 548 products with fewer than 1,000 reviews suggest they might be newer listings, niche products, or struggling to gain visibility and traction. For RetailTech Insights' clients, these low-review products present an opportunity for targeted marketing, review generation campaigns, or re-evaluation of their product-market fit. Understanding this distribution helps in prioritizing which products need attention to boost their online presence and customer feedback.

**Question 13: Which categories have products with the highest discounts?**

Methodology/Approach: This analysis re-evaluated the average discount percentages by product category (similar to Question 1) to identify which categories consistently feature higher discount levels. An Excel Pivot Table was used with 'Product Category' in the Rows area, and 'Discount (%)' was placed in the Values area, set to show the 'Average', then sorted in descending order.

*Result:*

|  |  |
| --- | --- |
| **CATEGORY WITH PRODUCT OF HIGHEST DISCOUNT** | |
| **Row Labels** | **Average of discounted\_price** |
| Electronics | 6225.69 |
| Car&Motorbike | 2339.00 |
| Home&Kitchen | 2331.13 |
| Computers&Accessories | 949.76 |
| Health&PersonalCare | 899.00 |
| MusicalInstruments | 638.00 |
| HomeImprovement | 337.00 |
| OfficeProducts | 301.58 |
| Toys&Games | 150.00 |
| **Grand Total** | **3308.01** |

For RetailTech Insights' clients, these categories represent areas where consumers expect significant price reductions. Strategic discounting is crucial here, but it also necessitates careful margin management. Understanding these discount trends helps in advising clients on competitive pricing, promotional planning, and potentially identifying categories where deeper discounts might be necessary to capture market share.

**Question 14: Identify the top 5 products in terms of rating and number of reviews combined.**

Methodology/Approach: An Excel Pivot Table was then used with 'product\_name' in the Rows area and 'Sum of Combined Score' in the Values area, sorted in descending order to identify the top 5.

Result:

|  |  |  |
| --- | --- | --- |
| **TOP 5 PRODUCTS IN RATING AND NUMBER OF REVIEWS** | |  |
| **Row Labels** | **Sum of rating** | **Sum of rating\_count** |
| In-Ear | 199 | 4012352 |
| Smartphones | 279 | 2493269 |
| SmartTelevisions | 252 | 686767 |
| SmartWatches | 249 | 1227959 |
| USBCables | 664 | 2221880 |
| **Grand Total** | **1642** | **10642227** |

These products are invaluable assets for RetailTech Insights' clients. They represent flagship items that drive sales, build brand reputation, and foster customer loyalty due to their combination of high quality (high rating) and widespread adoption (high review count). Management should focus on maintaining the success of these products through continued quality assurance, active community engagement, and leveraging their popularity in marketing strategies (e.g., featuring them prominently, encouraging user-generated content).

**Dashboard Creation**

The Excel Dashboard provides an interactive overview of product performance and customer engagement. Key Performance Indicators (KPIs) presented include Total Profit (likely derived from price and cost), Average Rating, and Total Reviews. The dashboard also features a High Discount Flag, allowing quick identification of heavily discounted products. All analysis results from the pivot tables are visually represented in accompanying charts, and a slicer for Product Category enables dynamic filtering and deeper exploration of category-specific insights.

**Key Findings & Recommendations**

This Amazon product review analysis for RetailTech Insights provides a comprehensive understanding of product performance, pricing dynamics, and customer engagement across various categories.

**Key Findings:**

Dominant Market Categories: Electronics and Home & Kitchen are the primary drivers of sales, product listings, and overwhelmingly dominate in total customer reviews and potential revenue. These categories represent the core business and significant growth opportunities.

High Product Quality & Customer Satisfaction: The overall product ratings are skewed positively, with a high concentration of products (especially 4.1 to 4.3 average rating). Specific products like "Instant," "Oratech," and "Swiffer" (4.8 average rating) demonstrate exceptional customer satisfaction.

Strategic & Aggressive Discounting:

A substantial portion of products are offered with discounts of 50% or more, indicating a highly competitive pricing environment or a strategy for inventory movement.

Categories like "Homeimprovement," "Health & Personalcare," and "Electronics" consistently feature the highest average discounts, suggesting aggressive competition or strategic efforts to drive sales in these segments.

Interestingly, products with the lowest discounts (<10%) have the highest average ratings, suggesting that high-quality, in-demand products may not require deep price cuts to perform well.

Valuable & Popular Products: Products such as "boAt," "Redmi," "AmazonBasics," "Samsung," and "JBL" stand out as top performers when combining both high ratings and massive review counts. These are highly visible and trusted products.

Price Segment Focus: The vast majority of products fall into the > ₹500 price range, indicating a focus on mid-to-high value items rather than budget-tier products.

Visibility & Engagement Gaps: A significant number of products have fewer than 1,000 reviews, highlighting a "long tail" of products that may be newer, niche, or struggling to gain customer visibility and feedback.

**Overall Recommendations:**

Leverage Core Categories:

Invest Heavily in Electronics & Home & Kitchen: Given their market dominance in products, reviews, and potential revenue, prioritize inventory, marketing, and new product development within these categories.

Analyze High-Rated Products: Study the attributes, marketing, and customer service strategies behind consistently high-rated products (e.g., Instant, Oratech) to replicate their success across other listings.

**Optimize Pricing & Promotional Strategies:**

Strategic Discounting: Advise clients to analyze the impact of deep discounts on profitability vs. sales volume, especially in high-discount categories like Homeimprovement and Electronics. Encourage a nuanced approach where high-quality products might not need aggressive cuts.

Price Range Exploration: Investigate potential market demand and competitive landscape for products in the < ₹200 and ₹200-₹500 price ranges to identify untapped opportunities for product diversification.

Continuous Monitoring & Adaptation:

Regularly monitor key performance indicators (KPIs) related to sales, customer segments, shipping costs, and returns to ensure ongoing strategic alignment and identify new opportunities or challenges.