Myntra Case study

ANALYSIS OF MYNTRA APPAREL   BY PARTH

I have take the data of first 1001 rows

Project Questions
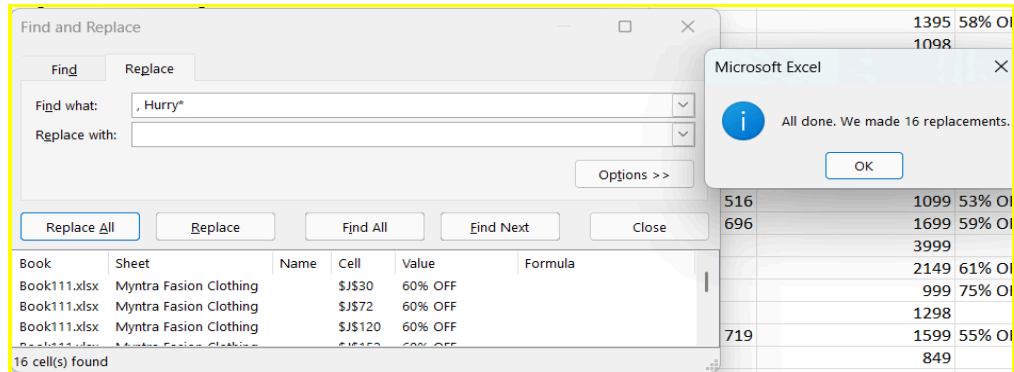
A. Data Cleaning and Preparation

1. Check for duplicate values in your dataset and remove them.
   - NO duplicates found

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

   step1

   - Remove , Hurry* using "ctrl+F "   -16 replacements / 1001 rows

   Step 2   -   Same way replace OFF using ctrl+f   963 replacements / 1001 rows

   Step 3 - Convert Rs to % or % to Rs ……..i take % to Rs .

   Step 4 - use formula to get the values

`=IF(ISNUMBER(SEARCH("RS",(J2))),(J2),(J2)*(I2))`

G	H	I	J	K	L	M	N	
	DiscountPric	OriginalPrice (in Rs)	Discounter	SizeOption	Ratings	Reviews	DISCOUNT IN RS	
it mid rise clean look jeans	824	1499	45%	28, 30, 32, 34, 36	3.9	999	674.550	P
solid slim fit track pants	517	1149	55%	S, M, L, XL	4	999	632	
ck geometric printed sustainable casual sh	629	1399	55%	38, 40, 42, 44, 46, 48	4.3	999	769	A
apewear zi3023core0nude	893	1295	31%	S, M, L, XL, XXL	4.2	999	401	
neck pure cotton t shirt		599	35%	XS, S, M, L, XL	4.2	999	210	
solid tank top		599	40%	XS, S, M, L, XL	4.4	999	240	
m fit solid regular trousers	599	1499	60%	30, 32, 34, 36	3.9	998	899	
e pure cotton top		1395	58%	S, M, L, XL	3.7	998	809	
lid t shirts		1098		XS, S, M, L, XL	4.3	997	0	
d rise clean look ankle length stretchable jeans		2749		28, 30, 32, 34, 36	3.5	996	0	
d dry training tights	1214	2699	55%	S, M, L, XL, XXL	4.4	996	1484	

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

(note :- if we replace zero directly using replace it will replace zero from the column where you have the value too ex - 270 ,400 it will remove there zero too) …Rows were both null ? Discount offer 94 rows with zero replace with null or blank

`=IF(AND(H2=" ",J2=" "),AVERAGEIF(D:D,D2,H:H),H2)`

Description	DiscountPrice (i	riginalPrice (i	Discounter	SizeOption	Rating	Reviews	unt % convert Rs(b	discounter	ge discount price in respective cate
roadster men navy blue slim fit mid rise clean look jeans	824	1499	45%	28, 30, 32, 34, 36	3.9	999	675	Low discount	824
locomotive men black white solid slim fit track pants	517	1149	55%	S, M, L, XL	4	999	632	High discount	517
roadster men navy white black geometric printed sustainable casual shirt	629	1399	55%	38, 40, 42, 44, 46, 48	4.3	999	769	High discount	629
zivame women black saree shapewear zi3023core0nude	893	1295	31%	S, M, L, XL, XXL	4.2	999	401	Low discount	893
roadster women white solid v neck pure cotton t shirt		599	35%	XS, S, M, L, XL	4.2	999	210	Low discount	0
mast harbour women yellow solid tank top		599	40%	XS, S, M, L, XL	4.4	999	240	Low discount	0
highlander men olive green slim fit solid regular trousers	599	1499	60%	30, 32, 34, 36	3.9	998	899	High discount	599
mayra pink embroidered a line pure cotton top		1395	58%	S, M, L, XL	3.7	998	809	High discount	0
roadster women pack of 2 solid t shirts		1098		XS, S, M, L, XL	4.3	997		Low discount	0
herenow men blue slim fit mid rise clean look ankle length stretchable jeans		2749		28, 30, 32, 34, 36	3.5	996		Low discount	0
hrx by hrithik roshan men rapid dry training tights	1214	2699	55%	S, M, L, XL, XXL	4.4	996	1484	High discount	1214
roadster men navy blue white striped polo collar pure cotton t shirt		699		XS, S, M, L, XL, XXL, 3XL, 4XL	4.1	996		Low discount	0
anubhutee women pink white printed kurta with trousers dupatta	1019	3399	70%	S, M, L, XL, XXL	4.2	996	2379	High discount	1019
athena women black solid basic jumpsuit		2499	50%	XS, S, M, L, XL	4.3	996	1250	Low discount	0
roadster women maroon solid round neck t shirt		799	60%	XS, S, M, L, XL	4	996	479	High discount	0
highlander men blue black slim fit printed casual shirt	516	1099	53%	39, 40, 42, 44	4.2	995	582	High discount	516
vishudh women blue pink floral print a line kurta	696	1699	59%	S, M, L, XXL	4.2	995	1002	High discount	696
sangria women green white printed kurta with palazzos dupatta		3999		XS, S, M, L, XL, XXL, 3XL	4.3	995		Low discount	0
tokyo talkies women olive green regular fit solid cargos		2149	61%	28, 30, 32, 34	4.1	995	1311	High discount	0
dressberry women lovely olive typography talking tee		999	75%	XS, S, M, L, XL	4.3	995	749	High discount	0
roadster women pack of 2 printed cotton t shirt		1298		XS, S, M, L, XL	4.4	994		Low discount	0
anouk women mustard yellow gold toned printed straight kurta	719	1599	55%	XS, S, M, L, XL, XXL	3.8	993	879	High discount	719
enamor black non wired removable pads high coverage medium impact sports bra sb08		849		S, M, L, XL	4.2	993		Low discount	0
all about you women mauve stretchable casual shirt		1499	50%	XS, S, M, L, XL	4.2	993	750	Low discount	0

