

Myntra Case Study

Presented by: Shivani Awasthi





Order of Business

Data Cleaning and Preparation

- Check for duplicate values in your dataset and remove them.
- Standardize the "DiscountOffer" column to a single format, ensuring all values are uniform.
- Identify rows where both "DiscountPrice" and "DiscountOffer" are null and fill the "DiscountPrice" with the average discount price of the respective category.
- Replace all null values in the "SizeOption" column with the text "Not Available."

Data Analysis

- Calculate the overall average original price for products with ratings greater than 4
- Count the number of products with a discount offer greater than 50% OFF.
- Count the number of products available in size "M."
- 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."

Data Retrieval

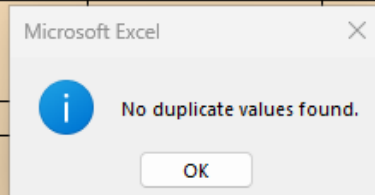
- Use VLOOKUP/XLOOKUP to find the product brand, price, and rating of the product with Product_id "11226634".
- Find the "DiscountPrice" for the product with the Product ID "6744434" using the INDEX and MATCH functions.
- Utilize nested xlookup to find any column's detail of a product with its product id.

DATA CLEANING AND PREPARATION

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

Resolution: There are no duplicates found.

harbour	0		599	
nder me	599	599	1499	
pink em	0		1395	
ter wom			1098	
ow men			2749	
hrithik r			2699	
ter men	883		699	
utee wo	1019	1019	3399	
a womer	0		2499	
ter wom	0		799	
nder me	516	516	1099	
h wome	696	696	1699	
a womer	776		3999	
talkies w	0		2149	
erry wo	0		999	



DATA CLEANING AND PREPARATION

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

Resolution: I have followed the following steps to standardize the DiscountOffer to a single format i.e. in "Rs.":

- Step1: Cleaned the Values "OFF" and "Hurry*" using "**Find and Replace**"

	I	J	K	L
	DiscountPrice (in Rs)	OriginalPrice (in Rs)	Stanadard Discount Offer in Rs	DiscountOffer
824	824	1499	675	45%
517	517	1149	632	55%
629	629	1399	769	55%
893	893	1295	401	31%
0		599	210	35%
0		599	240	40%
5			899	60%
10			809	58%
5			0	
10			0	
12			1484	55%
8			0	
10			2379	70%
0		2499	1250	50%
0		2499	1250	50%

	I	J	K	L
	DiscountPrice (in Rs)	OriginalPrice (in Rs)	Stanadard Discount Offer in Rs	DiscountOffer
0		599	240	40%
5			899	60%
10			809	58%
5			0	
10			0	
12			1484	55%
8			0	
10			2379	70%
0		2499	1250	50%
0		2499	1250	50%



DATA CLEANING AND PREPARATION

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

- Step 2: Inserted a new cell for standard values with below formula:

```
=IF(LEFT([@DiscountOffer], 4) = "Rs. ",  
VALUE(SUBSTITUTE([@DiscountOffer], "Rs. ", "")),  
ROUND([@[OriginalPrice (in Rs)]] *[@DiscountOffer],0))
```

=IF(LEFT([@DiscountOffer], 4) = "Rs. ", VALUE(SUBSTITUTE([@DiscountOffer], "Rs. ", "")), ROUND([@[OriginalPrice (in Rs)]] *[@DiscountOffer],0))							
	G	H	I	J	K	L	M
er	Description	Discount Price w	DiscountPrice (in Rs	OriginalPrice (in Rs	Stanadard Discount Offer in Rs	DiscountOffer	%Discount for all
	roadster men	824	824	1499	675	45%	
	locomotive me	517	517	1149	632	55%	
	roadster men	629	629	1399	769	55%	
	zivame womer	893	893	1295	401	31%	
	roadster wom	0	599	999	210	35%	
	mast harbour	0		599	240	40%	
	highlander me	599	599	1499	899	60%	
	mayra pink em	0		1395	809	58%	
	roadster wom	523		1098	0		
	herenow men	1036		2749	0		
	hrx by hrithik r	1214	1214	2699	1484	55%	

Myntra Fasion Clothing

Accessibility: Investigate



DATA CLEANING AND PREPARATION

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

Resolution: Created a new column with the below formula where:

