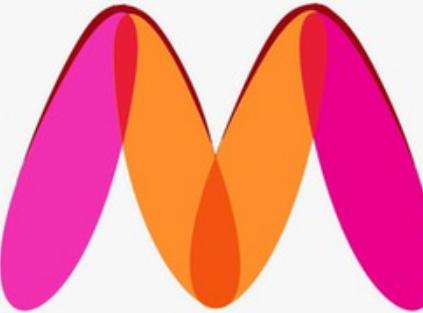




• • • •



Analysis of Myntra Apparel

• • • •

Presented By: Rahul Nagra

Introduction



Myntra is an Indian e-commerce company that sells fashion and lifestyle products online. It was founded in 2007-2008 to sell personalized gifts, but has since expanded to offer a wide range of products, including clothing, footwear, accessories, jewelry, and personal care products. Myntra works with over 6,000 brands, including H&M, Levis, Tommy Hilfiger, Nike, and Puma. The company is headquartered in Bengaluru, Karnataka, and services over 19,000 pin codes across India.



PROBLEMS

Project Questions

A. Data Cleaning and Preparation :-

1. 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



QA.1. :- Check for duplicate values in your dataset and remove them.

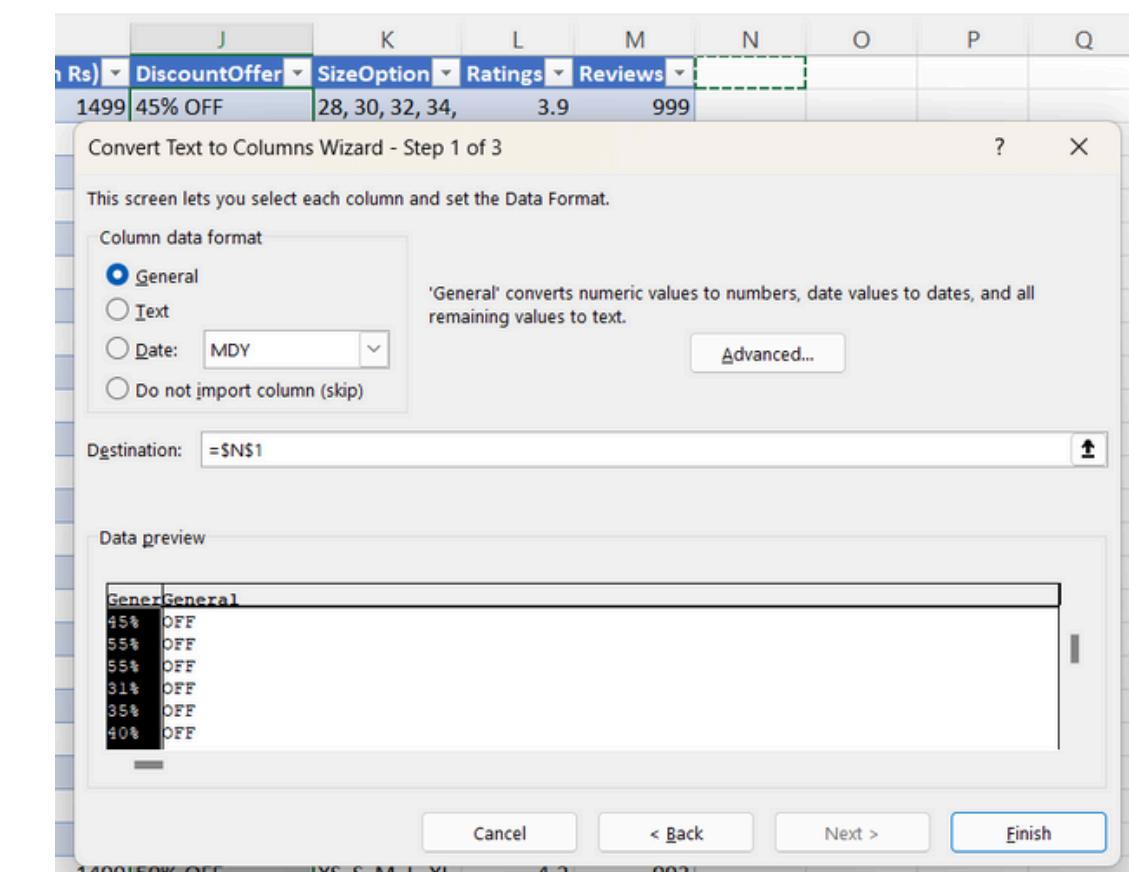
