MYNTRA PRODUCT ANALYSIS

CAPTURED BY SAPNA B

Data Fetching

Data Cleaning & Preparation

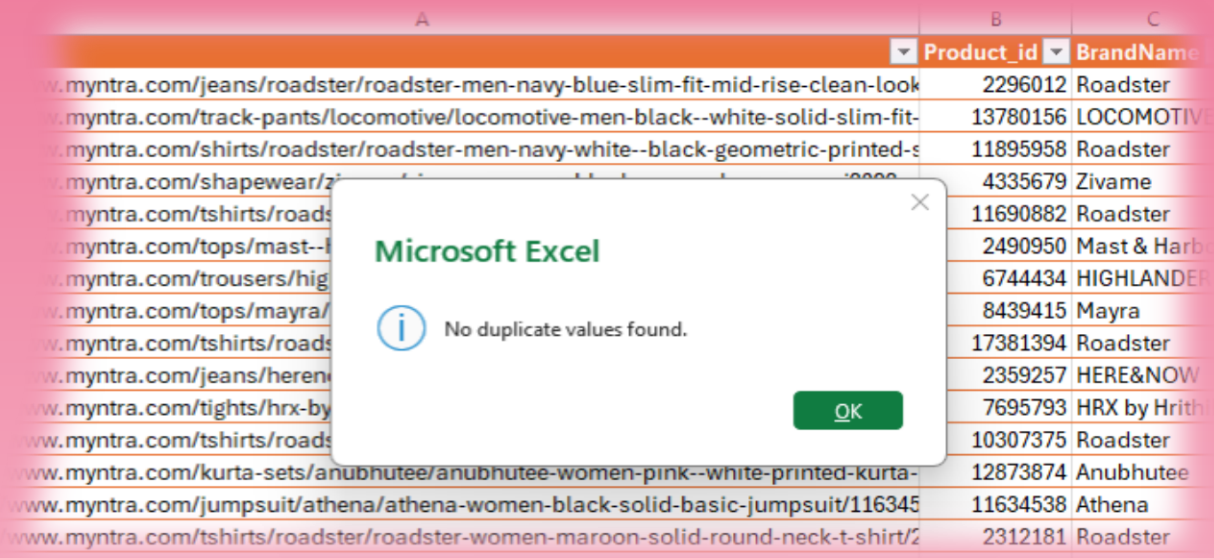
Data Retrieval & Lookup

Data Analysis

Data Cleaning & Preparation

Problem: Check for duplicate values in your dataset and remove them.

Solution: *A thorough examination of the dataset was conducted to identify any duplicate entries. The analysis confirmed that no duplicates were present, ensuring the dataset's accuracy and reliability for further processing..*



Data Cleaning & Preparation

Problem: Standardize the "DiscountOffer" column to a single format, ensuring all values are uniform.

J	K
Discounter	Std_Discount Offer (in Rs)
45%	675
55%	632
31%	401

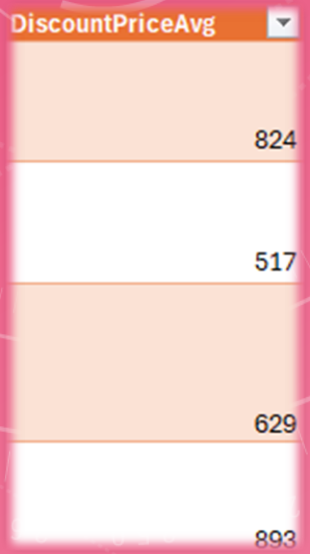
Solution: *Basically removed irrelevant text like "Off" and "Off Hurry!" to retain only the numeric portion of the discount.*

1. *The following Excel formula was applied to derive a clean numerical value:*

```
Std_Discount_Offer  =IF([@Discounter]<> "",
                        IF(ISNUMBER(SEARCH("Rs.", [@Discounter])),
                        VALUE(SUBSTITUTE([@Discounter], "Rs. ", "")),
                        [@[OriginalPrice (in
                        Rs)]]*VALUE(SUBSTITUTE([@Discounter], "%", " "))), 0)
```

Data Cleaning & Preparation

Problem: Identify rows where both "DiscountPrice" and "DiscountOffer" are null and fill the "DiscountPrice" with the average discount price of the respective category.