- I have first looked for the blank value in Discount price and discount offer,
- Then used Average function to fill the value in blank discount price
- I have then round of the value to get the whole numbers

=IF(AND(ISBLANK([@[DiscountPrice (in Rs)]]), [@[Stanadard Discount Offer in Rs]] = 0), ROUND(AVERAGEIFS(I:I,D:D,[@Category]),0), [@[DiscountPrice (in Rs)]])

=IF(AND(ISBLANK([@[DiscountPrice (in Rs)]]), [@[Stanadard Discount Offer in Rs]] = 0), ROUND(AVERAGEIFS(I:I,D:D,[@Category]),0), [@[DiscountPrice (in Rs)]])										
	D	E	F	G	H	I	J	K	L	M
Name	Category	Individual_category	category_by_Gen	Description	Discount Price w	DiscountPrice (in Rs)	OriginalPrice (in Rs)	Stanadard Discount Offer in Rs	DiscountOffer	%Discount for all
er	Bottom Wear	jeans	Men	roadster men	824	824	1499	675	45%	
OTIVE	Bottom Wear	track-pants	Men	locomotive m	517	517	1149	632	55%	
er	Topwear	shirts	Men	roadster men	629	629	1399	769	55%	
	Lingerie & S	shapewear	Women	zivame wome	893	893	1295	401	31%	
er	Western	tshirts	Women	roadster won	0		599	210	35%	
Harbou	Western	tops	Women	mast harbour	0		599	240	40%	
ANDER	Bottom Wear	trousers	Men	highlander m	599	599	1499	899	60%	
	Western	tops	Women	mayra pink er	0		1395	809	58%	
er	Western	tshirts	Women	roadster won	523		1098	0		
NOW	Bottom Wear	jeans	Men	herenow mer	1036		2749	0		
Hrithik	Sports Wear	tights	Men	hrx by hrithik	1214	1214	2699	1484	55%	



Myntra



DATA CLEANING AND PREPARATION

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

Resolution: There are no null or blank values found in the "SizeOption" Column. I have used find option and then also tested with the below formula:

=IF(COUNT(ISBLANK(O:O)) = 0, "No null values found", "Null values found")

=IF(COUNT(ISBLANK(O:O)) = 0, "No null values found", "Null values found")			O	P	Q	R	S	T
SizeOption	Ratings	Reviews						
28, 30, 32, 34, 36	3.9	999						
S, M, L, XL	4	999						
38, 40, 42, 44, 46, 48	4.3	999						
S, M, L, XL, XXL	4.2	999						
XS, S, M, L, XL	4.2	999						
XS, S, M, L, XL	4.4	999						



DATA ANALYSIS

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

Resolution: Calculated average using “AVERAGEIF” function as shown below:

=ROUND(AVERAGEIF(P:P, ">=4", J:J), 0)

=ROUND(AVERAGEIF(P:P, ">=4", J:J), 0)				R	S	T
SizeOption	Ratings	Reviews				
28, 30, 32, 34, 36	3.9	999				
S, M, L, XL	4	999				
38, 40, 42, 44, 46, 48	4.3	999		No null values found		
S, M, L, XL, XXL	4.2	999				
XS, S, M, L, XL	4.2	999				
XS, S, M, L, XL	4.4	999				
30, 32, 34, 36	3.9	998			Overall average original price for products with ratings greater than 4	1837



DATA ANALYSIS

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

Resolution: Please refer the below steps:

- Created a new column “%Discount for all” to get original % from “Standard Discount Offer in rs” column by using below formula: $=([@[Stanadard Discount Offer in Rs]] / [@[OriginalPrice (in Rs)]])$

M2	✕	✓	<i>fx</i>	$=([@[Stanadard Discount Offer in Rs]] / [@[OriginalPrice (in Rs)]])$			
	J	K	L	M			
	OriginalPrice (in Rs)	Stanadard Discount Offer in Rs	DiscountOffer	%Discount for all			
1	1499	675	45%	45%			
2	1149	632	55%	55%			
3	1399	769	55%	55%			
4	1295	401	31%	31%			
5	599	210	35%	35%			
6	599	240	40%	40%			
7							
8	1499	899	60%	60%			
9	1395	809	58%	58%			

- Now counted the product which has 50% or 50% plus off
 $=\text{COUNTIF}(M:M, ">=50\%")$