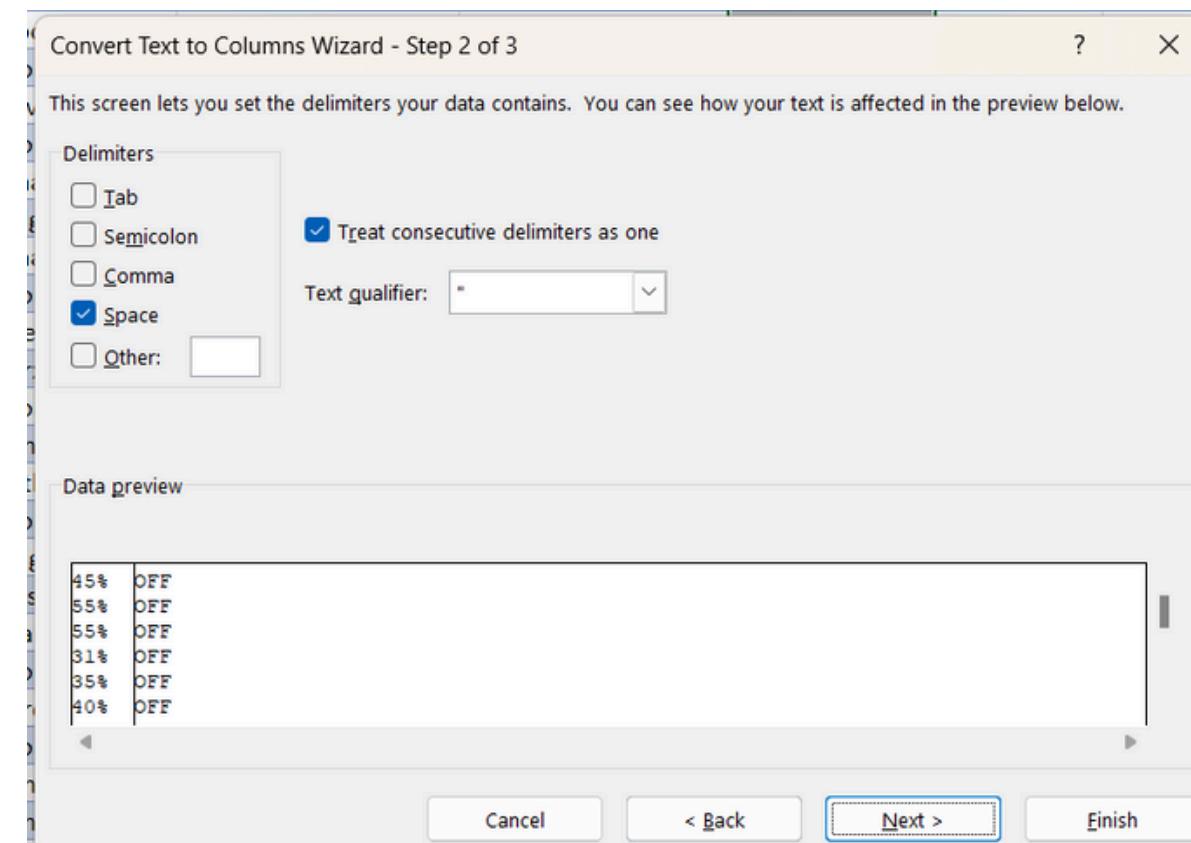
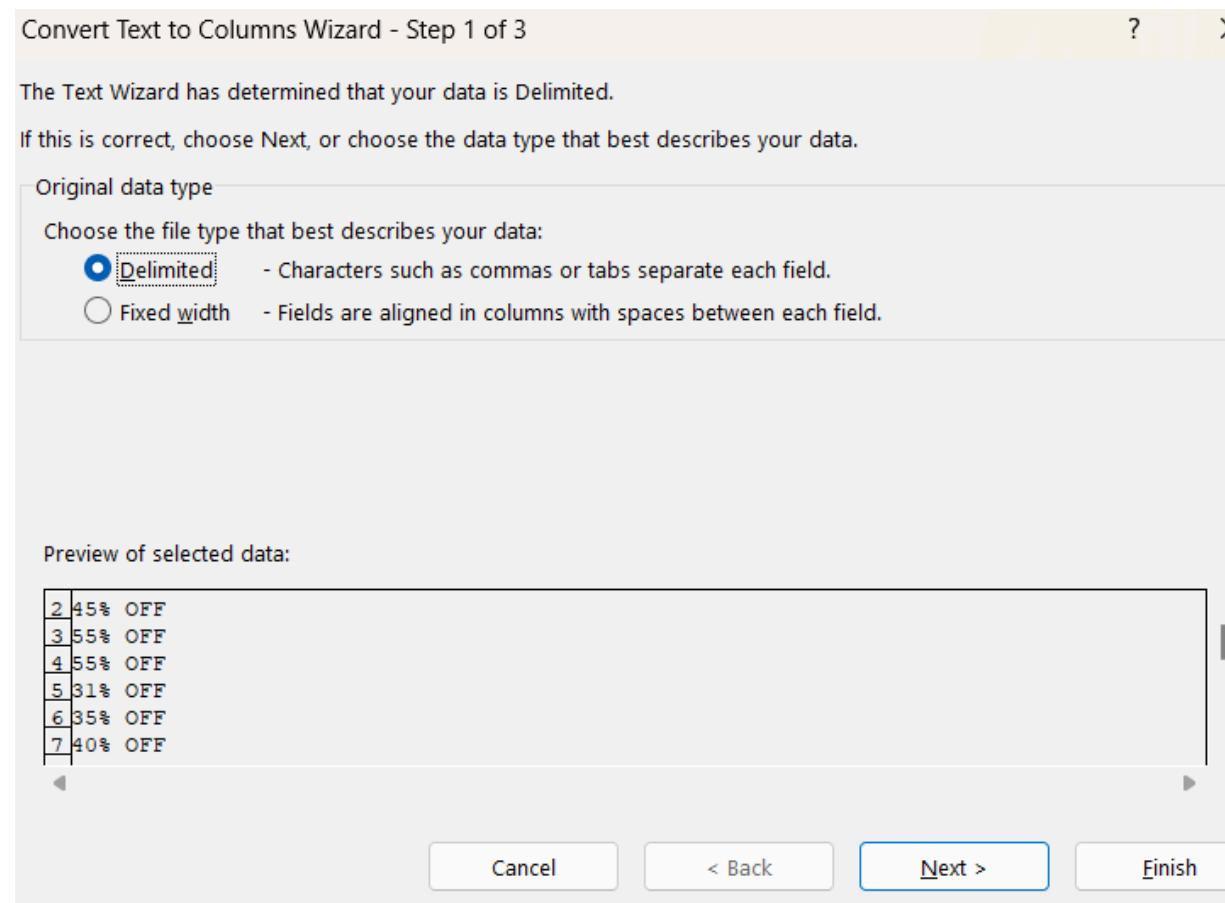
Select the dataset, go to Data > Remove Duplicates, and choose all columns to check for duplicates.

The screenshot shows a Microsoft Excel spreadsheet titled "Myntra Fasion Clothing.xlsx". The data consists of approximately 100 rows of clothing items with columns for URL, Product_id, BrandName, Category, Individual_c, categ, Description, DiscountPrice, Origin, DiscountOffer, SizeOption, Ratings, and Reviews. A 'Remove Duplicates' dialog box is overlaid on the spreadsheet, with the 'Select All' and 'My data has headers' options selected. The 'Columns' section of the dialog box has checkboxes for URL, Product_id, BrandName, Category, Individual_c, categ, and Description. A large blue arrow points from the bottom right towards the dialog box.

