
# Myntra

Presented by: Parshvi Jain

# About Myntra

Online Shopping Site for Fashion & Lifestyle in India. India's Fashion Expert brings you a variety of footwear, Clothing, Accessories and lifestyle products.

Myntra is a fashion destination that is committed to making you look good anytime, anywhere. Visit www.myntra.com or download the app to get the perfect look. It is an Indian fashion e-commerce company headquartered in Bengaluru, Karnataka, India. The company was founded in 2007-2008 to sell personalized gift items. In May 2014, Myntra.com was acquired by Flipkart.

# Project Questions

## A.Data Cleaning and Preparation

/. Check for duplicate values in your dataset and remove them.

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

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

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

## B. Data Analysis

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

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

3. Count the number of products available in size "M."

4. 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."

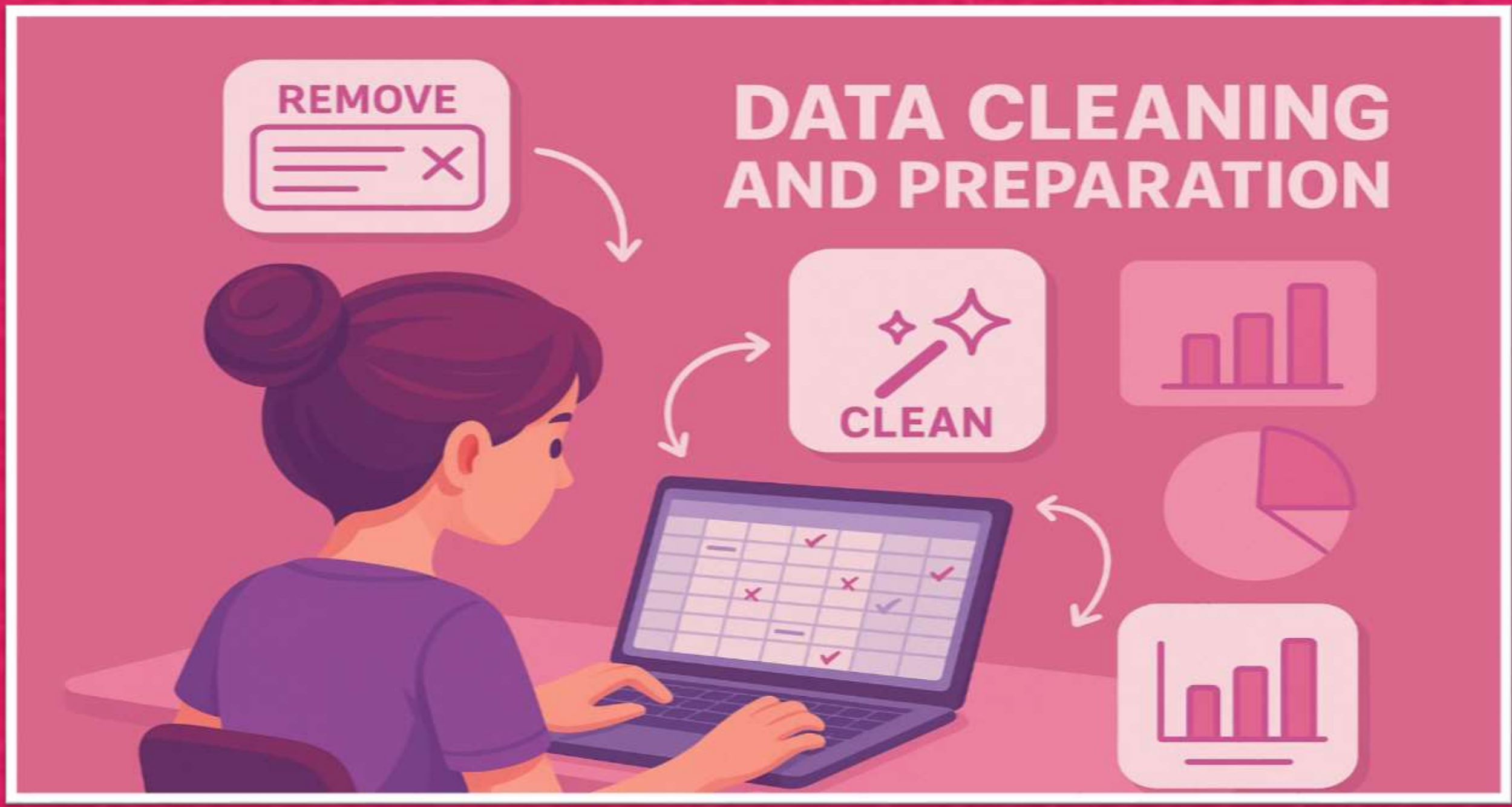
# C. Data Retrieval and Lookup

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

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

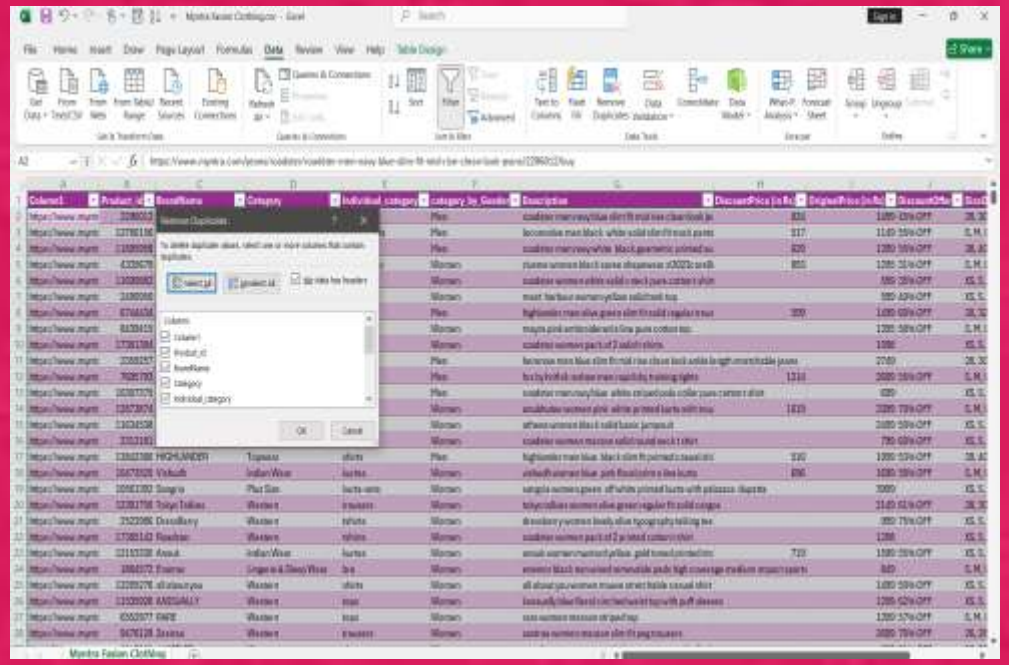
3. Utilize nested xlookup to find any column's detail of a product with it's product id.

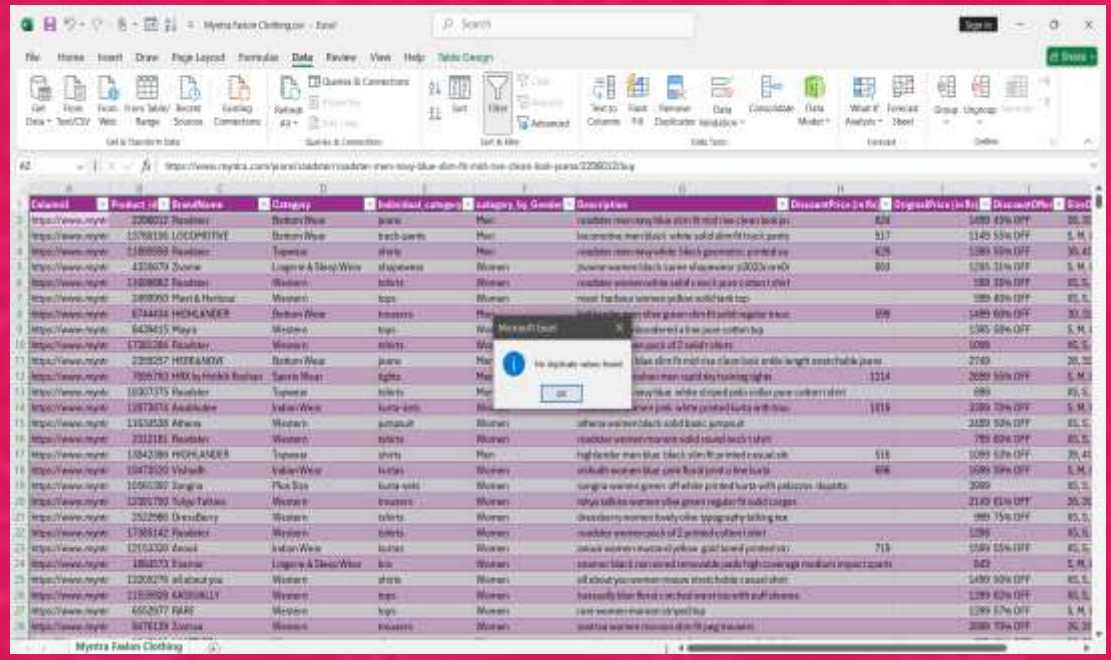
# DATA CLEANING AND PREPARATION



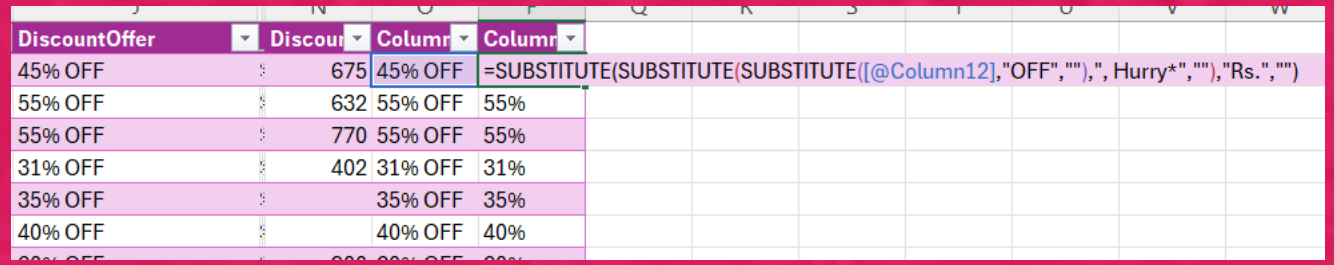
# Check for duplicate values in your dataset and remove them

~ Select the dataset, go to Data > Remove Duplicates, and choose all columns to check duplicates.





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



| DiscountOffer | Discoun | Columr | Columr |
|---|---|---|---|
| 45% OFF | 675 | 45% OFF | =SUBSTITUTE(SUBSTITUTE(SUBSTITUTE([@Column12],"OFF",""),", Hurry*",""),"Rs.","") |
| 55% OFF | 632 | 55% OFF | 55% |
| 55% OFF | 770 | 55% OFF | 55% |
| 31% OFF | 402 | 31% OFF | 31% |
| 35% OFF | | 35% OFF | 35% |
| 40% OFF | | 40% OFF | 40% |

To standardize discount offer firstly we have to remove OFF,Hurry,Rs by using substitute function.

