

# Analysis of Myntra Apparel

Presented by : Kinjan Chauhan



# PROJECT

**Problem Statement:** You are working at Myntra, a leading online fashion retailer. The management has asked you to analyze a dataset of various apparel items to gain insights into pricing, discounts, ratings, and available sizes.

## PROJECT QUESTIONS BASED ON

DATA CLEANING AND  
PREPARATION



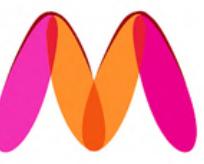
DATA  
ANALYSIS



DATA RETRIEVAL  
AND LOOKUP



# Data Cleaning and Preparation



# 1. Check for duplicate values in your dataset and remove them.

- STEP 1 : Select the table → Go → Data tools → Remove Duplicates
- RESULT : No duplicates find

Screenshot of Microsoft Excel showing the 'Data' tab selected. A 'Remove Duplicates' dialog box is open over a table of product data. The dialog box has 'My data has headers' checked and lists columns: Description, DiscountPrice (in Rs), OriginalPrice (in Rs), DiscountOffer, and SizeOption. An arrow points from the 'Data Tools' section of the ribbon to this dialog box.

Category	Individual_category	category_by_Gender	Description	DiscountPrice (in Rs)	OriginalPrice (in Rs)	DiscountOffer
Bottom Wear	jeans	Men	roadster men	516	1099	53% OFF
Bottom Wear	track-pants	Men	locomotive men	516	1099	53% OFF
Topwear	shirts	Men	roadster men	516	1099	53% OFF
Lingerie & Sleep Wear	shapewear	Women	zivame women	516	1099	53% OFF
Western	tshirts	Women	roadster women	516	1099	53% OFF
Western	tops	Women	mast harbour	516	1099	53% OFF
Bottom Wear	trousers	Men	highlander men	516	1099	53% OFF
Western	tops	Women	mayra pink	516	1099	53% OFF
Western	tshirts	Women	roadster women	516	1099	53% OFF
Bottom Wear	jeans	Men	herenow men	516	1099	53% OFF
Sports Wear	tights	Men	hrx by hrithik	516	1099	53% OFF
Topwear	tshirts	Men	roadster men	516	1099	53% OFF
Indian Wear	kurta-sets	Women	anubhutee women	516	1099	53% OFF
Western	jumpsuit	Women	athena women	516	1099	53% OFF
Western	tshirts	Women	roadster women	516	1099	53% OFF
Topwear	shirts	Men	highlander men	516	1099	53% OFF
Indian Wear	kurtas	Women	vishudh women	696	1699	59% OFF
Plus Size	kurta-sets	Women	sangria women green off white printed kurta sets	696	1699	59% OFF
Western	trousers	Women	tokyo talkies women olive green regular fit	696	1699	59% OFF
Western	tshirts	Women	dressberry women lovely olive typography	696	1699	59% OFF
Western	tshirts	Women	roadster women pack of 2 printed cotton	696	1699	59% OFF
Indian Wear	kurtas	Women	anouk women	719	1599	55% OFF
Lingerie & Sleep Wear	bra	Women	enamor black non wired removable pads	849	1699	50% OFF
Western	shirts	Women	all about you women mauve stretchable c	1499	2999	50% OFF
Western	tshirts	Women	lazypants blue floral printed t shirt	1499	2999	50% OFF

Screenshot of Microsoft Excel showing the 'Table Design' ribbon tab selected. A message box in the center says 'No duplicate values found.' with an 'OK' button. The background shows a table of product data with columns: Individual\_category, category\_by\_Gender, Description, DiscountPrice (in Rs), OriginalPrice (in Rs), and DiscountOffer. A large purple arrow points from the 'Data Tools' section of the ribbon towards the message box.

