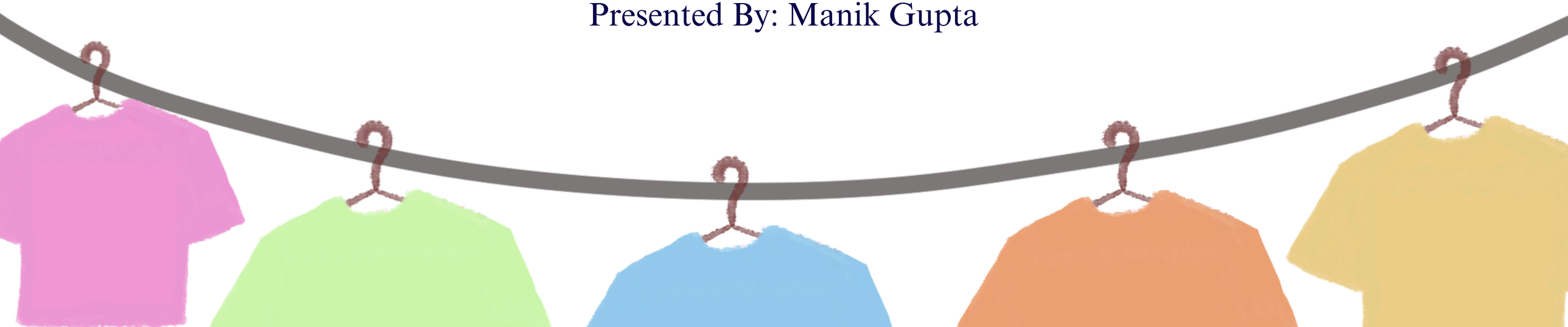
Analysis of Myntra Apparel

Presented By: Manik Gupta



Introduction

Myntra is a leading online fashion and lifestyle e-commerce platform in India, established in 2007. It offers a diverse range of products, including clothing, footwear, accessories, and beauty items from numerous national and international brands. Known for its customer-centric approach, Myntra leverages data analytics to enhance user experience, optimize inventory management, and drive personalized marketing strategies. This project aims to solve problems related to Data Cleaning and Preparation, Data Analysis and Data Retrieval & Lookup.

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."