1.First, where "Discount Price in Rupees" was blank, we calculated it using:
Original Price – Discount Price = Discount in Rupees.
2.Then, wherever the "Discount Offer" was blank, we checked:

1. If the "Discount Price" was available, we used that.
2. If not, we kept the existing "Discount Offer" value.

3.After that, we used the SUBSTITUTE function to clean the text — removing words like "off", "₹", and "Flat".

4.Finally, to complete the standardization of the "Discount Offer" column, we focused on converting all remaining flat discount values (in rupees) into percentage format.

To do this, we used logical functions like SEARCH, IF, and IFERROR:

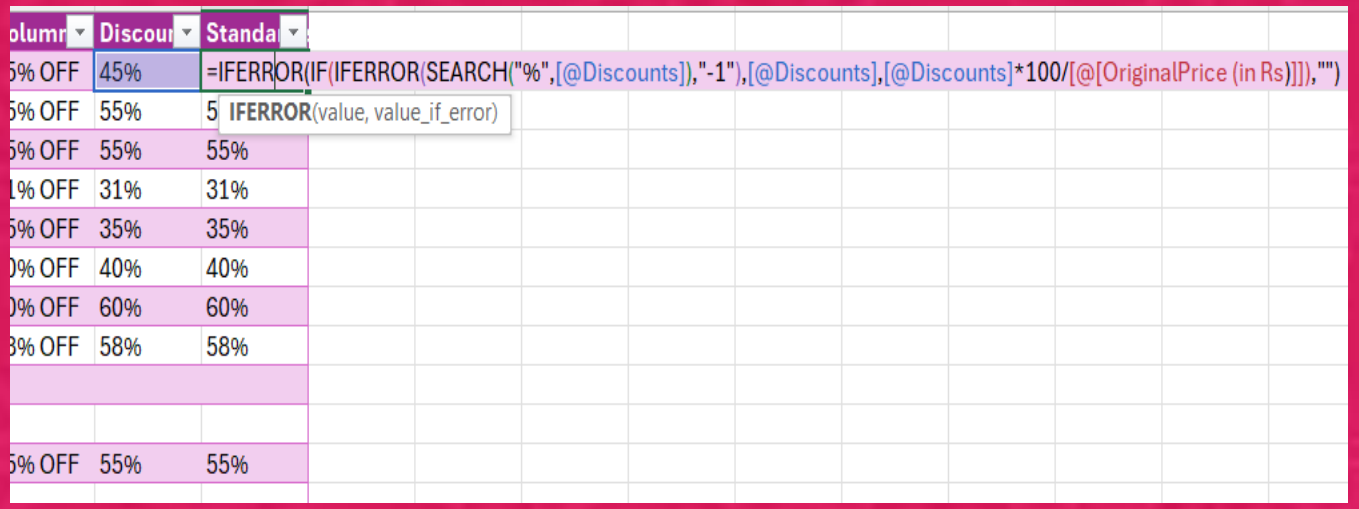
•First, we checked whether the value already contained a percentage symbol (like "45%").

If yes, we kept it unchanged, as it was already in the correct format.

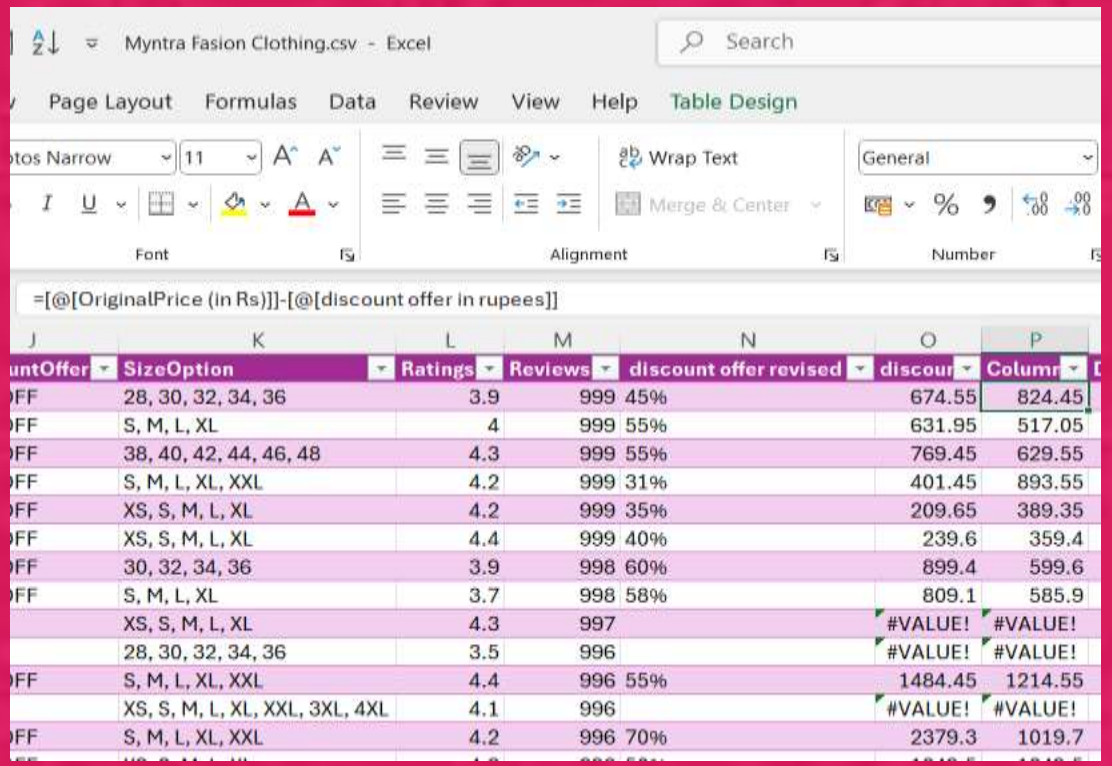
•If it did not contain a percentage, it meant the value was a flat discount in rupees (like "₹500").