C	D	E	F	G	H	I
Individual_category	category_by_Gender	Description	DiscountPrice (in Rs)	OriginalPrice (in Rs)	DiscountOffer	
Near	jeans	Men	roadster men	824	1499	45% OFF
Near	track-pants	Men	locomotive men	517	1149	55% OFF
Near	shirts	Men	roadster men	629	1399	55% OFF
Lingerie & Sleep Wear	shapewear	Women	zivame women	893	1295	31% OFF
Near	tshirts	Women	roadster women white solid v neck pure c	599	999	35% OFF
Near	tops	Women	mast harbour women yellow solid tank to	599	999	40% OFF
Near	trousers	Men	highlander men	599	1499	60% OFF
Near	tops	Women	microsoft excel	599	1395	58% OFF
Near	tshirts	Women	re cotto	1098	1999	45% OFF
Near	jeans	Men	clean	2749	4299	33% OFF
Near	tights	Men	1214	2699	55% OFF	
Near	tshirts	Men	ed pol	699	1299	45% OFF
Near	kurta-sets	Women	anubhutee won	1019	3399	70% OFF
Near	jumpsuit	Women	athena women black solid basic jumpsuit	2499	4999	50% OFF
Near	tshirts	Women	roadster women maroon solid round nec	799	1699	60% OFF
Near	shirts	Men	highlander men	516	1099	53% OFF
Near	kurtas	Women	vishudh women	696	1699	59% OFF
Near	kurtas	Women	sangria women green off white printed ku	3999	7999	50% OFF
Near	trousers	Women	tokyo talkies women olive green regular fi	2149	4299	61% OFF
Near	tshirts	Women	dressberry women lovely olive typography	999	1999	75% OFF



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

- STEP 1 : Cntrl+f → Replace Rs. by blank
- STEP 2 : Data → Data tools → Text to column → Delimited → space

	finalPrice	DiscountOffer	SizeOption	Rating	Review	
	1499	45% OFF	28, 30, 32, 34, 36	3.9	999	
	1149	55% OFF	S, M, L			
	1399	55% OFF	8, 40, 42			
	1295	31% OFF	S, M, L			
	599	35% OFF	XS, S, M			
	599	40% OFF	XS, S, M			
	1499	60% OFF	30, 32			
	1395	58% OFF	S, M, L			
	1098		XS, S, M			
	2749		28, 30, 32, 34, 36	3.5	996	
	2699	55% OFF	S, M, L, XL, XXL	4.4	996	

Replace Rs. by blank

The screenshot shows two overlapping Excel dialogs. The top dialog is the 'Find and Replace' dialog, which has 'Replace' selected. It contains fields for 'Find what:' (set to 'Rs.') and 'Replace with:' (empty). Below these are buttons for 'Replace All', 'Replace', 'Find All', 'Find Next', and 'Close'. A message box from Microsoft Excel is overlaid on the left, stating 'All done. We made 230 replacements.' with an 'OK' button. The bottom dialog is the 'Convert Text to Columns Wizard - Step 2 of 3'. It shows a preview of the data in columns E through I. The 'Delimiters' section is open, with 'Space' checked and 'Treat consecutive delimiters as one' checked. A 'Text qualifier' dropdown is set to '\"'. The 'Data preview' section shows the 'DiscountOffer' column with values like '45%', '55%', etc., followed by 'OFF'. The 'Convert' button at the bottom right of the wizard is highlighted.

Delimited by space

- STEP 3 : Now delete the extra column which are not of use
- STEP 4 : By using 'if' function we standardised the " Discount offer "column to the new column " Discount offer (in Rs.) "

The screenshot shows a Microsoft Excel spreadsheet with data in columns H through O. The formula bar displays two formulas:

- `=ISBLANK([@Column2]), "",[@Column1]*[@[OriginalPrice (in Rs)]]])`
- `=IF(ISNUMBER([@Column2]),[@Column2],IF(ISBLANK([@Column2]), "",[@Column1]*[@[OriginalPrice (in Rs)]]))`

A red arrow points from the text-to-column step to the formula application step.

	H	I	J	K	L	M	N	O	P	Q	R	S
ice	DiscountOffer	SizeOption	Rating	Review	Column1	Column2						
99	45% OFF	28, 30, 32, 34, 36	3.9	999	45%	OFF						
49	55% OFF	S, M, L, XL	4	999	55%	OFF						
99	55% OFF	8, 40, 42, 44, 46, 4	4.3	999	55%	OFF						
95	31% OFF	S, M, L, XL, XXL	4.2	999	31%	OFF						
99	35% OFF	XS, S, M, L, XL	4.2	999	35%	OFF						
99	40% OFF	XS, S, M, L, XL	4.4	999	40%	OFF						

Through Text to column --column  
got separated

Through Text to column --column  
got separated



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

- STEP 1 : First we fill the blank value of "Discount Price" by subtracting "Discount offer (in Rs)" from "Original price" in new column by formula