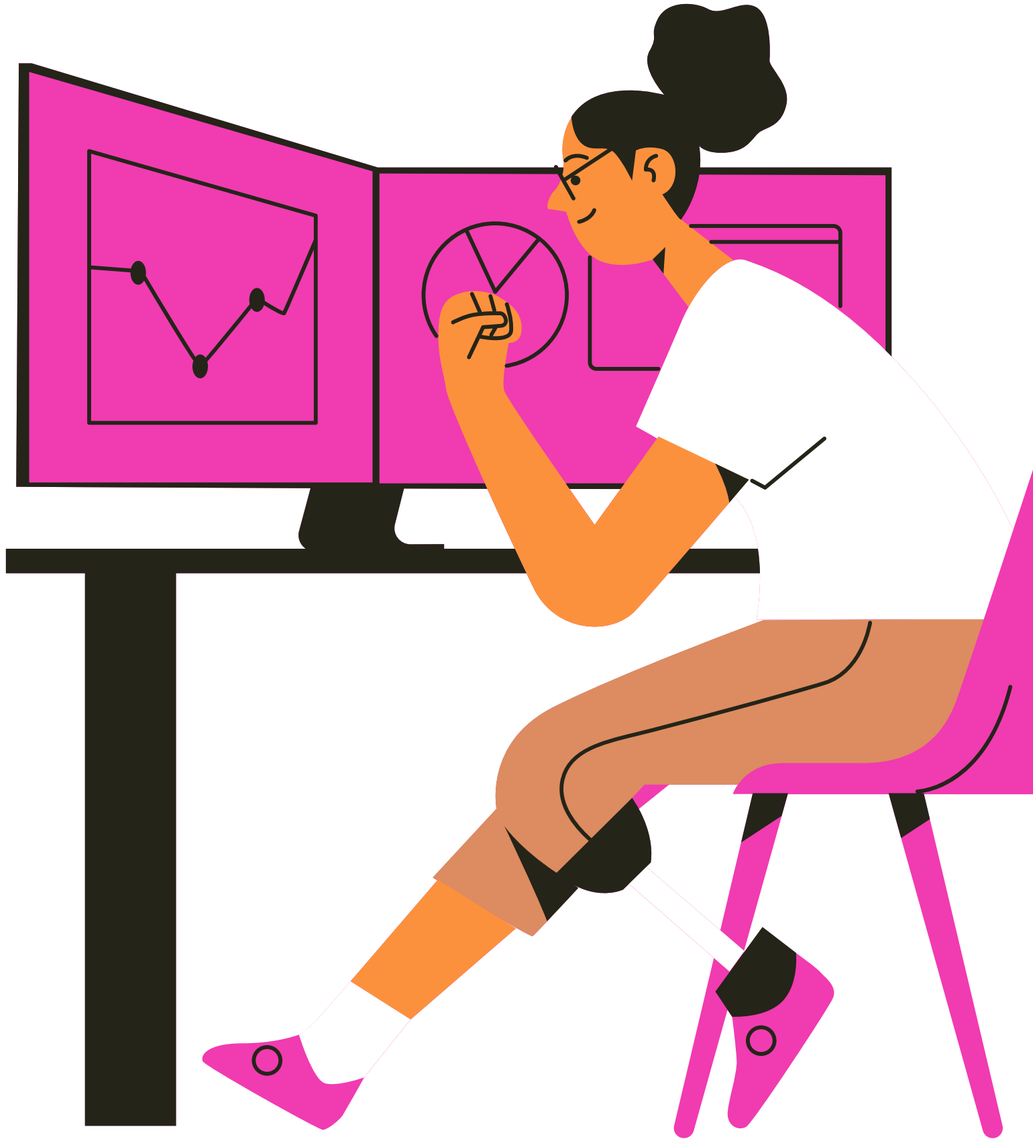
B. Data Analysis

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

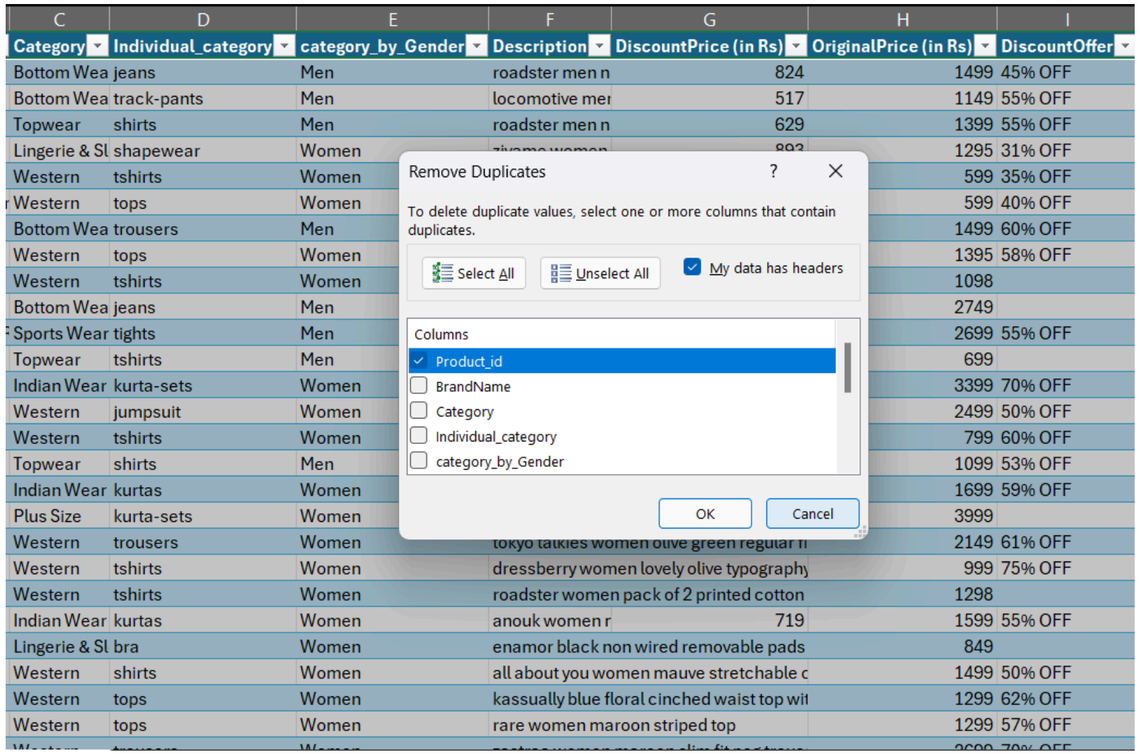
1. Use VLOOKUP/XLOOKUP to find the product brand, price, and rating of the product with Product_id "11226634".
2. Find the "DiscountPrice" for the product with the Product ID "6744434" using the INDEX and MATCH functions.
3. Utilize nested xlookup to find any column’s detail of a product with it’s product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