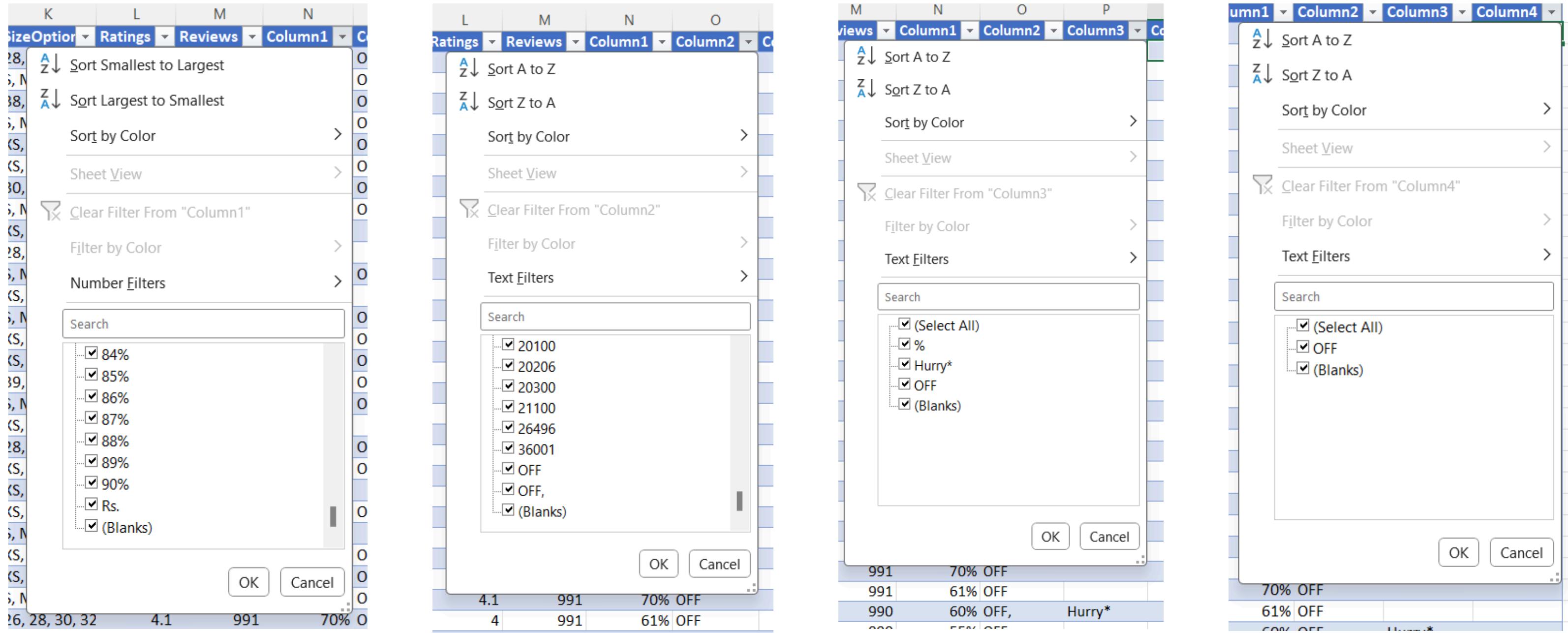
The screenshot shows the same Microsoft Excel spreadsheet as the previous one, but the 'Remove Duplicates' dialog box is now closed. A message box in the foreground says "No duplicate values found." with an "OK" button. The data table below it remains the same, showing the list of clothing items.

QA.2. :- Standardize "DiscountOffer" column to uniform format.

To standardize the column we first separate our column by using convert “Text to columns” option from Data.