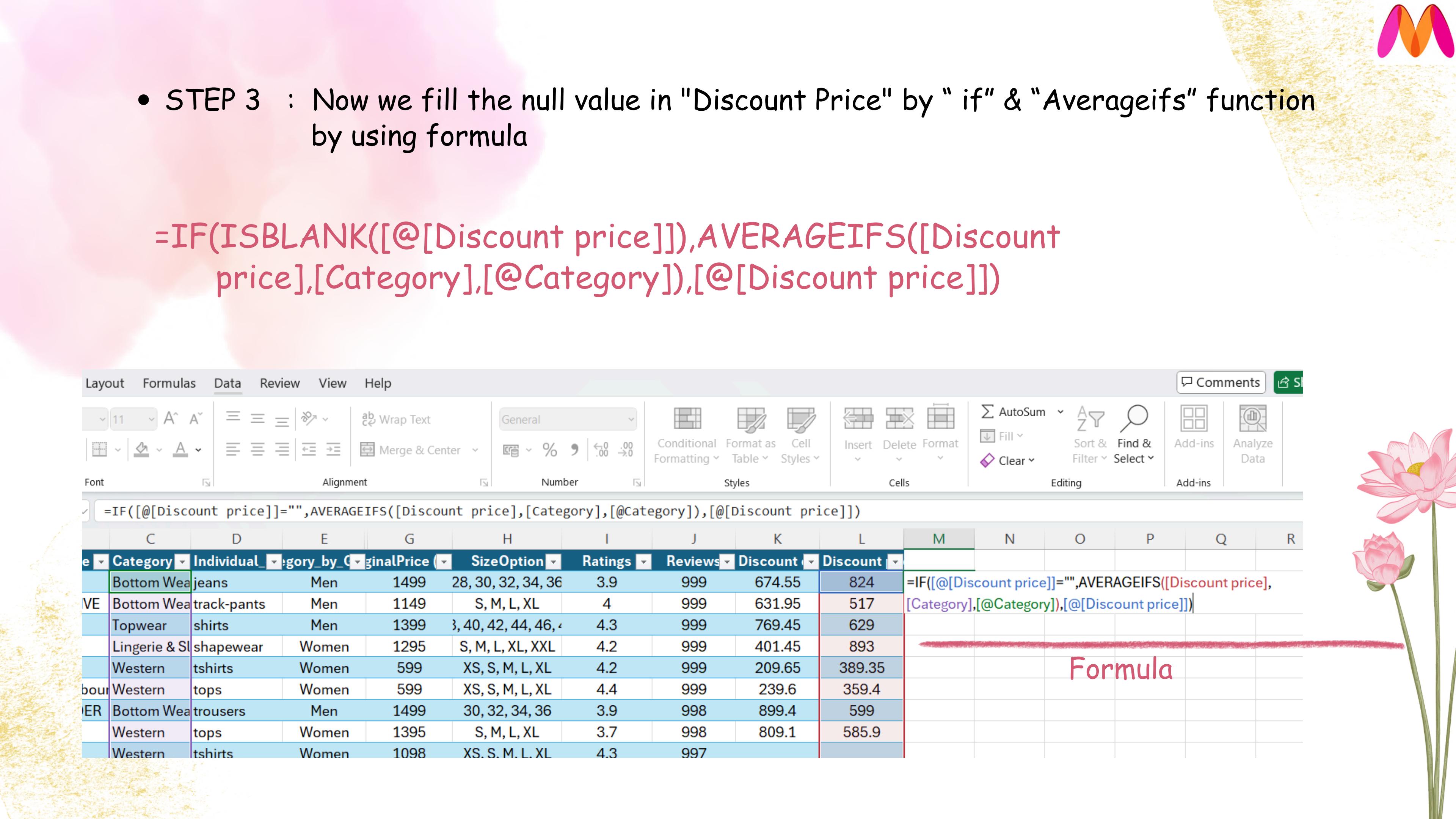
```
=IF(ISBLANK([@DiscountPrice (in Rs)]),[@OriginalPrice (in Rs)]-[@[Discount price (in Rs.)]],[@DiscountPrice (in Rs)])
```

- STEP 2 : After this we found an error in blank cell so we use "if " function with "iserror" function to fix the error

```
=IF(ISERROR([@Discount price]),"",[@Discount price])
```

- STEP 3 : Now we fill the null value in "Discount Price" by " if" & "Averageifs" function by using formula

=IF(ISBLANK([@Discount price]),AVERAGEIFS([Discount price],[Category],[@Category]),[@Discount price])



The screenshot shows a Microsoft Excel spreadsheet with a decorative background featuring a pink lotus flower and a yellow flower.