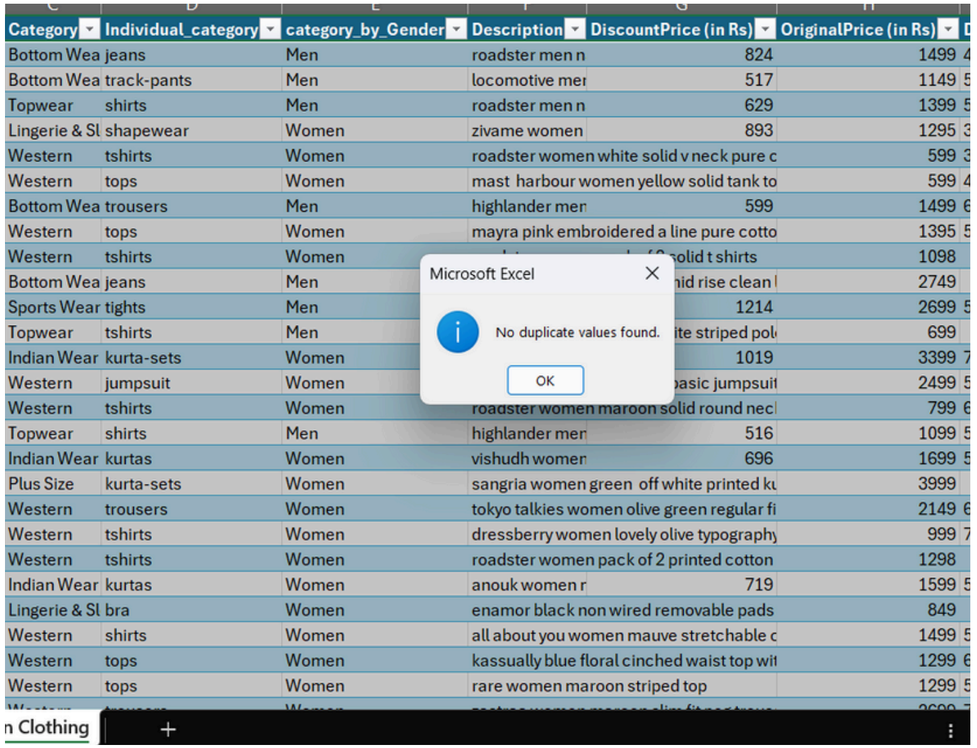
Data Cleaning and Preparation



Check for duplicate values in your dataset and remove them.

Select the unique column i.e. Product ID, go to Data > Remove Duplicates, and check only the Product ID and then click OK.





Standardize the "DiscountOffer" column to a uniform format

Step 1 : We have to separate all the values using Text to column using blank space delimited in Discount Offer column and then it will create 3 new columns

Step 2 : Delete the last 2 columns that contains “OFF”, “Hurry*” which is not required to standardize the data.