T8	✕	✓	<i>fx</i>	$=\text{COUNTIF}(M:M, ">=50\%")$			
	M	P	Q	R	S	T	
	%Discount for all	Ratings	Reviews				
1	45%	3.9	999				
2	55%	4	999				
3	55%	4.3	999				
4	31%	4.2	999				
5	35%	4.2	999				
6	40%	4.4	999				
7							
8	60%	3.9	998				
9	58%	3.7	998				
10	0%	4.3	997				

Count the number of products with a discount offer greater than 50% OFF38978



DATA ANALYSIS

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

Resolution: Counted the value using below formula

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

✕ ✓ <i>fx</i>	=COUNTIF(O:O,"*M*")						
N	O	P	Q	R	S	T	
	SizeOption	Ratings	Reviews				
	28, 30, 32, 34, 36	3.9	999				
	S, M, L, XL	4	999				
	38, 40, 42, 44, 46, 48	4.3	999	No null values found			
	S, M, L, XL, XXL	4.2	999				
	XS, S, M, L, XL	4.2	999				
	XS, S, M, L, XL	4.4	999				
	30, 32, 34, 36	3.9	998				
	S, M, L, XL	3.7	998				
	XS, S, M, L, XL	4.3	997				
	28, 30, 32, 34, 36	3.5	996				
	S, M, L, XL, XXL	4.4	996	Count the number of products available in size "M."		39205	
	XS, S, M, L, XL, XXL, 3XL, 4XL	4.1	996				
	S, M, L, XL, XXL	4.2	996				
	XS, S, M, L, XL	4.3	996				
	XS, S, M, L, XL	4	996				



DATA ANALYSIS

Problem 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."

Resolution: Used the below formula to provide a label based on discount offer:

=IF(NOT([@[%Discount for all]] = 0),IF([@[%Discount for all]]>=50%, "High Discount", "Low Discount"), "No Discount")

=IF(NOT([@[%Discount for all]] = 0),IF([@[%Discount for all]]>=50%, "High Discount", "Low Discount"), "No Discount")				
Standard Discount Offer in Rs	DiscountOffer	%Discount for all	Discount Status	SizeOf
675	45%	45%	Low Discount	28, 30
632	55%	55%	High Discount	S, M, L
769	55%	55%	High Discount	38, 40
401	31%	31%	Low Discount	S, M, L
210	35%	35%	Low Discount	XS, S,
240	40%	40%	Low Discount	XS, S,
899	60%	60%	High Discount	30, 32
809	58%	58%	High Discount	S, M, L
0		0%	No Discount	XS, S,
0		0%	No Discount	28, 30
1484	55%	55%	High Discount	S, M, L
0		0%	No Discount	XS, S,
2379	70%	70%	High Discount	S, M, L



DATA RETRIEVAL AND LOOKUP

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

Resolution: I have tried with both. Please refer the below formulas and snapshots:

Using VLOOKUP

Brand: =VLOOKUP(T17,B:P,2,FALSE)

=VLOOKUP(T17,B:P,2,FALSE)			R	S	T
	L	M			
	DiscountOffer	%Discount			
675	45%	45%			
632	55%	55%			
769	55%	55%			
779	60%	60%			
682	53%	53%			
002	59%	59%			
0		0%			
811	61%	61%			
749	75%	75%			
0		0%			

Original Price: =VLOOKUP(T17,B:J,9,FALSE)

=VLOOKUP(T17,B:J,9,FALSE)								
	L	M		R		S		T
is	DiscountOffer	%Discou						
675	45%	45%						
632	55%	55%						
769	55%	55%			No null values found			
479	60%	60%						
582	53%	53%				Product ID	11226634	
1002	59%	59%						
0		0%						
						Details Using VLOOKUP		
1311	61%	61%				BrandName	Maniac	
749	75%	75%				Original Price	1199	
0		0%				Ratings	3.9	

Rating: =VLOOKUP(T17,B:P,15, FALSE)

=VLOOKUP(T17,B:P,15, FALSE)			R	S	T
	L	M			
	DiscountOffer	%Discount			
675	45%	45%			
632	55%	55%			
769	55%	55%			
479	60%	60%			
582	53%	53%			
002	59%	59%			
0		0%			
311	61%	61%			
749	75%	75%			
0		0%			



DATA RETRIEVAL AND LOOKUP

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

Resolution: I have tried with both. Please refer the below formulas and snapshots:
Using XLOOKUP

