**Project Topic: Amazon Product Review Analysis**

**Review**

The Amazon data for this project is scraped from Amazon product pages.

**Aim: To analyze product and customer review data**

**Objectives:**

To generate insight that can guide product improvement, marketing strategies

And improvement of customer engagement.

**Data Description**

The Amazon data includes:

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

• Customer engagement: user reviews, titles, and content

• Each row represents a unique product, with aggregated reviewer data stored as comma-separated values

Total Records: 1,465 rows

Total Fields: 16 columns

**Methodology**

* **Data cleaning and calculated columns**
* **Exploratory data analysis**
* **Reporting and visualization of findings (Dashboard)**
* **Discussion, conclusion and recommendation**

**Tools**

**Ms excel**

Lets start by reviewing the data set to see what we will be needing, and to know areas that requires cleaning

1. **Data cleaning** 
   1. First remove duplicate from the product Id by: Control + A 🡪 Data tap🡪 Data tool🡪 Click on remove duplicate, a dialogue box pops up🡪 Unselect all then select products ID only then ok. Results shows 114 duplicate with 1351 unique value left
   2. The category column comprises of both category name and sub- category name which needs to be separate. To do this, copy the category column to a different sheet and clean.

Highlight category column 🡪 Data tap 🡪 Data tool 🡪 text to column🡪 Click delimiter and next on the dialogue box 🡪 another dialogue box pops up, select others and paste the delimiter then next 🡪 finish

Results shows category separated from sub- categories up to six.

Go back to the Amazon data insert columns equivalent to number of categories and sub-categories splitted. Copy the separated category and sub- category columns and paste. Then we delete the old combined category.

* 1. Next filter discount, actual price, discount percentage, rating and rating count columns to check for any irregularities in data input. By highlighting each column 🡪data tap 🡪click on filter, under sort and filter, scroll through the values for any irregular or error in data inputs. Which usually appears at the bottom, also check for blank space.

**Results** : discounted price checked, actual price had error in one of the data, 1,39,900 entered instead of 139,900 in the last cell. To correct this? Click on the number ->ok and correct it, discount percentage checked, in rating column,

A symbol (/ ) was detected and the entire roll with the symbol was deleted.

The rating count column, Has two missing values which were deleted

**A2. Calculated columns**

1. Insert a column beside rating count, to capture the total potential revenue. Which is Actual price (K2)\* rating count (P2), (given) I.e.

= K2\* P2

Where K2 is the first row under Actual price column and P2 first row under rating count column

Then autofill down the column

1. Insert another column beside discount percentage, to capture Discount percentage of 50% and above.

Click on the first row and enter the IF formula

=if(M2,>= 50%, True, if(M2<= 50%, False)) then click enter and auto fill the column down.

M2 here represent discount percentage. Where discount percentage is 50% or more, returns true while less than 50% returns false.

To input the Indian currency symbol on Discounted price, account price and total potential revenue: Highlight Discounted\_price, account\_price, press control and click on the total potential revenue to get the 3 columns highlighted,

Click on the Home tap 🡪Click on dollar sign under custom 🡪More accounting format. A dialogue box pops up

Scroll down the Symbol and select the English Indian currency, then ok. This applies the currency on the highlighted columns.

1. Insert a column to capture price range Bucket.

Click on the first row and insert the formula

=IF(K2<200,"<₹200",IF(AND(K2>=200,K2<=500),"₹200 - ₹500",">₹500")) then click enter

Note K2= first row on actual\_ price column.

1. Insert the Range of rating column and insert the if formula = IF(P2<=1.4,1,IF(P2<=2.4,2,IF(P2<=3.4,3,IF(P2<=4.4,4,IF(P2<=5.4,5,6))))) Then enter. Note P2= rating column
2. Insert a calculated column to capture Ranges of number of review = IF(R2<1000,”Fewer than 1000”,

=IF(R2>=1000,”more than 1000 reviews")), Enter and auto fil down the column

1. Insert calculated column to combine rating and review.

=Average Rating + (Rating Count / Scaling Factor), (choose a factor like 1000 to balance weight).

After cleaning save as table by highlighting the entire data, then Control A and use control S to save.

1. **Data Analysis**
2. To determine the average Discount Percentage by category

Go to insert--->pivot table ---> pivot table ---> fill in the dialogue box with range and click on new sheet then ok.