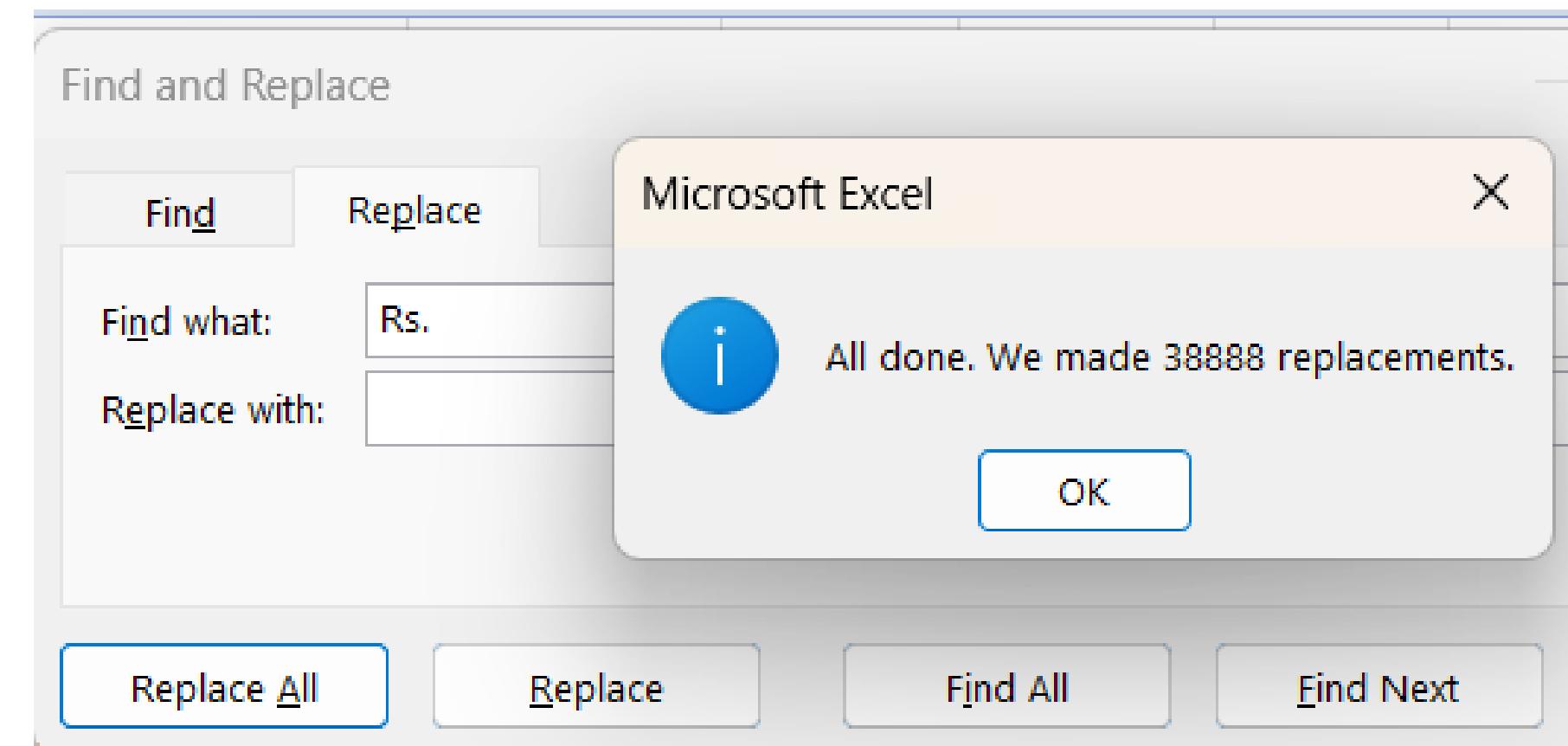
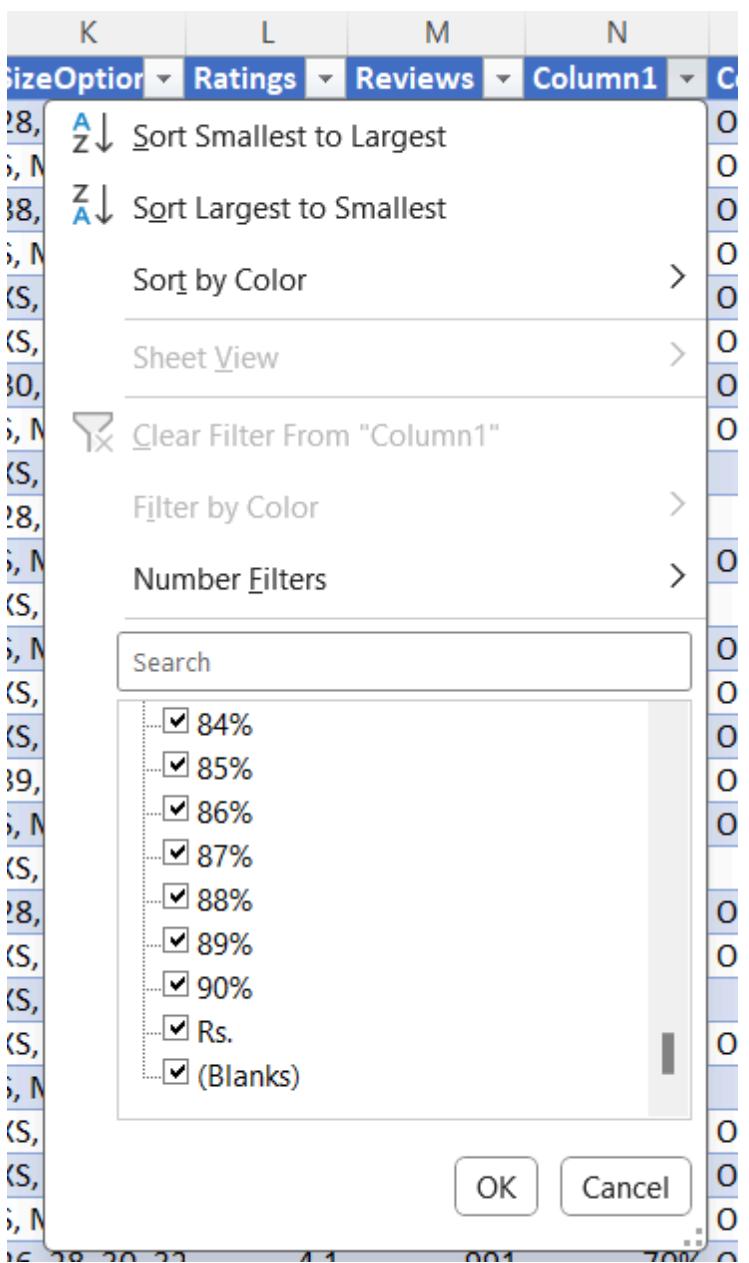
Standardize "DiscountOffer" column to uniform format.



By separating our column using space delimiter character from Text to Column. We have four additional column as a result of it. And as we can see in the last two columns the data is not useful to us so we can remove those columns.

Standardize "DiscountOffer" column to uniform format.

We have “Rs.” values in the 1st column we will replace them all by a space so that our column1 has the same kind of values as %.



Find and Replace All “Rs.” into
“Space Value”.

Standardize "DiscountOffer" column to uniform format.

Now we left with two columns one with % values and other one with all the amounts.

Number Filters

Search

- 83%
- 84%
- 85%
- 86%
- 87%
- 88%
- 89%
- 90%
- (Blanks)

OK

Column1

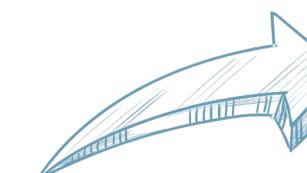
Text Filters

Search

- 20100
- 20206
- 20300
- 21100
- 26496
- 36001
- OFF
- OFF,
- (Blanks)

OK

Column2



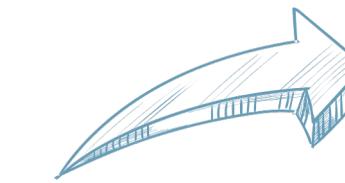
Where ever we have “OFF” in the column2 we have percentage value in column1. And all amounts in the column2 with the OFF. To standardize all these we are going to use a formula.

To standardize the "DiscountOffer" column to a single format, ensuring all values are uniform we are going to use this formula.