=IF(AND(ISBLANK([@[DiscountPrice (in Rs)]]), [@[Std_Discount Offer (in Rs)]] = 0), IF(COUNTIFS(D:D, [@Category], I:I, "<>") > 0, ROUND(AVERAGEIFS(I:I, D:D, [@Category], I:I, "<>"), 0), ""), [@[DiscountPrice (in Rs)]])

Solutions:

Check if Discount Price is missing and Standard Discount Offer is zero:

- *ISBLANK([@[DiscountPrice (in Rs)]])*
- *[@[Std_Discount Offer (in Rs)]] = 0*

If both conditions are met:

- *Look for other non-blank discount prices within the same Category.*
- *Use AVERAGEIFS to calculate the average of available prices in that category.*
- *ROUND the average to the nearest whole number.*

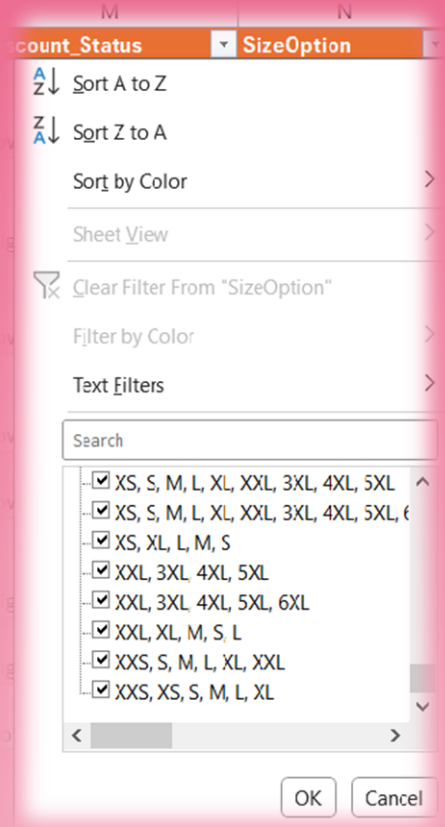
If conditions are not met:

- *Use the existing Discount Price from the row.*

Data Cleaning & Preparation

Problem: Replace all null values in the "SizeOption" column with the text "Not Available."

Solution: *A data filter was applied to the "SizeOption" column to identify any NULL (empty or missing) entries.*
After filtering, it was confirmed that ***no NULL values are present*** *in the column. Therefore,* ***no replacement was necessary*** *as the data is already complete in this field.*



Data Analysis

Problem: Calculate the overall average original price for products with ratings greater than 4.

Solution: *To solve this, we used the AVERAGEIF function in Excel to filter and calculate the required value:*

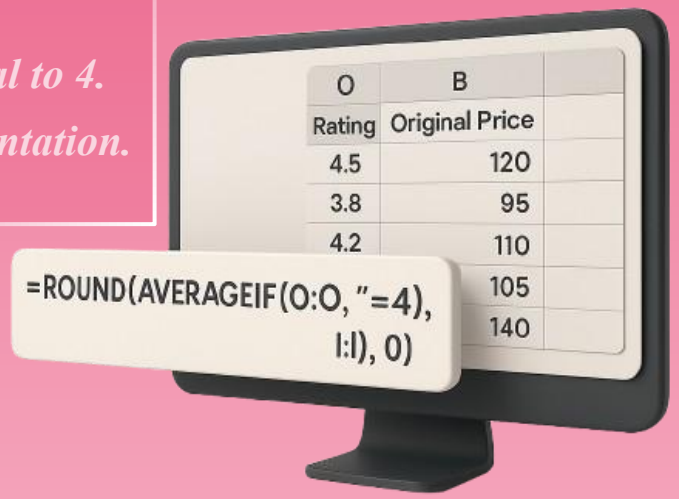
AvgPrice_4PlusRating=ROUND(AVERAGEIF(O:O, ">=4", I:I), 0)

O:O refers to the column containing product ratings.

I:I refers to the column with the original prices.

The formula calculates the average original price only for products with a rating greater than or equal to 4.

The ROUND(..., 0) function is used to round the result to the nearest whole number for cleaner presentation.



AvgPrice_4PlusRating	1638

Data Analysis

Problem: Count the number of products with a discount offer greater than 50% OFF.

Solutions: *We use the COUNTIF function in Excel to count how many products have a discount value greater than or equal to 50% (i.e., >= 0.5).*

Discount_Above50_Count=COUNTIF(K:K,">=0.5")

Discount_Above50_Count	3544

Explanation:

K:K refers to the column containing the discount percentage for each product (in decimal form, e.g., 0.50 for 50%).

The formula counts how many rows in column K have values greater than or equal to 0.5, indicating products with more than 50% discount.

Data Analysis

Problem: Count the number of products available in size "M."