`=IF(AND(H2=" ",J2=" "),AVERAGEIF(D:D,D2,H:H),H2)`

Here …..in this formula we can use three functions i.e IF AND AVERAGEIF

STEPS TO FOLLOW

-   Create a new column use if function
-   Here H2 is the column of discount price in Rs .And J2 is discount in %.
-   Now for Averageif select the entire column of category i.e D:D then select the specific range of the category i.e D2 than select the entire column of discount price , close the bracket and fill it with discount price if it is not null .

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.

=AVERAGEIF(L:L,">4",[OriginalPrice (in Rs)])

1669.24

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

=COUNTIF([DiscountOffer], ">50%OFF")

534 /1001

3. Count the number of products available in size "M."
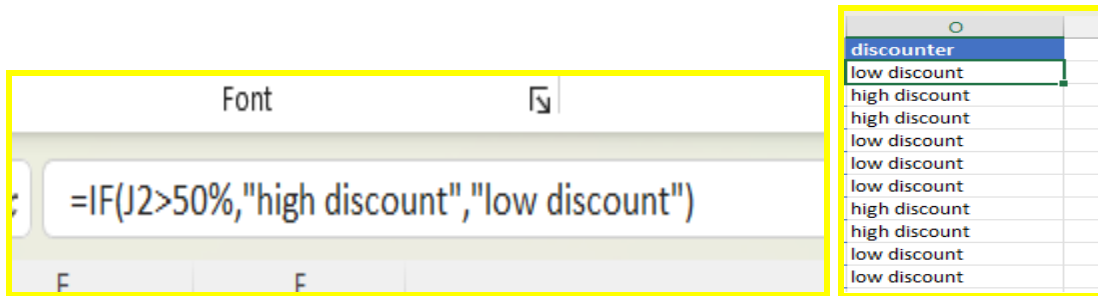
=COUNTIF(K:K,"*M*")

667/1001

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."

=IF(J:J>50%,"high discount","low discount")

E	F	G	H	I	J	K	L	M	N	O
al_catego	https://youtu	Description	iscountPrice (i	riginalPrice (i	Discounter	SizeOption	Rating	Reviews	unt % convert Rs(b	discounter
	Men	roadster men navy blue slim fit mid rise clean look jeans	824	1499	45%	28, 30, 32, 34, 36	3.9	999	675 ⚠	#SPILL!
nts	Men	locomotive men black white solid slim fit track pants	517	1149	55%	S, M, L, XL	4	999	632	
	Men	roadster men navy white black geometric printed sustainable casual shirt	629	1399	55%	38, 40, 42, 44, 46, 48	4.3	999	769	

Note - don't select the entire column

Font				discounter
				low discount
				high discount
				high discount
=IF(J2>50%,"high discount","low discount")				low discount
				low discount
				low discount
				high discount
F		F		high discount
				low discount
				low discount

— Select that particular column only i.e J2

C. Data Retrieval and Lookup

1. Use VLOOKUP/ XLOOKUP to find the product brand, price, and rating of the product with Product_id "11226634" 12391750.

   The Id "11226634" is not in my dataset so I can replace it with "12391750".

   - table value wahh se start hoge ja se lookup value hoge

			12391750
1 Original price	=VLOOKUP(R6,B2:P1002,8,FALSE)	original price	2149

2 Brand	=VLOOKUP(R6,B2:P1002,2,FALSE)	Product Brand	Tokyo Talkies

3 Rating	=VLOOKUP(R6,B2:P1002,11,FALSE)	rating	4.1

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

		6744434
=INDEX(H:H,MATCH(R13,B:B,0))	indexmatch	599