H	I	J	K	L	M	N	O	P	Q	R	S	T
untPr	OriginalPrice (in Rs)	DiscountOffer	SizeOption	Ratings	Reviews	Column1	Column2	Column3				
824	1499	45% OFF	28, 30, 32, 34	3.9	999	45%	OFF	=IF(ISNUMBER([@Column2]),[@Column2],IF(
517	1149	55% OFF	S, M, L, XL	4	999	55%	OFF	ISBLANK([@Column1]," ",[@Column1]*				
629	1399	55% OFF	38, 40, 42, 44	4.3	999	55%	OFF	[@Column1]*[@[OriginalPrice (in Rs)]])				
893	1295	31% OFF	S, M, L, XL, XX	4.2	999	31%	OFF					
	599	35% OFF	XS, S, M, L, XL	4.2	999	35%	OFF	209.65				
	599	40% OFF	XS, S, M, L, XL	4.4	999	40%	OFF	239.6				
599	1499	60% OFF	30, 32, 34, 36	3.9	998	60%	OFF	899.4				
	1395	58% OFF	S, M, L, XL	3.7	998	58%	OFF	809.1				
	1098		XS, S, M, L, XL	4.3	997							
	2749		28, 30, 32, 34	3.5	996							
1214	2699	55% OFF	S, M, L, XL, XX	4.4	996	55%	OFF	1484.45				
	699		XS, S, M, L, XL	4.1	996							
1019	3399	70% OFF	S, M, L, XL, XX	4.2	996	70%	OFF	2379.3				
	2499	50% OFF	XS, S, M, L, XL	4.3	996	50%	OFF	1249.5				
	799	60% OFF	XS, S, M, L, XL	4	996	60%	OFF	479.4				
516	1099	53% OFF	39, 40, 42, 44	4.2	995	53%	OFF	582.47				
696	1699	59% OFF	S, M, L, XL, XX	4.2	995	59%	OFF	1002.41				
	3999		XS, S, M, L, XL	4.3	995							
	2149	61% OFF	28, 30, 32, 34	4.1	995	61%	OFF	1310.89				
	000	75% OFF	XS, S, M, L, XL	4.2	995	75%	OFF	740.25				

QA.3. :- Identify rows where both "DiscountPrice" and "DiscountOffer" are null.

L	M	N	O	P	Q	R	S
Reviews	Discount	Column	ActualF				
3.9	999	674.55	=IF(ISBLANK([@DiscountPrice (in Rs)]),[@[OriginalPrice (in Rs)]]-[@DiscountAmount],[@DiscountPrice (in Rs)])				
4	999	631.95	[@DiscountAmount],[@DiscountPrice (in Rs)])				
4.3	999	769.45					
4.2	999	401.45	893	893			
4.2	999	209.65	389.35	389.35			
4.4	999	239.6	359.4	359.4			
3.9	998	899.4	599	599			
3.7	998	809.1	585.9	585.9			
4.3	997	#VALUE!					
3.5	996	#VALUE!					
4.4	996	1484.45	1214	1214			
4.1	996	#VALUE!					
4.2	996	2379.3	1019	1019			
4.3	996	1249.5	1249.5	1249.5			
4	996	479.4	319.6	319.6			
4.2	995	582.47	516	516			
4.2	995	1002.41	696	696			
4.3	995	#VALUE!					
4.1	995	1310.89	838.11	838.11			
4.2	995	749.25	249.75	249.75			



M	N	O	P	Q	R	S
Discount	Column	ActualF				
99	674.55	824	=IF(ISERROR([@Column2]),"",[@Column2])			
99	631.95	517	517			
99	769.45	629	629			
99	401.45	893	893			
99	209.65	389.35	389.35			
99	239.6	359.4	359.4			
98	899.4	599	599			
98	809.1	585.9	585.9			
97		#VALUE!				
96		#VALUE!				
96	1484.45	1214	1214			
96		#VALUE!				
96	2379.3	1019	1019			
96	1249.5	1249.5	1249.5			
96	479.4	319.6	319.6			
95	582.47	516	516			
95	1002.41	696	696			
95		#VALUE!				
95	1310.89	838.11	838.11			
95	749.25	249.75	249.75			
94		#VALUE!				

As we have standardized the “DiscountOffer” Column now we can get the most of the values of “DiscountPrice” column by subtracting “DiscountOffer” from “OriginalPrice” and for those cells that are blank we use the iserror formula.

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

K	L	M	N	O	P	Q
Reviews	Discount_amount	ActualPrice	Column			
3.9	999	674.55	824	=IF([@Discount_amount]","",AVERAGEIFS(
4	999	631.95	517	[Discount_amount],[Category],[@Category]),		
4.3	999	769.45	629	[@Discount_amount]))		
4.2	999	401.45	893	401.45		
4.2	999	209.65	389.35	209.65		
4.4	999	239.6	359.4	239.6		
3.9	998	899.4	599	899.4		
3.7	998	809.1	585.9	809.1		
4.3	997			970.1588		
3.5	996			1049.203		
4.4	996	1484.45	1214	1484.45		
4.1	996			925.7254		
4.2	996	2379.3	1019	2379.3		
4.3	996	1249.5	1249.5	1249.5		
4	996	479.4	319.6	479.4		
4.2	995	582.47	516	582.47		
4.2	995	1002.41	696	1002.41		
4.3	995			1281.074		
4.1	995	1310.89	838.11	1310.89		
4.3	995	749.25	249.75	749.25		
4.4	994			970.1588		

To fill up the “DiscountAmount” and “DiscountPrice” with the average discount amount and price with the respective category. We use this formula.

=IF(@[Discount_amount]","",AVERAGEIFS([Discount_amount],[Category],[@Category]),
[@[Discount_amount]])

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

Description	OriginalPrice (in Rs)	SizeOption	Ratings
roadster			3.9
locomot			4
roadste			4.3
zivame			4.2
roadste			4.2
mast h			4.4
highlan			3.9
mayra p			3.7
roadste			4.3
hereno			3.5
hrx by h			4.4
roadste			4.1
anubhu			4.2
athena			4.3
roadste			4
highlan			4.2
vishudh			4.2
sangria			4.3
tokyo t			4.1
dressbe			4.3
roadste			4.4
anouk v			3.8
enamor			4.2
all abou			4.2
kassually blue t	1299	XS, S, M, L, XL,	4.3
	1299	S, M, L, XL	4.3