Solution: *We use the COUNTIF function in Excel to count the number of entries in the dataset where size "M" is mentioned.*

=COUNTIF(O:O, "*M*")

Total_Size_M_Items	3447

- *O:O is the column containing size information (e.g., "S", "M", "L", or combinations like "S,M,L").*
- *"*M*" looks for any cell that contains "M" (regardless of whether it's alone or part of a group).*
- *The formula returns the total count of products where size "M" is available.*

Data Analysis

Problem: Create a new column to label the products as "High Discount" if the discount offer is greater than 50% OFF, otherwise label them as "Low Discount."

Solution: ***D****iscount Status We will use an IF formula in Excel to derive the new label based on the discount value in the Discounter column.*

Formula (for the column: Discount_Status):

```
=IF(NOT(ISBLANK([@Discounter])),
        IF([@Discounter]>=50%, "High Discount", "Low Discount"),
        "No Discount")
```

- *NOT(ISBLANK([@Discounter])): Checks if the discount cell is not empty.*
- *IF([@Discounter]>=50%, "High Discount", "Low Discount"): Applies the discount logic.*
- *If the discount is blank or null, it assigns "No Discount".*

Data Retrieval and Lookup

Problem: Use VLOOKUP/XLOOKUP to find the product brand, price, and rating of the product with Product_id "11226634".

Solution: *We will use the XLOOKUP function to find and return the respective details for the product with Product ID "11226634" stored in cell V6*

Retrieve Brand Name:

=XLOOKUP(V6, B:B, C:C)

This returns the brand name from column C by matching the Product ID in column B.

Retrieve Price:

=XLOOKUP(V6, B:B, I:I)

This returns the price of the product from column I.

Retrieve Rating:

=XLOOKUP(V6, B:B, O:O)

This returns the customer rating from column O

XLOOKUP ID	11226634
Brand Name	Maniac
Price	1199
Rating	3.9

Data Retrieval and Lookup

Problem: Find the "DiscountPrice" for the product with the Product ID "6744434" using the INDEX and MATCH functions.

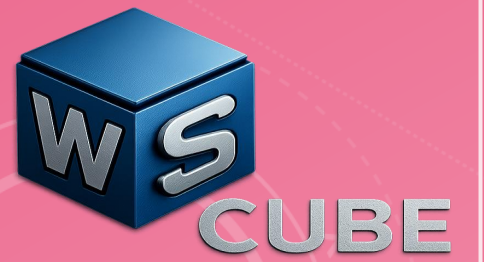
Solution: *We will use a combination of INDEX and MATCH to retrieve the Discount Price and round it to the nearest whole number.*

=ROUND(INDEX(K:K, MATCH(V12, B:B, 0)), 0)

INDEX and MATCH ID	6744434
Discount Price	899

Explanation:

MATCH(V12, B:B, 0) finds the row number where the Product ID in cell V12 (which is "6744434") matches the Product ID in column B.
INDEX(K:K, ...) retrieves the DiscountPrice from column K based on the row found by MATCH.
ROUND(..., 0) rounds the result to the nearest whole number for a cleaner presentation.
This formula ensures accuracy and flexibility in pulling data from non-adjacent columns.



Data Retrieval and Lookup

Problem: Utilize nested xlookup to find any column's detail of a product with it's product id.

Solution: We use nested XLOOKUP to first locate the product row using the Product ID and then look up the desired column name dynamically.

= XLOOKUP(TRIM(Y3), TRANSPOSE(A1:P1), TRANSPOSE(XLOOKUP(X3, B2:B5160, A2:P5160,"Not Found")),"Not Found")

Product ID	Product Details	Output
12222072	BrandName	SASSAFRAS

X3: Contains the Product ID to search for.

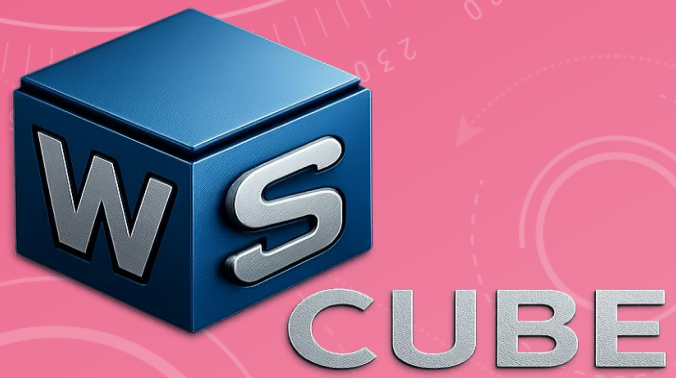
Y3: Contains the Column Name the user wants to retrieve (e.g., "Brand").

The inner XLOOKUP finds the full product row based on the Product ID.

The outer XLOOKUP locates the correct column value based on the column name entered.

TRANSPOSE is used to align data orientation between the header and the result row..

Thank You



Mentor: Ayushi Jain