DSA DATA ANALYSIS CAPSTONE PROJECT

CASE STUDY 1: AMAZON PRODUCT REVIEW ANALYSIS

JULY 4TH 2025

JONATHAN OMHANS OKHUNKHUN

LAGOS, NIGERIA

**DOCUMENTATIONS ON AMAZON PRODUCT REVIEW ANALYSIS.**

**OBJECTIVE:**

The objective of the analysis is to generate insights that can guide product improvement, market strategies, and customer engagement.

**UNDERLYING ASSUMPTIONS OF THE ANALYSIS**

**a.** The spotlight is on pricing, ratings, discount pattern, and category insights. Therefore, non-relevant columns on the original data like; Product Name, Product Sub Categories, Review ID, User Name, Review Title, Review Content, were removed.

**b.** Since each row contains a unique product that has a unique ID, all duplicates of product ID were removed. (This reduced the row from 1465 to 1351).

**c.** Potential Revenue is assumed to be the product of Actual Price and Rating Count [Actual Price \* Rating Count].

**d**. For analysis purposes,price ranges are as follows; **below $200, $200-$500, 501-$1000, $1001-$10,000, and above $10,000.**

**e**. Rating Count is the same as Number of Reviews

**FUNCTIONS APPLIED**

**i. Text to Columns Function** was used t*o* split the product categories into Main and subs. The only the Main category was retained while order removed

**Steps: Click Data Tab →** Text to Columns **→** (on the Text to Columns Wizard) Select Delimited and click Next **→** under Delimiter, check (√) others and Type in “|” on the box **→** select “Finish”.

**ii. Delete Function** wasusedto remove all the non-relevant Columns.

**Steps:** Select all thenon-relevant Columns and press the “Delete Key”

**iii.** Data Validation for columns (to for check errors, inconsistencies, null cell) was done using “the Filter Function”; two null rows on Column “Rating Count” were replaced with “0” (Zero), a row with 13,9,900 on Column “Actual Price” was corrected to 139,900, and a row with “|” on Column “Rating” was also marked N/A.

**iv.** Convert the relevant columns to table

**Steps:** Click Insert Tab **→** Click Table **→** check (√) the box “My table has headers” **→** Click Ok.

**v.** Create a **Pivot Table** that will be used for the analysis

**Steps:** Select allthe relevant columns **→** Click Insert Tab **→** Click Pivot Table **→** on the Wizard, Select “New worksheet” and Click Ok.

**ANALYSIS TASKS**

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

|  |  |
| --- | --- |
| **1. AVERAGE DISCOUNT PERCENTAGE BY PRODUCT CATEGORY** | |
| **Product Category** | **Average Discount %** |
| Electronics | 51% |
| Computers & Accessories | 54% |
| Musical Instruments | 46% |
| Office Products | 12% |
| Home & Kitchen | 40% |
| Home Improvement | 58% |
| Toys & Games | 0% |
| Car & Motorbike | 42% |
| Health & Personal Care | 53% |

**2. How many products are listed under each category?**

|  |  |
| --- | --- |
| **2. NUMBER OF PRODUCTS BY CATEGORY** | |
| **Product Category** | **No of Products** |
| Electronics | 526 |
| Computers & Accessories | 453 |
| Musical Instruments | 2 |
| Office Products | 31 |
| Home & Kitchen | 334 |
| Home Improvement | 2 |
| Toys & Games | 1 |
| Car & Motorbike | 1 |
| Health & Personal Care | 1 |

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

|  |  |
| --- | --- |
| **3. NUMBER OF REVIEWS PER CATEGORY** | |
| **Product Category** | **Number of Reviews** |
| Electronics | 15,778,848 |
| Computers & Accessories | 7,728,689 |
| Musical Instruments | 88,882 |
| Office Products | 149,675 |
| Home & Kitchen | 2,729,922 |
| Home Improvement | 8,566 |
| Toys & Games | 15,867 |
| Car & Motorbike | 1,118 |
| Health & Personal Care | 3,663 |