**Formula Bar:**

```
=IF([@Discount price]="",AVERAGEIFS([Discount price],[Category],[@Category]),[@Discount price])
```

**Excel ribbon:**

- Layout
- Formulas
- Data
- Review
- View
- Help
- Comments
- Find & Select
- Add-ins
- Analyze Data

**Table Structure:**

	C	D	E	G	H	I	J	K	L	M	N	O	P	Q	R
e	Category	Individual	Category_by_C	OriginalPrice	SizeOption	Ratings	Reviews	Discount	Discount						
VE	Bottom Wear	jeans	Men	1499	28, 30, 32, 34, 36	3.9	999	674.55	824	=IF([@Discount price]="",AVERAGEIFS([Discount price],[Category],[@Category]),[@Discount price])					
VE	Bottom Wear	track-pants	Men	1149	S, M, L, XL	4	999	631.95	517						
	Topwear	shirts	Men	1399	3, 40, 42, 44, 46, 4	4.3	999	769.45	629						
	Lingerie & Sl	shapewear	Women	1295	S, M, L, XL, XXL	4.2	999	401.45	893						
bou	Western	tshirts	Women	599	XS, S, M, L, XL	4.2	999	209.65	389.35						
bou	Western	tops	Women	599	XS, S, M, L, XL	4.4	999	239.6	359.4						
PER	Bottom Wear	trousers	Men	1499	30, 32, 34, 36	3.9	998	899.4	599						
PER	Western	tops	Women	1395	S, M, L, XL	3.7	998	809.1	585.9						
	Western	tshirts	Women	1098	XS, S, M, L, XL	4.3	997								

Formula

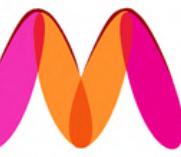
## 4. Replace all null values in the "SizeOption" column with the text "Not Available."

Category	Individual	Category_by_L	OriginalPrice	SizeOption	Ratings	Review	Dis
Bottom Wear:jeans					3.9	999	
Bottom Wear:track-pants					4	999	
Topwear shirts					4.3	999	
Lingerie & Sleepwear:shapewear					4.2	999	
Western tshirts					4.2	999	
Western tops					4.4	999	
Bottom Wear:trousers					3.9	998	
Western tops					3.7	998	
Western tshirts					4.3	997	
Bottom Wear:jeans					3.5	996	
Sports Wear:tights					4.4	996	
Topwear tshirts					4.1	996	
Indian Wear:kurta-set					4.2	996	
Western jumpsuit					4.3	996	
Western tshirts					4	996	
Topwear shirts					4.2	995	
Indian Wear:kurtas					4.2	995	
Plus Size kurta-set					4.3	995	
Western trousers					4.1	995	
Western tshirts					4.3	995	
Western tshirts					4.4	994	
Indian Wear:kurtas					3.8	993	
Lingerie & Sleepwear:bra					4.2	993	
Western shirts					4.2	993	
Western tops					4.3	992	
Western trousers					4.2	991	
Western shorts					4.1	991	
Plus Size kurta-set					4	991	
Western shorts	Women	1199	26, 28, 30, 32, 34, 36		4.3	990	
Western tshirts	Women	1598	L, M, S, XS, XL		4.2	990	
Western dresses	Women	2349	S, M, L, XL, XXL		4.4	990	

- STEP 1 : Select the column "Size option ".
- STEP 2 : click to downdrop.
- STEP 3 : Type "null " to the search box.
- STEP 4 : we find that there is no match found.
- so, there is no "null" value find in the ""Size option " column.



# Data Analysis



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

- STEP 1 : We will use "Averageif" function to find the average original price for products with ratings greater than 4.

OriginalPrice	SizeOption	Ratings	Reviews	Discount	Discount price2	
1499	28, 30, 32, 34, 36	3.9	999	674.55	824	
1149	S, M, L, XL	4	999	631.95	517	Calculate the overall average original price
1399	38, 40, 42, 44, 46, 48	4.3	999	769.45	629	for products with ratings greater than 4.
1295	S, M, L, XL, XXL	4.2	999	401.45	893	
599	XS, S, M, L, XL	4.2	999	209.65	389.35	=ROUND(AVERAGEIF(Table1[Ratings],">4",
599	XS, S, M, L, XL	4.4	999	239.6	359.4	Table1[OriginalPrice (in Rs)]),2)
1499	30, 32, 34, 36	3.9	998	899.4	599	

Formula

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

- STEP 1 : First we convert the "Discount offer (in Rs.)" to "Discount offer (in %)" in new column.