So, we calculated the discount percentage using the available discount amount and original price.

•In case of any missing or incorrect values, we used error-handling logic to avoid blank or broken results.

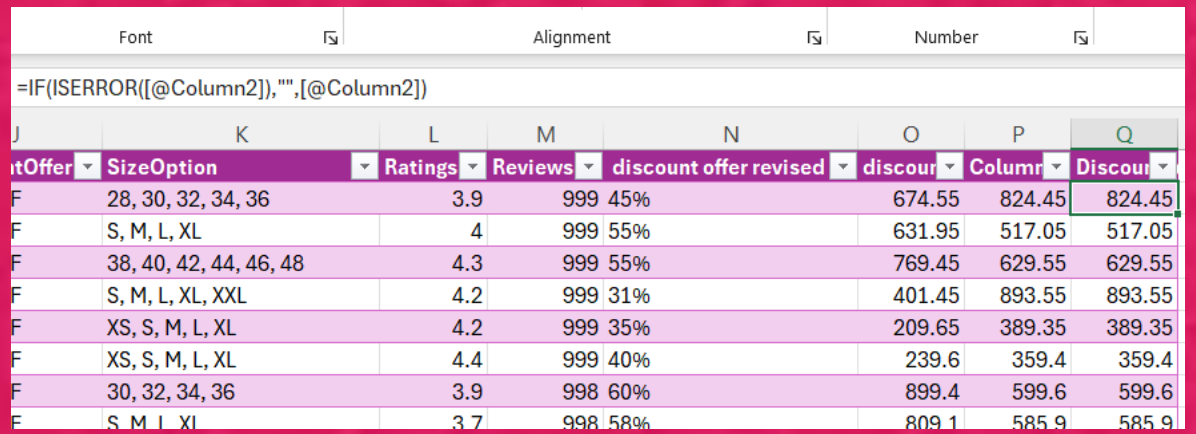


# Identify rows where both "DiscountPrice"and "DiscountOffer" are null and fill the "DiscountPrice" with the average discount price of the respective category.



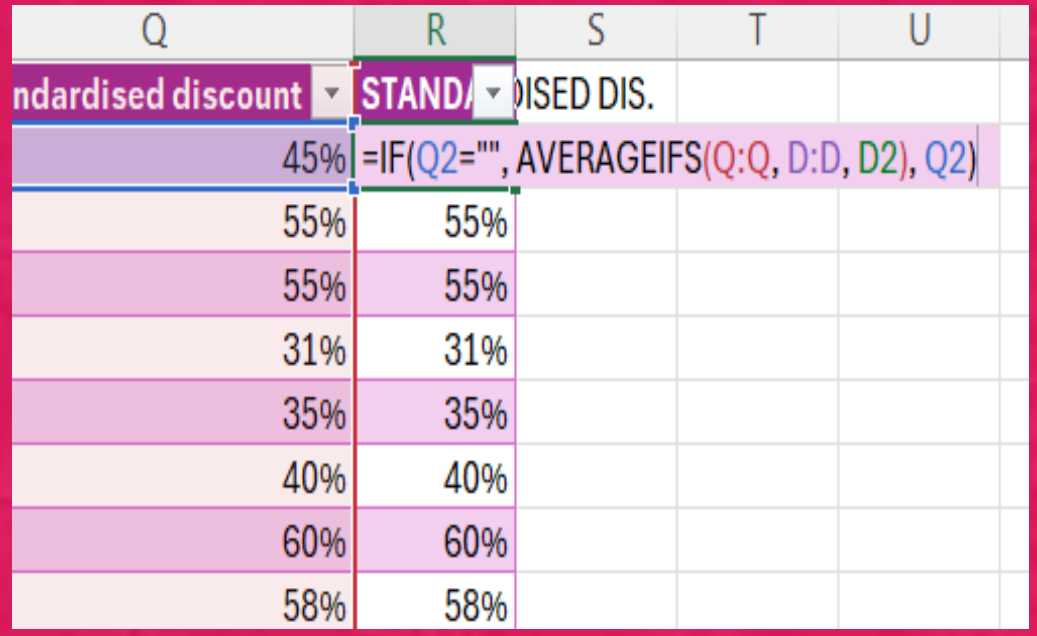

| J | K | L | M | N | O | P |
|---|---|---|---|---|---|---|
| untOffer | SizeOption | Ratings | Reviews | discount offer revised | discour | Columr |
| FF | 28, 30, 32, 34, 36 | 3.9 | 999 | 45% | 674.55 | 824.45 |
| FF | S, M, L, XL | 4 | 999 | 55% | 631.95 | 517.05 |
| FF | 38, 40, 42, 44, 46, 48 | 4.3 | 999 | 55% | 769.45 | 629.55 |
| FF | S, M, L, XL, XXL | 4.2 | 999 | 31% | 401.45 | 893.55 |
| FF | XS, S, M, L, XL | 4.2 | 999 | 35% | 209.65 | 389.35 |
| FF | XS, S, M, L, XL | 4.4 | 999 | 40% | 239.6 | 359.4 |
| FF | 30, 32, 34, 36 | 3.9 | 998 | 60% | 899.4 | 599.6 |
| FF | S, M, L, XL | 3.7 | 998 | 58% | 809.1 | 585.9 |
| | XS, S, M, L, XL | 4.3 | 997 | | #VALUE! | #VALUE! |
| | 28, 30, 32, 34, 36 | 3.5 | 996 | | #VALUE! | #VALUE! |
| FF | S, M, L, XL, XXL | 4.4 | 996 | 55% | 1484.45 | 1214.55 |
| | XS, S, M, L, XL, XXL, 3XL, 4XL | 4.1 | 996 | | #VALUE! | #VALUE! |
| FF | S, M, L, XL, XXL | 4.2 | 996 | 70% | 2379.3 | 1019.7 |