**4. Which products have the highest average ratings?**

|  |  |
| --- | --- |
| **4. TOP 10 PRODUCTS WITH THE HIGHEST AVERAGE RATINGS** | |
| **Product ID** | **Average Rating** |
| B0BP7XLX48 | 5.0 |
| B0BQRJ3C47 | 5.0 |
| B091JF2TFD | 5.0 |
| B0BCKJJN8R | 4.8 |
| B08YK7BBD2 | 4.8 |
| B09MFR93KS | 4.8 |
| B00B7GKXMG | 4.7 |
| B07Q4QV1DL | 4.7 |
| B09W9V2PXG | 4.7 |
| B09WN3SRC7 | 4.7 |

**5. What is the average actual price versus the discount price by category?**

|  |  |  |
| --- | --- | --- |
| **5. THE AVERAGE ACTUAL PRICE VERSUS THE DISCOUNTED PRICE BY CATEGORY** | | |
| **Product ID** | **Average of Actual Price** | **Average of Discounted Price** |
| Electronics | 10,127 | 5,966 |
| Computers & Accessories | 1,684 | 843 |
| Musical Instruments | 1,347 | 638 |
| Office Products | 397 | 302 |
| Home & Kitchen | 4,193 | 2,319 |
| Home Improvement | 799 | 337 |
| Toys & Games | 150 | 150 |
| Car & Motorbike | 4,000 | 2,339 |
| Health & Personal Care | 1,900 | 899 |

**6. Which products have the highest number of reviews?**

|  |  |
| --- | --- |
| **6. TOP 11 PRODUCTS WITH THE HIGHEST NUMBER OF REVIEWS** | |
| **Product ID** | **Sum of Number of Reviews** |
| B014I8SSD0 | 426,973 |
| B07KSMBL2H | 426,973 |
| B014I8SX4Y | 426,973 |
| B07YL54NVJ | 426,972 |
| B0B3D39RKV | 363,713 |
| B07GPXXNNG | 363,713 |
| B08FTFXNNB | 363,711 |
| B0BF4YBLPX | 313,836 |
| B09GFLXVH9 | 313,836 |
| B09PLFJ7ZW | 313,832 |
| B07T5DKR5D | 313,832 |

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

|  |  |
| --- | --- |
| **7. NUMBER OF PRODUCTS THAT HAVE A DISCOUNT OF 50% OR MORE** | |
| **Discounts** | **Number of Products** |
| 50% or More | 712 |
| Below 50% | 639 |

**8. What is the distribution of product ratings?**

|  |  |
| --- | --- |
| **8. THE DISTRIBUTION OF PRODUCT RATINGS** | |
| **Ratings** | **Number of Products** |
| 2 | 1 |
| 2.3 | 1 |
| 2.8 | 2 |
| 3 | 4 |
| 3.1 | 3 |
| 3.2 | 1 |
| 3.3 | 15 |
| 3.4 | 9 |
| 3.5 | 21 |
| 3.6 | 27 |
| 3.7 | 40 |
| 3.8 | 79 |
| 3.9 | 109 |
| 4 | 169 |
| 4.1 | 228 |
| 4.2 | 212 |
| 4.3 | 216 |
| 4.4 | 117 |
| 4.5 | 70 |
| 4.6 | 15 |
| 4.7 | 5 |
| 4.8 | 3 |
| 5 | 3 |
| 0 | 1 |

**9. What is the total potential revenue?**

|  |
| --- |
| **9. THE TOTAL POTENTIAL REVENUE** |
| **Sum of Potential Revenue** |
| $ 120,030,078,505 |

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

|  |  |
| --- | --- |
| **10. THE NUMBER OF UNIQUE PRODUCTS PER PRICE RANGE BUCKET** | |
| **PRICE RANGE** | **Number of Products** |
| $1001-$10,000 | 664 |
| $501-$1000 | 300 |
| Above $10,000 | 180 |
| $200-$500 | 171 |
| Below $200 | 36 |

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

**12. How many products have fewer than a thousand (1000) reviews?**

**13. Which categories have products with the highest discount?**

**14. Identify the top five (5) products in terms of rating and number of reviews Combined.**