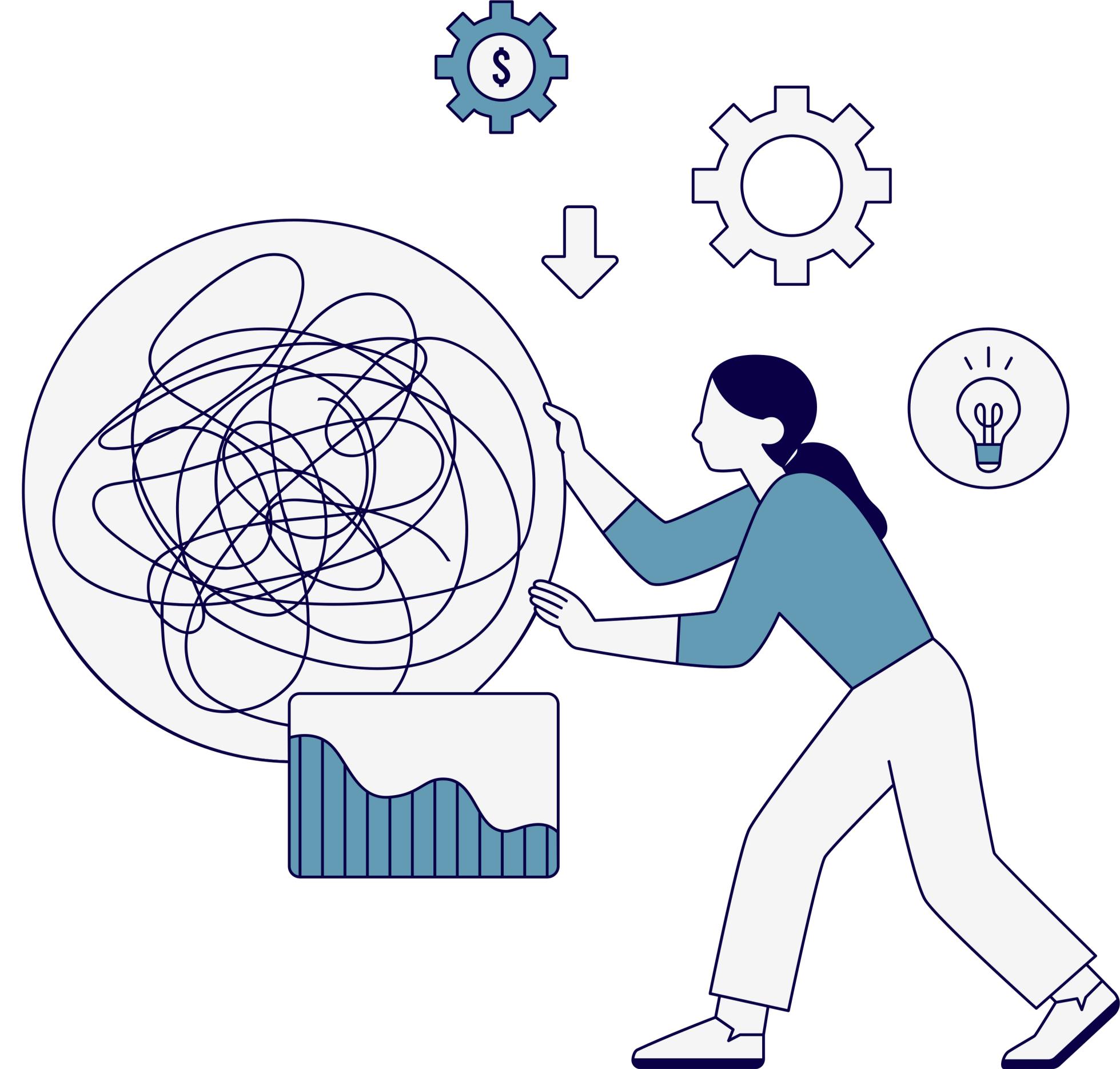
=COUNTA(Table1[SizeOption])
AH
AI

SizeOption
526564

SizeOp	Ratings
vy blue	28, 30, 32
black w	S, M, L, XL
vy whit	38, 40, 42
black sa	S, M, L, XL
Microsoft Excel	
OK	Cancel
No cells were found.	

We can confirm with above three methods that there are no null values are present in the column "Size Option".

DATA ANALYSIS



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

				=AVERAGEIF(I:I,>4",M:M)
				AI
				Average of original price whose rating is more than 4.
				2452.784698

Calculate the average original price for ratings above 4 using the formula
=AVERAGEIF(I:I,>4",M:M)

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

We will utilize the COUNTIFS function to identify products with discount offers exceeding 50% using this formula:

=COUNTIF(Table1[Discount Percentage], ">50%")

The screenshot shows the Microsoft Excel interface. In the formula bar at the top, the formula =COUNTIF(Table1[Discount Percentage], ">50%") is entered. Below the formula bar, the status bar displays the text "Count the number of products with a discount offer greater than 50% OFF." To the right of the status bar, the value 268039 is shown. At the bottom of the screenshot, the formula =COUNTIF(Table1[Discount Percentage], ">50%") is repeated, enclosed in a red rectangular box.

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

Here we are using the COUNTIF function to count the number of products with
the condition where the size is "M".

=COUNTIF(Table1[SizeOption],"M")