Screenshot of Microsoft Excel showing a table with columns: Category\_by\_C, OriginalPrice, SizeOption, Ratings, Reviews, Discount offer(in Rs.), Discount price, and Discount offer(%). The formula for the Discount offer(%) column is =IF([@Discount offer(in Rs.)]="", "",[@Discount offer(in Rs.)]/[@OriginalPrice (in Rs)])

Category_by_C	OriginalPrice	SizeOption	Ratings	Reviews	Discount offer(in Rs.)	Discount price	Discount offer(%)
Men	1499	28, 30, 32, 34, 36	3.9	999	674.55	824	=IF([@Discount offer(in Rs.)]:"", "",[@Discount offer(in Rs.)]/[@OriginalPrice (in Rs)])
Men	1149	S, M, L, XL	4	999	631.95	517	
Men	1399	38, 40, 42, 44, 46, 48	4.3	999	769.45	629	
Women	1295	S, M, L, XL, XXL	4.2	999	401.45	893	
Women	599	XS, S, M, L, XL	4.2	999	209.65	389.35	

Formula

- STEP 2 : Then we use "Countif" function to Count the number of products with a discount offer greater than 50% OFF.

Screenshot of Microsoft Excel showing the same table as above, but with more rows. The last row shows a formula: =COUNTIF(Table1[Discount offer(%)], ">50%")

OriginalPrice	SizeOption	Ratings	Reviews	Discount offer(in Rs.)	Discount price	Discount offer(%)
1499	28, 30, 32, 34, 36	3.9	999	674.55	824	45%
1149	S, M, L, XL	4	999	631.95	517	55%
1399	38, 40, 42, 44, 46, 48	4.3	999	769.45	629	55%
1295	S, M, L, XL, XXL	4.2	999	401.45	893	31%
599	XS, S, M, L, XL	4.2	999	209.65	389.35	35%
599	XS, S, M, L, XL	4.4	999	239.6	359.4	40%
1499	30, 32, 34, 36	3.9	998	899.4	599	60%
1395	S, M, L, XL	3.7	998	809.1	585.9	58%
1098	XS, S, M, L, XL	4.3	997	739.615796		
2749	28, 30, 32, 34, 36	3.5	996	908.8077558		
2699	S, M, L, XL, XXL	4.4	996	1484.45	1214	55%
699	, S, M, L, XL, XXL, 3XL, 4	4.1	996	698.2084967		

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

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

- STEP 1 : First click the " Size option " → filters → Text filters → Contains → "M"
- STEP 2 : After apply filter the " Size option " column contains the product available in size "M" and to count the number see the downward bar which shows the count

E	F	G	H	I	J	K
Category_by_	Description	OriginalPrice	SizeOption	Ratings	Review	Discount of
Men	A	1149	4	999	631.95	
Women	Z	1295	4.2	999	401.45	
Women	A	599	4.2	999	209.65	
Women	Z	599	4.4	999	239.6	
Women	A	1395	3.7	998	809.1	
Women	Z	1098	4.3	997		
Men	A	2699	4.4	996	1484.45	
Men	Z	699	4.1	996		
Women	A	3399	4.2	996	2379.3	
Women	Z	2499	4.3	996	1249.5	
Women	A	799	4	996	479.4	
Women	Z	1699			1002.41	
Women	A	3999	Equals...		49.25	
Women	Z	3999	Does Not Equal...		79.45	
Women	A	1298	Begins With...		749.5	
Women	Z	1298	Ends With...		749.5	
Women	A	1599	Contains...		05.38	
Women	Z	1599	Custom...		40.43	
Women	A	849	Contains...		539.4	
Women	Z	849	Does Not Contain...		799	
Women	A	1499	Custom Filter...		526.85	
Women	Z	1499	Custom Filter...		591.85	
Women	A	1299	4.3	990	1627.45	
Women	Z	1299	4.1	990	539.4	
Women	A	1299	4.3	989	587.51	
Women	Z	1299	4	988	1139.4	
Men	A	2959	4	988	1104.35	