* Click and drag category to rows
* Click and drag discount % to the values -> click on the default sum ---> value field settings --->click average then ok.

1. To determine the number of product per category?

Copy previous pivot table and paste where you want your result to appear.

Click on any cell and then fill the pivot table box.

Click and drag category into rows and product\_id into the value . This counts the number of product\_id which is a unique identifier of the product. Though both products id and product name gives same result

1. To determine the total number of Reviews per product category

Copy and paste pivot table, click and drag category to rows and rating\_count to values.

Note. Each product has aggregated reviewers with a Review id, it is assumed that rating count is equivalent to number of reviews per product.

1. To determine Which products have the highest average ratings

Pivot table-> click and drag product id to rows and ratings to values (go to value field and set to average -> ok

Sort Average Rating (from largest to smallest) and record findings

1. To determine the average actual price vs discounted price by category

Pivot Table -> drag Category to rows-> Drag Actual Price and Discounted Price to values→values field and set to Average for both value columns

1. To determine which products have the highest number of reviews

Sort Rating Count column in descending order or go to pivote-> drag product category and products id to rows, -> rating count to values.

Sort from highest to lowest.

Go to pivot table tool-> Design-> report layout-> show as outline form.

1. How many products have a discount of 50% or more

Pivot table-> drag discount % of 50 or more to row and products id to value-> count

Restful shows Trues for products that meets the if criteria and false for those that did not.

1. What is the Distribution of product ratings

Pivot Table: drag Ranges of Rating to row and product id to value-> count

1. What is the total potential revenue by category

Pivot table –>drag products category to rows, Total potential revenue (calculated) to values-> sum

1. What is the total number of unique products per price range

Pivot table-> drag price range (calculate) to rows-> product id to value-> count

1. How does the rating relate to the level of discount.

Highlight the percentage discount column, control and highlight rating column.

Go to insert-> scatter chart.

Format the chat

1. How many products have fewer than 1000 reviews

Pivot table-> drag Ranges of number of review (calculated) to rows, drag product id-> to value-> count.

1. Which categories have products with the highest discounts

Pivot Table:

Rows: Category, Values: Discount % → sum-> show values as %grand total

1. Identify the top 5 products in terms of rating and number of reviews combined.

Create calculated column to capture rating and number of reviews

=Average Rating + (Rating Count / Scaling Factor), (choose a factor like 1000 to balance weight). Then sort descending and pick top 5.

Or pivot table, drag product id to rows, sum of rating and number of reviews (calculated) then sort from largest to lowest, pick top 5.

1. **Results/Findings**
2. Products category, Home improvement has the highest average discount of 0.58%, followed by computers&Accessories, and Health &Personal care, with average discount percentage of 0.53% both, Toys&Games has the least average % discount of 0%
3. Electronics category has the highest number of product (490), Home&Kitchen (447), computer&Accessories (373), office products (31), Home Improvement and musical instruments, both have 2 product each, while Toys&Games, Car&Motorbike, as well as Health&PersonalCare have 1 product each.
4. Electronics, Computer%Accessories, Home%Kitchen, the highest number of reviews respectively with Car%Motorbike having the least reviews.
5. Product Id B0BP7XLX48 and B09ZHCJDPI have the highest average ratings of 5 each.
6. Electronics category has the highest actual price of 10,418, with an average discounted price of 6,226, while Toys&Games with the least actual price of 150, has discounted price of 150
7. Products id (B014I8SSD0, BO14ISX4Y, and B07KSMBL2H) All belonging to the Electronics category have the highest number of reviews, with 426973 reviews each.
8. 660 products have percentage discount of 50% and above, while 688 products have percentage discount of less than 50%.
9. About 90% (1213/1348), of the Amazon products have been rated between 3.5 to 4.4, and 7% (95/1348) have ratings between 4.5 to 5.4. While, 2.8% (38/1348) of the products, where rated between 2.5 to 3.4, and only 2 products where rated the least between 1.5 to 2.4.
10. Electronics, Computer&Accessories and Home&Kitchen categories, have the highest total potential revenues of (91, 11 and 10 billions) Respectively, while Car&Motorbikes, and Toy&Games, have the least total potential revenue of (4.5 and 2.4 millions) respectively.
11. A total of 1163 products fall within price range of more than 500, 151 products fall within price range of 200 to 500 while 34 products fall within price range of less than 200.
12. Rating has a linear relationship with level of discount i.e the is no significant relationship between rating of Amazon products with the discount percentage or level of Amazon products.
13. A total of 1041 products have more than 1000 reviews while 307 products have less than a thousand reviews.
14. Electronics category has the highest discounts percentage, with sum of discount percentage of 38.88%, Computer&Accessories (31.48%) and Home&Kitchen with 28.55% Discount. Health &personalCare, Car&Motorbike and Toys&Games have the least discount % of 0.08%, 0.07% and 0.00% respectively
15. The top 5 products in terms of rating and number of reviews combined are all within the Electronics category with products ids, B, B and B having the highest of 431 points while product id, B and B have 368.