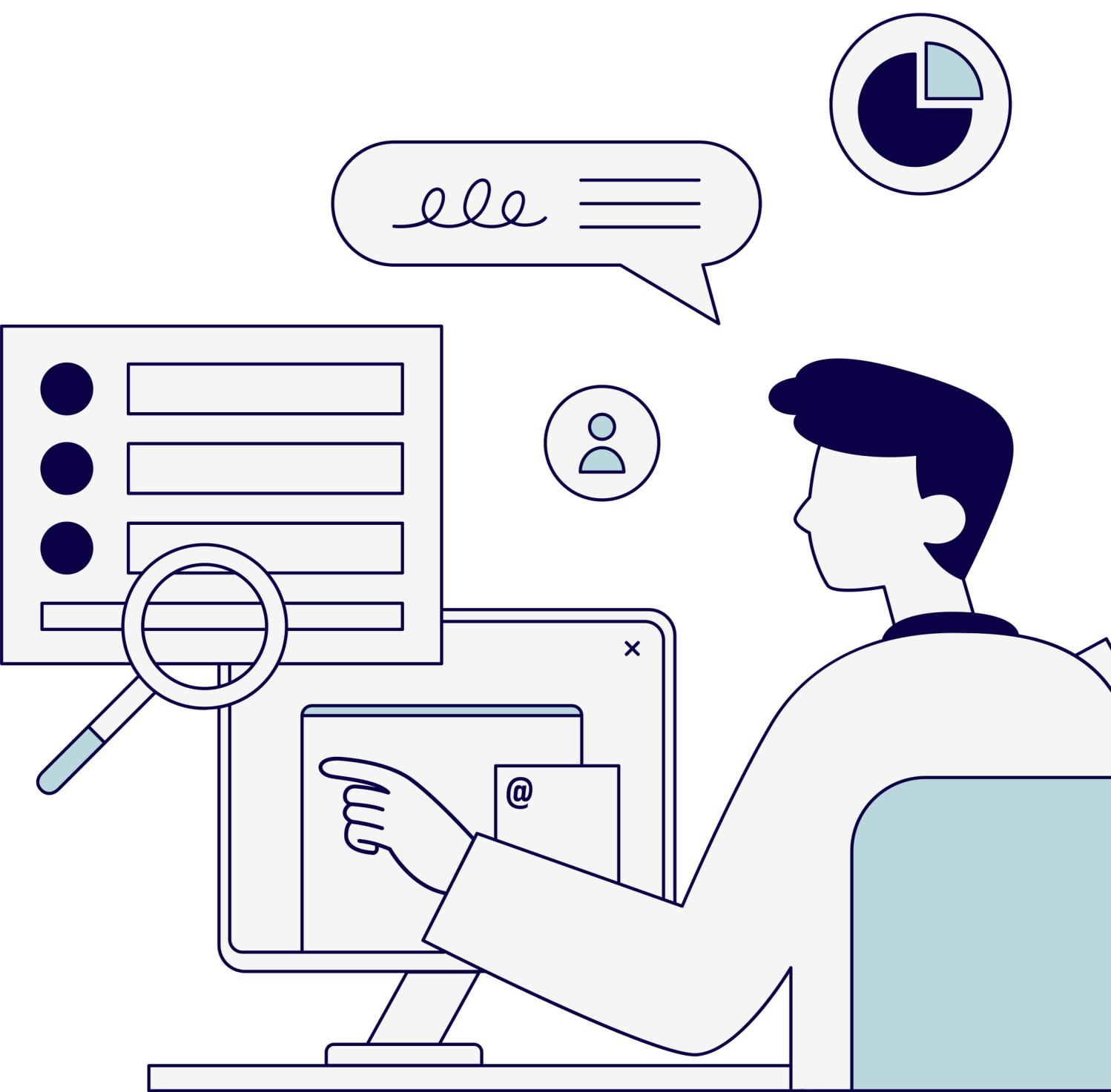
fx	=COUNTIF(Table1[SizeOption],"M")
	S
	Count the number of products available in size "M."
	656
	=COUNTIF(Table1[SizeOption],"M")

QB.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."

N	O	P
Discount Percent	High Discount	
45%	=IF([@Discount Percentage]>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	
88%	High Discount	
38%	Low Discount	
55%	High Discount	
132%	High Discount	
70%	High Discount	
50%	Low Discount	
60%	High Discount	
53%	High Discount	
59%	High Discount	
32%	Low Discount	
61%	High Discount	
75%	High Discount	

Create a column labeling offers as "High Discount" for over 50% and "Low Discount" for under 50% using the "If" function.

DATA RETRIEVAL AND LOOKUP



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

W	X	Y	Z	A
Lookup value				
11226634				
<pre>=VLOOKUP(W11,Table1[[#All],[Product_id]:[High Discount]],[2,7,9],FALSE)</pre>				
Brand	Price	Rating		
Maniac	1199	3.9		

VLOOKUP to find the product brand, price, and rating of the product with Product_id "11226634"

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

	=XLOOKUP(W11,Table1[Product_id],Table1[BrandName],,0)		
W	X	Y	
Lookup value			
11226634			
	=XLOOKUP(W11,Table1[Product_id],Table1[BrandName],,0)		
Brand	Price	Rating	
Maniac	1199	3.9	

XLOOKUP to find the product brand, price, and rating of the product with Product_id "11226634"