Brand: =XLOOKUP(T17,B:B,C:C)

=XLOOKUP(T17,B:B,C:C)			S	T
675	DiscountOffer	%Discou		
632				
769			No null values found	
479				
582			Product ID	11226634
002				
750			Details Using XLOOKUP	
805			Brand	Maniac
740			Original Price	1199
889			Rating	3.9
548				

Original Price: =XLOOKUP(T17,B:B,J:J)

=XLOOKUP(T17,B:B,J:I)								
	L	M	R	S	T			
5	Discount Offer	%Discount						
675	45%	45%						
632	55%	55%						
769	55%	55%		No null values found				
479	60%	60%						
582	53%	53%		Product ID	11226634			
002	59%	59%						
750	50%	50%		Details Using XLOOKUP				
805	62%	62%		Brand	Morice			
740	57%	57%		Original Price	1199			
889	70%	70%		Rating	3.9			
522	60%	60%						

Rating: =XLOOKUP(T17,B:B,P:P)

=XLOOKUP(T17,B:B,P:P)			
	L	M	
s	DiscountOffer	%Discou	
675	45%	45%	
632	55%	55%	
769	55%	55%	No null values found
479	60%	60%	
582	53%	53%	
002	59%	59%	
750	50%	50%	
805	62%	62%	
740	57%	57%	
889	70%	70%	
549	61%	61%	



DATA RETRIEVAL AND LOOKUP

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

Resolution: Please refer the below formula:

=INDEX(H:H,MATCH(T31,B:B,0))

=INDEX(H:H,MATCH(T31,B:B,0))			P	Q	R	S	T
SizeOption	Ratings	Reviews					
XS, S, M, L, XL, XXL, 3XL	4	990					
26, 28, 30, 32, 34, 36	4.3	990					
L, M, S, XS, XL	4.2	990					
S, M, L, XL, XXL	4.4	990					
S, M, L, XL, XXL	4.3	990					
XS, S, M, L, XL	4.3	990					
XS, S, M, L, XL	4.1	990					
Onesize	4.2	989					
S, M, L, XL	4.3	989					
28, 30, 32, 34, 36, 38	4	988					

Product ID	6744434
Discount Price	599



DATA RETRIEVAL AND LOOKUP

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

Resolution: I have achieved it with 3 steps. Please refer the details below:

I. Created a list of "EmployeeID" using data validation

The screenshot shows an Excel spreadsheet with a column header 'V' and a dropdown menu labeled 'Select Product ID'. The dropdown currently displays '12873874'. To the right, the 'Data Validation' task pane is open, showing the 'Settings' tab. The 'Allow' dropdown is set to 'List', and the 'Source' field contains the formula '=\$B:\$B'. The 'Ignore blank' and 'In-cell dropdown' checkboxes are both checked.

V
Select Product ID
12873874

Data Validation Settings

Validation criteria

Allow: List

Data: between

Source: =\$B:\$B

☒ Ignore blank

☒ In-cell dropdown

II. Created a list of "Column Headers" using data validation

The screenshot shows an Excel spreadsheet with a column header 'vv' and a dropdown menu labeled 'Select Column Headers'. The dropdown currently displays 'Ratings'. To the right, the 'Data Validation' task pane is open, showing the 'Settings' tab. The 'Allow' dropdown is set to 'List', and the 'Source' field contains the formula '=\$C\$1:\$Q\$1'. The 'Ignore blank' and 'In-cell dropdown' checkboxes are both checked.

vv
Select Column Headers
Ratings

Data Validation Settings

Validation criteria

Allow: List

Data: between

Source: =\$C\$1:\$Q\$1

☒ Ignore blank

☒ In-cell dropdown



DATA RETRIEVAL AND LOOKUP

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

Resolution: Please refer the below step 3:

III. Used that list in nested XLOOKUP formula to get dynamic value for any columns:

=XLOOKUP(V2,B:B,XLOOKUP(W2,Table1[#Headers],[Product_id]:[Reviews]],B:Q):[Reviews]],B:Q))

=XLOOKUP(V2,B:B,XLOOKUP(W2,Table1[#Headers],[Product_id]:[Reviews]],B:Q):[Reviews]],B:Q))		
V	W	X
Select Product ID	Select Column Headers	Value using Nested XLOOKUP
12873874	Ratings	4.2





Thank You

Presented by: Shivani Awasthi