OK Cancel

F	G	H	I	J	K	L	M	N
Description	OriginalPrice	SizeOption	Ratings	Review	Discount of	Discount	Discount offe	Column1
comotive nr	1149	S, M, L, XL	4	999	631.95	517	55%	High Discount
name women	1295	S, M, L, XL, XXL	4.2	999	401.45	893	31%	Low Discount
adster wom	599	XS, S, M, L, XL	4.2	999	209.65	389.35	35%	Low Discount
ast harbou	599	XS, S, M, L, XL	4.4	999	239.6	359.4	40%	Low Discount
ayra pink ei	1395	S, M, L, XL	3.7	998	809.1	585.9	58%	High Discount
adster wom	1098	XS, S, M, L, XL	4.3	997		739.6158		High Discount
x by hrithik	2699	S, M, L, XL, XXL	4.4	996	1484.45	1214	55%	High Discount
adster men	699							High Discount
ubhutee wi	3399							High Discount
hena wome	2499							Low Discount
adster wom	799							High Discount
shudh wom	1699							High Discount
ngria wom	3999							High Discount
essberry wo	999							High Discount
adster wom	1298							High Discount
ouk wome	1599							High Discount
amor black	849							High Discount
about you	1499							Low Discount
ssually blu	1299							High Discount
re women r	1299							High Discount
ngria wom	899							High Discount
adster wom	1598	L, M, S, XS, XL	4.2	990	799	799	50%	Low Discount
shudh wom	2349	S, M, L, XL, XXL	4.4	990	1526.85	822.15	65%	High Discount
ttingfab oliv	2449	S, M, L, XL, XXL	4.3	990	1591.85	857.15	65%	High Discount
iero wome	2959	XS, S, M, L, XL	4.3	990	1627.45	1331.55	55%	High Discount

Custom Autofilter ? X

Show rows where:

SizeOption

contains m

And Or

Use ? to represent any single character  
Use \* to represent any series of characters

OK Cancel



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."

- STEP 1 : We will use "if" function for label "High Discount" & "Low Discount" .

L	M	N	O	P
Count	Discount offer	Column1		
517	55%	=IF([@Discount offer(%)]>50%, "High Discount", "Low Discount")		
893	31%			
89.35	35%	Low Discount		
359.4	40%	Low Discount		
585.9	58%	High Discount		
9.6158		High Discount		

Formula



# Data Retrieval and Lookup



# 1. Use VLOOKUP/XLOOKUP to find the product brand, price, and rating of the product with Product\_id "11226634".

- We will use "vlookup" function to find the product brand, price, and rating of the product with Product\_id "11226634" by using formula

=VLOOKUP(P7,Table1,{2,7,9},FALSE) → Formula

- Result

	11226634		
Brand	Original price	Rating	
Maniac	1199	3.9	

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

6744434
=INDEX((Table1[Discount price]),MATCH(P14,Table1[Product_id],0))
um, [column_num])
ow_num, [column_num], [area_num]

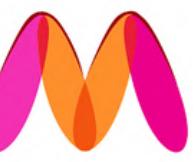
=INDEX((Table1[Discount  
price]),MATCH(P14,Table1[Product\_id],0))

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

- STEP 1 : Select any product id from product id column for which we want to find product detail.
- STEP 2 : Go to → data → data tools → data validation → select " list" in allow box → ok
- STEP 3 : Select all headers of the column in table as a source value. It will create a dropdown of list of headers in the column ex. Brand , Category etc.

The screenshot shows the Microsoft Excel ribbon at the top with various tabs like Help, Connections, Data Tools, Forecast, and Outline. Below the ribbon is a table with columns G through N. Cell P1 contains the value '11895958'. The 'Data Tools' tab is selected, and a 'Data Validation' dialog box is open over the table. In the dialog box, under 'Validation criteria', the 'Allow' dropdown is set to 'List'. Under 'Source', the formula '=A\$1:\$N\$1' is entered. The 'Data' dropdown is set to 'between'. The 'OK' button is highlighted. A purple arrow points from the 'Source' dropdown in the dialog box towards the list of column headers on the right.

Product_id	Category
11895958	Category
	Product_id
	BrandName
	Category
	Individual_category
	category_by_Gender
	Description
	OriginalPrice (in Rs)
	SizeOption
	Ratings
	Reviews
	Discount offer(in Rs.)
	Discount price



- STEP 4 : Now we use the nested "xlookup" function to find the different details of "product id " from different columns.

```
=XLOOKUP(P4,Table1[Product_id],XLOOKUP(Q4,Table1[#Headers],Table1))
```

- Result :

Product_id
11895958
Category
Topwear

Product_id
11895958
SizeOption
38, 40, 42, 44, 46, 48

Product_id
11895958
Ratings
4.3

Thank you