As we have standardize the "Discount Offer" Column now we can get the most of the values of "DiscountPrice" column by subtracting "Discount Offer " from "OriginalPrice" and for those cells that are blank we use the iserror formula.




| J | K | L | M | N | O | P | Q |
|---|---|---|---|---|---|---|---|
| tOffer | SizeOption | Ratings | Reviews | discount offer revised | discour | Columr | Discour |
| F | 28, 30, 32, 34, 36 | 3.9 | 999 | 45% | 674.55 | 824.45 | 824.45 |
| F | S, M, L, XL | 4 | 999 | 55% | 631.95 | 517.05 | 517.05 |
| F | 38, 40, 42, 44, 46, 48 | 4.3 | 999 | 55% | 769.45 | 629.55 | 629.55 |
| F | S, M, L, XL, XXL | 4.2 | 999 | 31% | 401.45 | 893.55 | 893.55 |
| F | XS, S, M, L, XL | 4.2 | 999 | 35% | 209.65 | 389.35 | 389.35 |
| F | XS, S, M, L, XL | 4.4 | 999 | 40% | 239.6 | 359.4 | 359.4 |
| F | 30, 32, 34, 36 | 3.9 | 998 | 60% | 899.4 | 599.6 | 599.6 |
| F | S, M, L, XL | 3.7 | 998 | 58% | 809.1 | 585.9 | 585.9 |

Fill the "DiscountPrice in rupees" with the average discount price in rupees of the respective category.



| Q | R | S | T | U |
|---|---|---|---|---|
| ndardised discount | STANDA | )ISED DIS. | | |
| 45% | =IF(Q2="", AVERAGEIFS(Q:Q, D:D, D2), Q2) | | | |
| 55% | 55% | | | |
| 55% | 55% | | | |
| 31% | 31% | | | |
| 35% | 35% | | | |
| 40% | 40% | | | |
| 60% | 60% | | | |
| 58% | 58% | | | |

=IF(Q2="", AVERAGEIFS(Q:Q, D:D, D2), Q2)

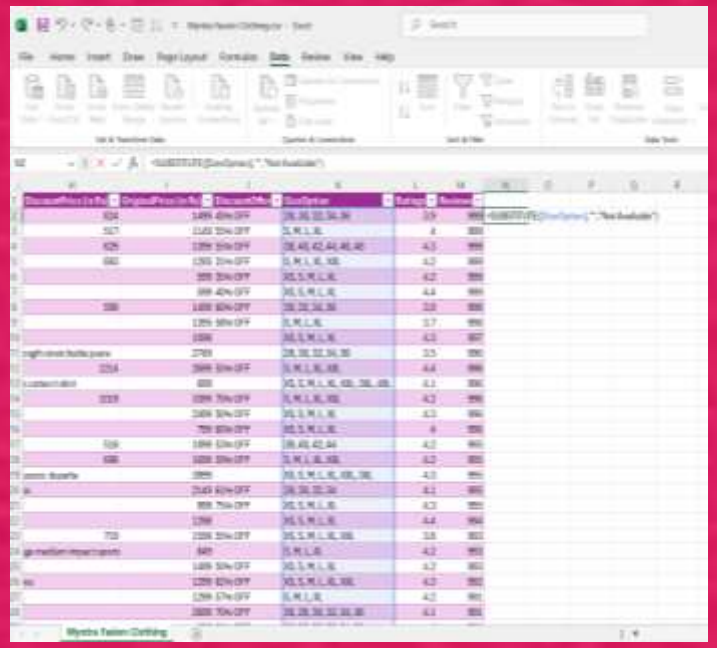
- Q2="" → Checks if the "Standardise" value is blank.
- AVERAGEIFS(Q:Q, D:D, D2) → Calculates the average of all
- "Standardise" values where the category (from column D) matches the current r
- Q2 → If the value is not blank, it keeps the existing value.

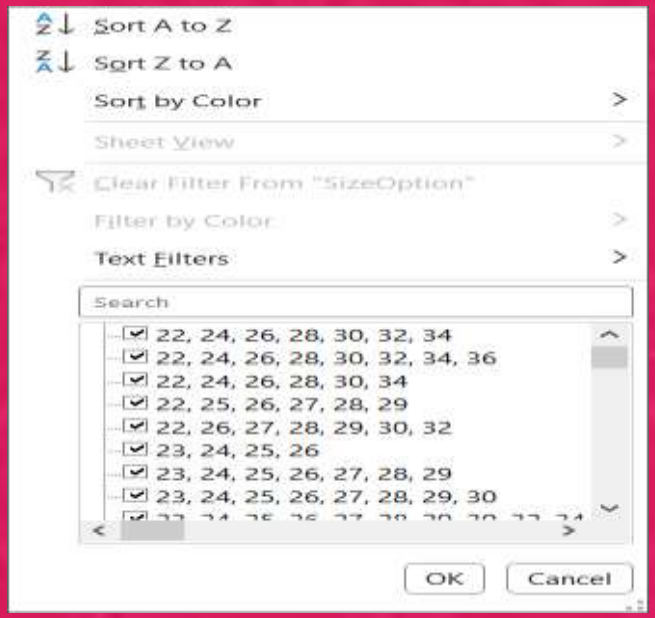
Now all blank "Standardise" cells are filled using the average discount of their respective category!