**Discussion**

Amazon has a total of 1348 products, distributed across 9 products categories, with price range of less than 200, 200- 500, and more than 500. **Table 2&10,** respectively. Amazon products have a total average discount percentage of 0.47%,**Table 1.** With a good number of products (660/1348), having a discount percentage of more than 50%, **figure Q7**. Electronics category has the highest discounts percentage, with the sum of discount percentage of 38.88%, and average discount of 0.50%, Computer&Accessories has total discount percentage of 31.48%, with an average discount percentage of 0.53%, and Home&Kitchen with 28.55% Discount, with an average percentage discount of 0.4%, **.** This might contribute to the high customer engagements resulting to high number of reviews. While Health&personalCare, Car&Motorbike and Toys&Games have the least discount % of 0.08%, 0.07% and 0.00% respectively, this indicates the need to review the discount level of products with less percentage discounts as they contribute less to the company’s revenue.

A total of 1041 products have more than 1000 reviews, with the Electronics category, Computer&Accessories, and the Home&Kitchen categories having the highest number of reviews respectively, This indicates high customer engagement with respect to these product categories and could be attributed to high demands and percentage discounts of this products categories as seen in **table 3&13.** These products also have high earnings capacity as seen in **Table 9.**

About 90% (1213/1348), of the Amazon products have been rated highly between 3.5 to 4.4, **Table 8.** This indicates customers satisfaction with the Amazon products and services, however there is still room for improvement to archive highest product ratings. This can be done via customers feed backs and products evaluation surveys, to help Amazon to better understand customers needs and how to meet those needs, in order to serve customers better. The top 5 products in terms of rating and number of reviews combined are all within the Electronics category with products ids, **B014I8SSD0, B014I8SX4Y and B07KSMBL2H** having the highest of 431 points each, while product id, **B07GPXXNNG**, **B07GQD4K6L and** **B071Z8M4KX**, have 368 points each. However, there is no significant relationship between rating of Amazon products and the discount percentage or level, of Amazon products, **figure Q11.**

Electronics, Computer&Accessories and Home&Kitchen categories, have the highest total potential revenues of (91, 11 and 10 billions) Respectively, while Car&Motorbikes, and Toy&Games, have the least total potential revenue of (4.5 and 2.4 millions) respectively, **Table 9**. This indicates maximum earning capacity of products with high potential revenue, which should be the focus for resource allocations towards maximizing profit, and more of the products should be made available to meet customers demand. However, products with low potential revenue indicates areas for growth and improvement, to improve the earning capacity of such products. As well as the need to put in place market strategies to create more awareness and improve customer engagements of such products.

**Conclusion**

Amazon products demonstrate high quality products, indicated by high customer engagement via reviews and high ratings. With good percentage discount on products. Products with high potential revenue should be prioritized in terms of resources allocation and production to meet demands and maximize profit. However there is still room for more reviews and improvement on product categories such as Toys&Games, and other products with low discount percentage, low ratings, low reviews and low potential revenue.

**Recommendation**

Products with low percentage discount should be reviewed, to ensure the organization is not running on a loss.

Conduct more surveys and get customers feedback on products with low ratings and low customer engagement, this will help Amazon to better understand customer needs and areas of product improvement.

Review best practice and build on it to sustain and improve quality of highly rated products with high potential revenue, and also to maintain the high customer engagements.

For Analysis Dashboard, [Click Here]()

Acknowledged and Appreciation

I will like to Appreciate God almighty for the grace and good health that sustained me through this course. A big thank you to God's servant Daddy G.O for making this program freely available and accessible for someone like me. And the facilitators, Mr Mushin, Thank you, and to all the DSA crew and those who contributed in one way or the other. May God bless you all.