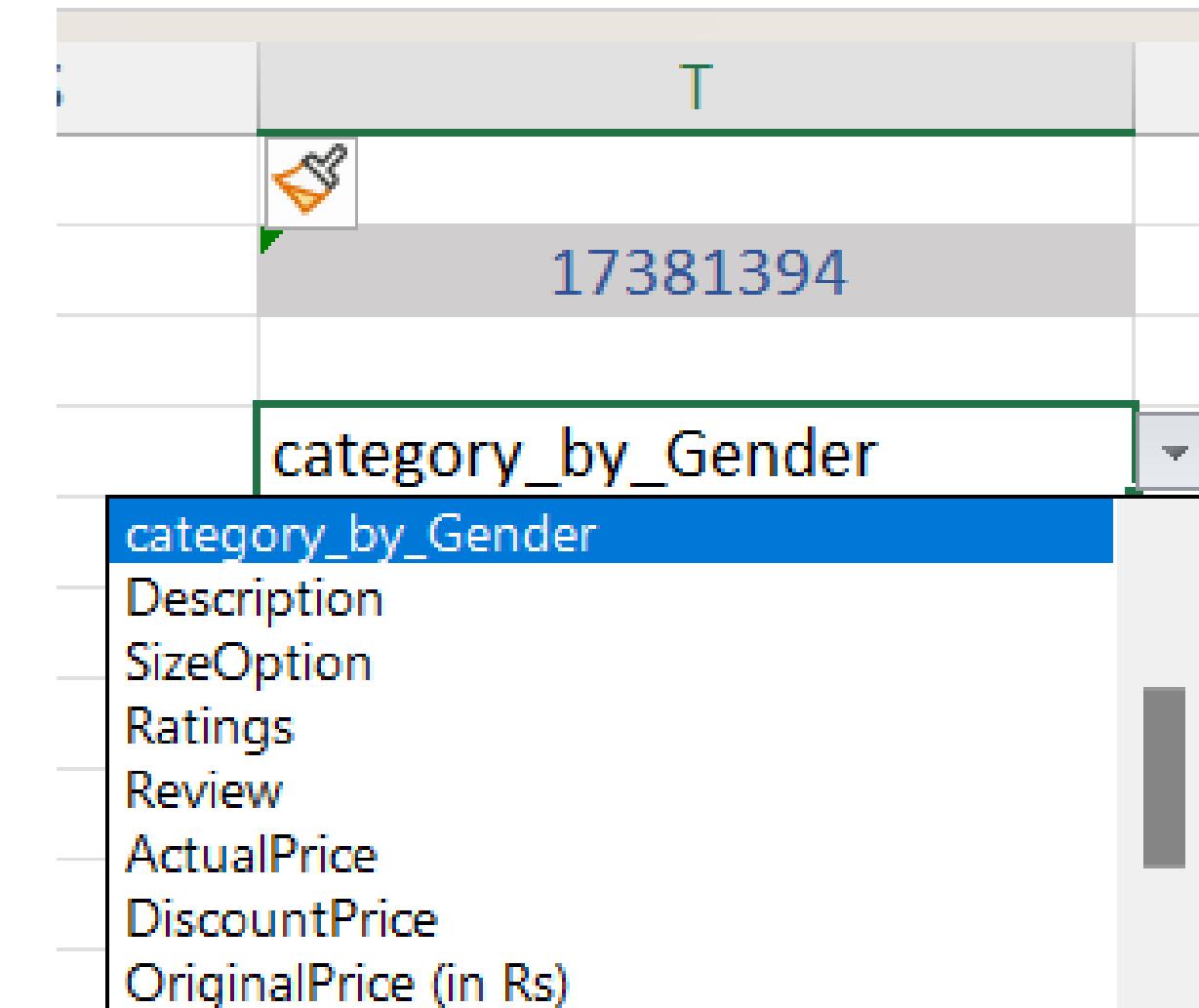
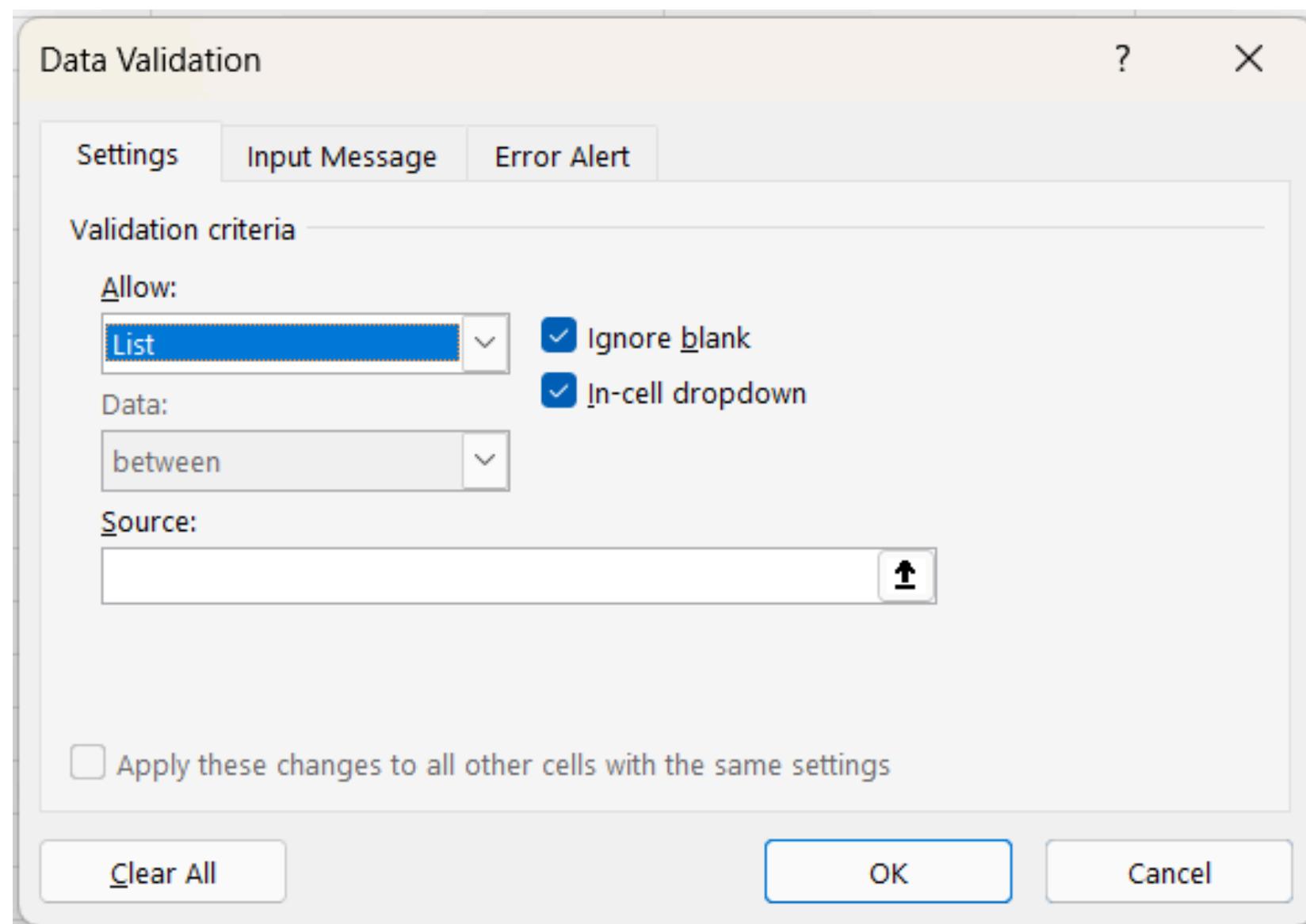
QC.2. :- Find the "DiscountPrice" for the product with the Product ID "6744434" using the INDEX and MATCH functions.

AB	AC	AD	AE	AF	AG
			=INDEX(Table1[DiscountPrice],MATCH(AC5,Table1[Product_id],0))		

Product ID	Discount Price
6744434	899.4

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

DATA>DATA VALIDATION> SELECT LIST>FILL SOURCE>CLICK OK



The product list has been curated utilizing data validation techniques.

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

	=XLOOKUP(T15,Table1[Product_id],XLOOKUP(T17,Table1[[#Headers],[URL]:[High Discount]],Table1[[URL]:[High Discount]]))		
T	17381394	U	V
category_by_Gender			W
	Women		

This is the result we get once we utilize the nested xlookup to find any column's detail of a product with it's product id.

For Product ID "17381394" column "category_by_Gender" returns the value as "Women".



Thank you