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

1.Identifying the null values of" SIZE OPTION" and replacing it with "Not Available".
Using Formula
=SUBSTITUTE([SizeOption],"","Not Available")

2.Upon checking the dataset, there are no null or blank values in the "SizeOption" column.
Hence, no replacement is required.



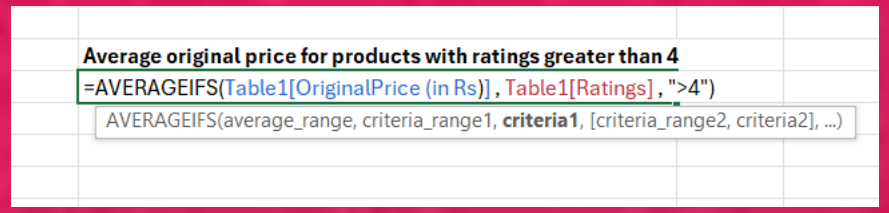


# DATA ANALYSIS

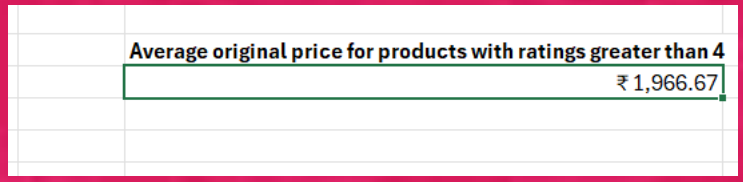
# Calculate the overall average original price for products with ratings greater than 4.

By using Averageifs formula -

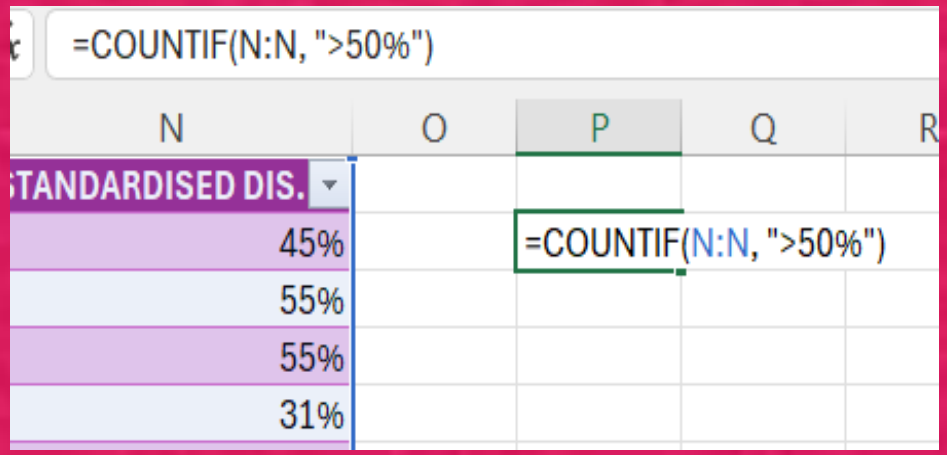
=AVERAGEIFS(Table1[OriginalPrice (in Rs)], Table1[Ratings], ">4")




The average original price of products with ratings greater than 4 is ₹1966.67.



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

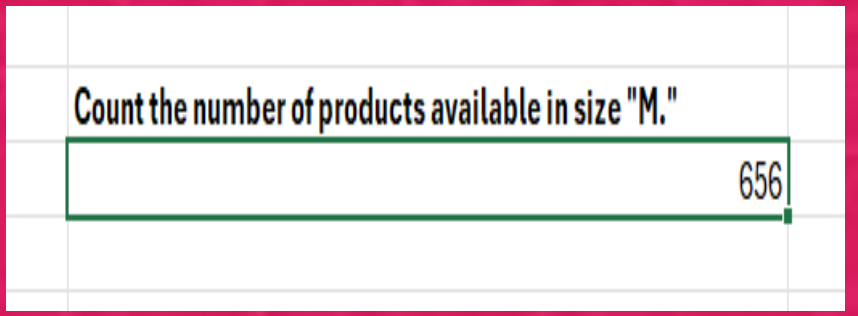



- This formula checks each value in column N
- It counts how many products have a discount greater than 50%.
- We used the Standardise column which contains clean percentage values for discounts.
- The result gives the total number of products with more than 50% OFF.

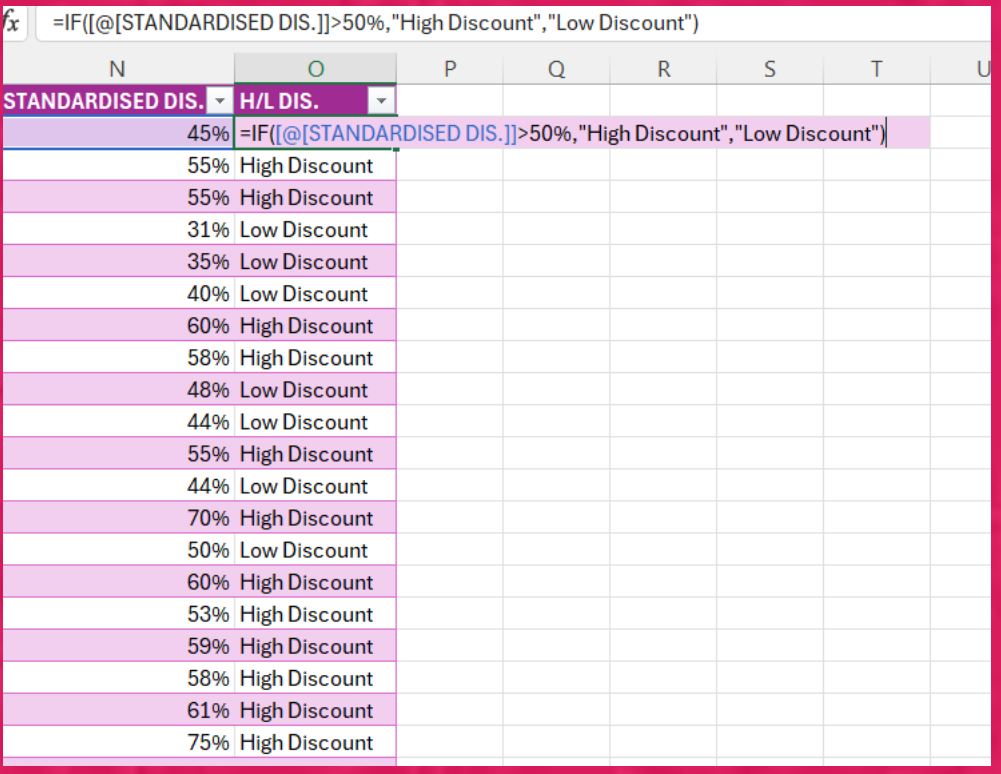
# Count the number of products available in size "M."

We will use this formula to find the number of products available in size M:
=COUNTIF(Table1[SizeOption],"M")


~The number of products available in size M is 656.



# 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."




| STANDARDISED DIS. | H/L DIS. |
|---|---|
| 45% | =IF([@[STANDARDISED DIS.]]>50%,"High Discount","Low Discount") |
| 55% | High Discount |
| 55% | High Discount |
| 31% | Low Discount |
| 35% | Low Discount |
| 40% | Low Discount |
| 60% | High Discount |
| 58% | High Discount |
| 48% | Low Discount |
| 44% | Low Discount |
| 55% | High Discount |
| 44% | Low Discount |
| 70% | High Discount |
| 50% | Low Discount |
| 60% | High Discount |
| 53% | High Discount |
| 59% | High Discount |
| 58% | High Discount |
| 61% | High Discount |
| 75% | High Discount |

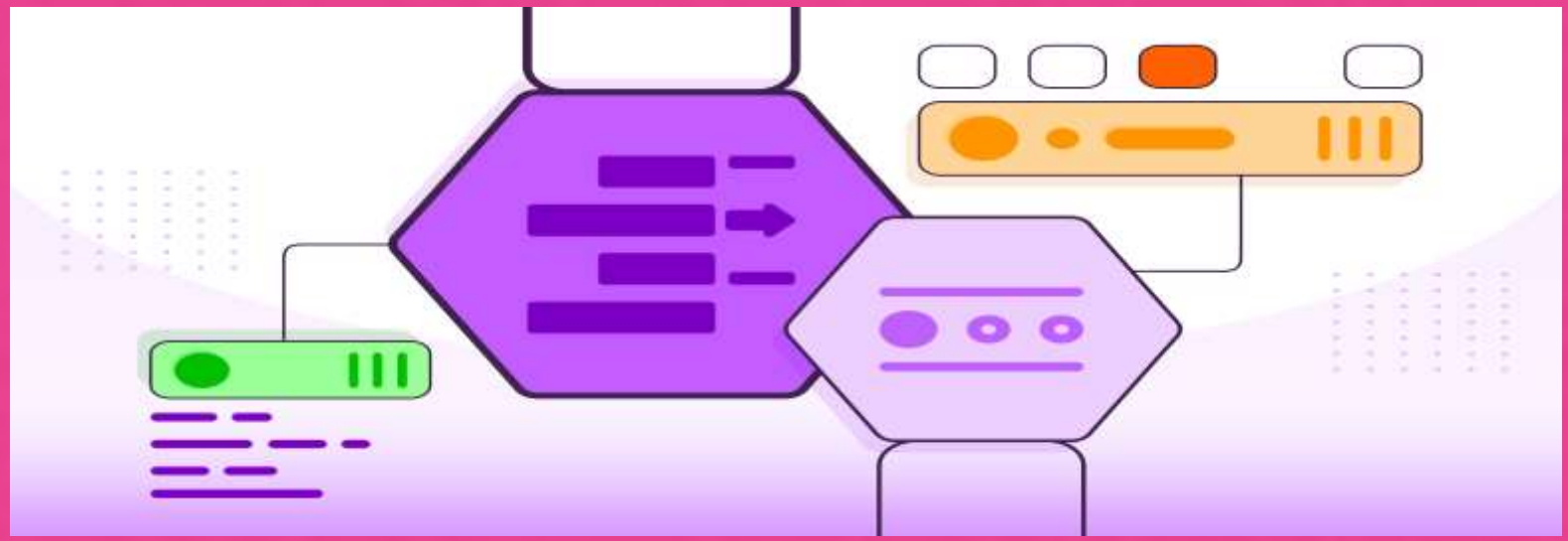
- A formula was applied to label the discount either as “High Discount”, “Low Discount”.

=IF([@[STANDARDISED DIS.]]>50%, "High Discount", "Low Discount")

◆ Outcome:

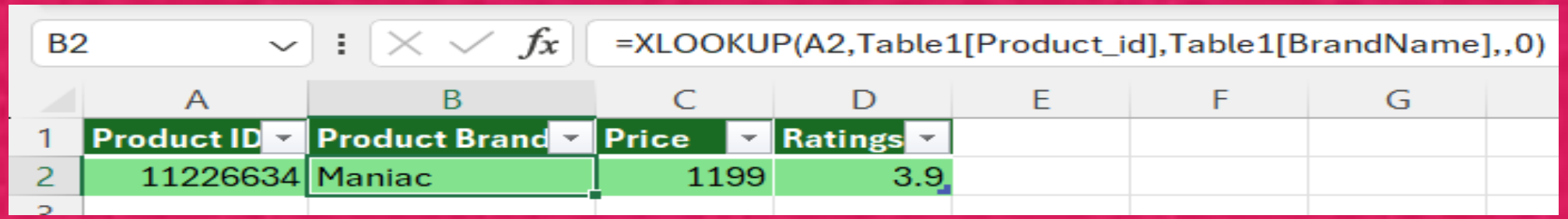
- If the discount is greater than 50%, it is labeled as High Discount.
- If the discount is 50% or less, it is labeled as Low Discount.
- A formula was applied to label the discount as either “High Discount” or “Low Discount.”

# Data Retrieval and Lookup



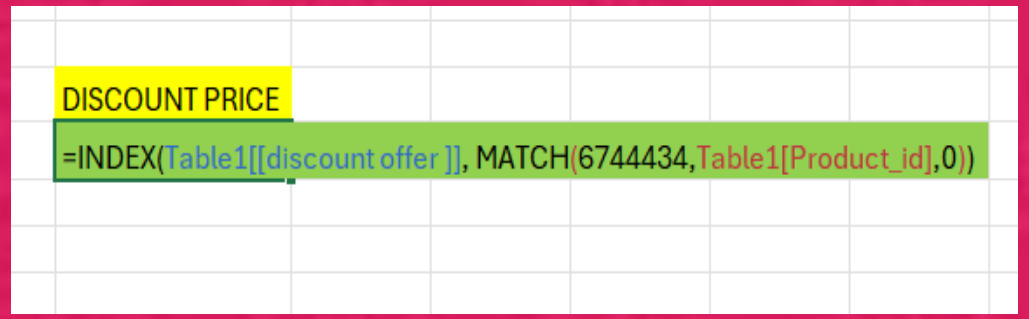
# Use VLOOKUP/XLOOKUP to find the product brand, price, and rating of the product with Product_id "11226634"

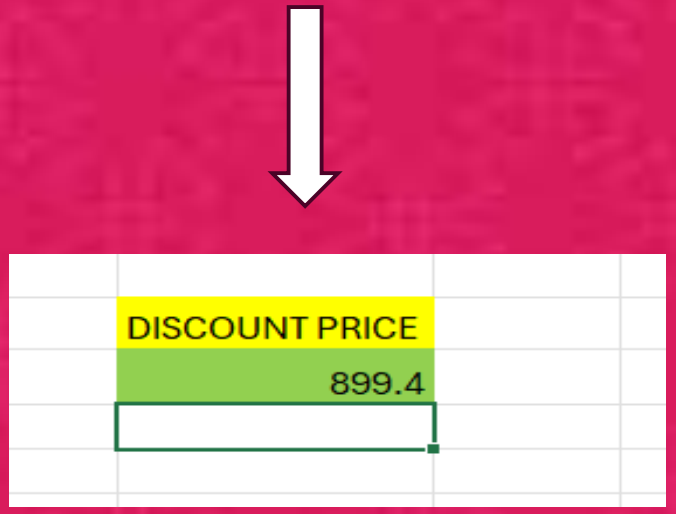
We used XLOOKUP to find product brand, ratings, price for product id 11226634




=XLOOKUP(A2,Table1[Product_id],Table1[BrandName],,0)

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

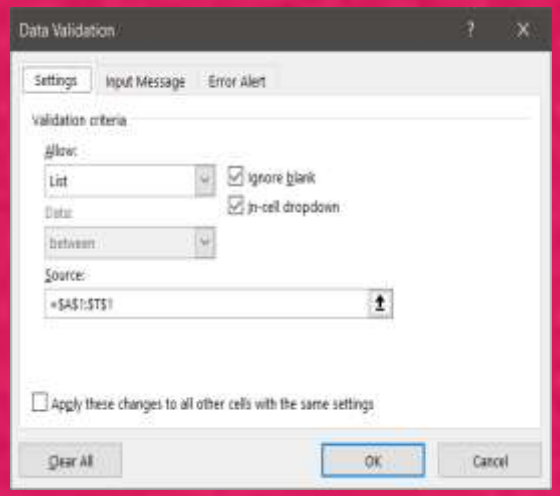


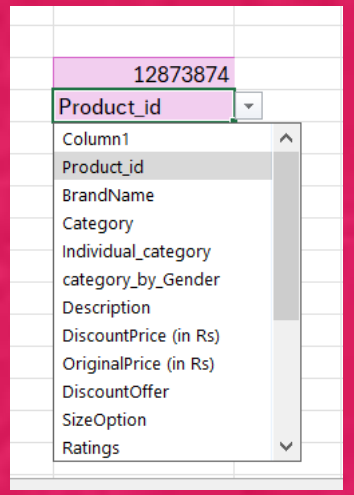



# Utilize nested xlookup to find any column's detail of a product with it's product id.

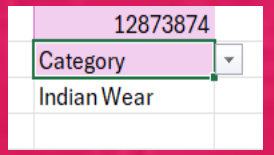
First, select the Product ID cell,
then go to the Data tab,
click on Data Validation,
choose "List",
and in the Source,
select all column headers from your table.

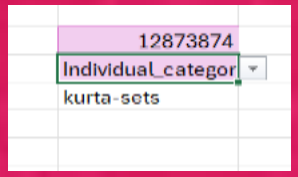
This sets up a dropdown where you can choose any column (like Brand, Rating, Price), and then use nested XLOOKUP to fetch values based on selected Product ID and selected column.








Thank
You!