Step 3 : Type the following formula to extract the Standardize Discount Offer in a new column:

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

I	J	K	L	M	N	O	P	Q
DiscountOffer	SizeOption	Ratings	Reviews	Column1	Column2	Standardize Discount Amount		
45% OFF	28, 30, 32, 34, 3	3.9	999	45%	OFF	=IF(ISNUMBER([@Column2]),[@Column2],IF(ISBLANK(		
55% OFF	S, M, L, XL	4	999	55%	OFF	[@Column1]),"",[@Column1]*[@[OriginalPrice (in Rs)]]))		
55% OFF	38, 40, 42, 44, 4	4.3	999	55%	OFF	769.45		
31% OFF	S, M, L, XL, XXL	4.2	999	31%	OFF	401.45		
35% OFF	XS, S, M, L, XL	4.2	999	35%	OFF	209.65		
40% OFF	XS, S, M, L, XL	4.4	999	40%	OFF	239.6		
60% OFF	30, 32, 34, 36	3.9	998	60%	OFF	899.4		
58% OFF	S, M, L, XL	3.7	998	58%	OFF	809.1		
	XS, S, M, L, XL	4.3	997					
	28, 30, 32, 34, 3	3.5	996					
55% OFF	S, M, L, XL, XXL	4.4	996	55%	OFF	1484.45		
	XS, S, M, L, XL, X	4.1	996					
70% OFF	S, M, L, XL, XXL	4.2	996	70%	OFF	2379.3		
50% OFF	XS, S, M, L, XL	4.3	996	50%	OFF	1249.5		
60% OFF	XS, S, M, L, XL	4	996	60%	OFF	479.4		
53% OFF	39, 40, 42, 44	4.2	995	53%	OFF	582.47		

IF(logical_test, [value

Standardized Discount Offer into Rs column

Filling the “Discount Price” with category average

In order to fill out all the null value on the basis of the category average, we have use the following formula:

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

dardize Discount Amount],[Category],[@Category]),[@[Standardize Discount Amount]])

I	J	K	L	M	N	O
Ratings	Reviews	Standardize Discount Amoun	Actual Price	Column1		
3.9	999	674.55	824	=IF(ISBLANK([@[Standardize Discount		
4	999	631.95	517	Amount]]),AVERAGEIFS([Standardize		
4.3	999	769.45	629	Discount Amount],[Category],		
4.2	999	401.45	893	[@Category]),[@[Standardize Discount		
4.2	999	209.65	389.35	Amount]])		
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			900.8663407		
3.5	996			1101.002024		
4.4	996	1484.45	1214	1484.45		
4.1	996			927.8471517		
4.2	996	2379.3	1019	2379.3		
4.3	996	1249.5	1249.5	1249.5		

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

In order to replace all null values in the “SizeOption” column with the text “Not Available”, then we can use the below formula in a new column:

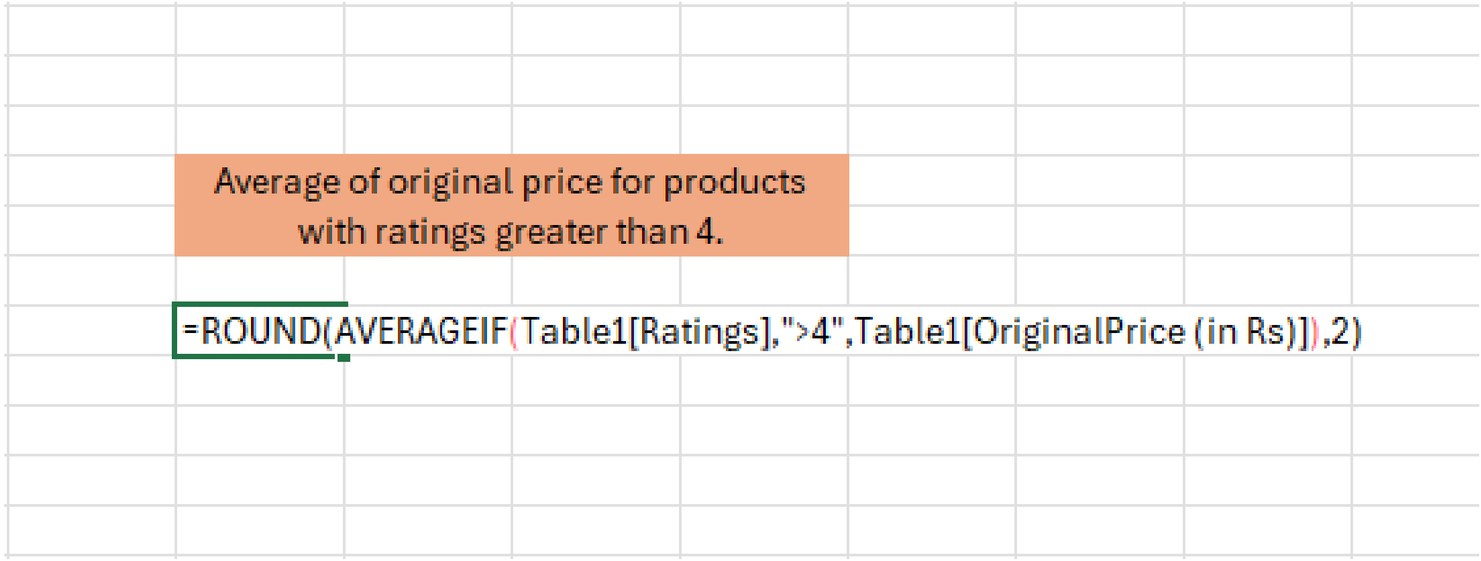
=IF(ISBLANK([@SizeOption]),"NOT AVAILABLE",[@SizeOption])

H	I	J	K	L	
SizeOption	Ratings	Reviews	Actual Price	Discount Amount	SizeOption2
28, 30, 32, 34, 3	3.9	999	824	674.55	=IF(ISBLANK([@SizeOption]),"NOT AVAILABLE",[@SizeOption])
S, M, L, XL	4	999	517	631.95	S, M, L, XL
38, 40, 42, 44, 4	4.3	999	629	769.45	38, 40, 42, 44, 46, 48
S, M, L, XL, XXL	4.2	999	893	401.45	S, M, L, XL, XXL
XS, S, M, L, XL	4.2	999	389.35	209.65	XS, S, M, L, XL
XS, S, M, L, XL	4.4	999	359.4	239.6	XS, S, M, L, XL
30, 32, 34, 36	3.9	998	599	899.4	30, 32, 34, 36
S, M, L, XL	3.7	998	585.9	809.1	S, M, L, XL
XS, S, M, L, XL	4.3	997		900.87	XS, S, M, L, XL
28, 30, 32, 34, 3	3.5	996		1101	28, 30, 32, 34, 36

In above sheet, if there will be any null values available in SizeOption, it will convert them into the “NOT AVAILABLE” in SizeOption2 column where we have applied the formula.

Data Analysis

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



To calculate the overall average original price for products with ratings greater than 4, we can use the below formula:

=ROUND(AVERAGEIF(Table1[Ratings],">4",Table1[OriginalPrice (in Rs)]),2)

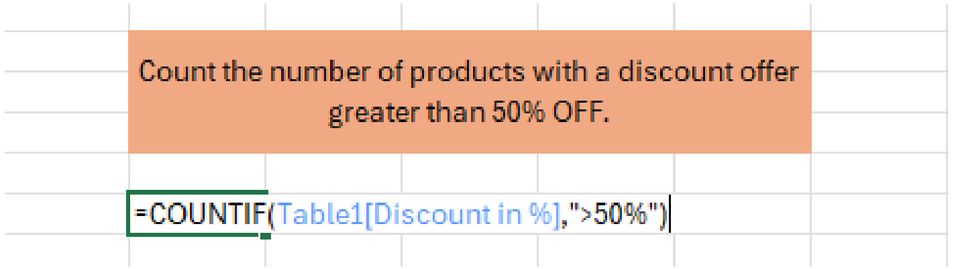
Here, ROUND is used to get the value upto 2 Decimal.

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

Step 1: Firstly, we need to convert Discount Offer which is in rupees to percentage form using the formula: =[@[Discount Amount]]/[@[OriginalPrice (in Rs)]] and then select %.

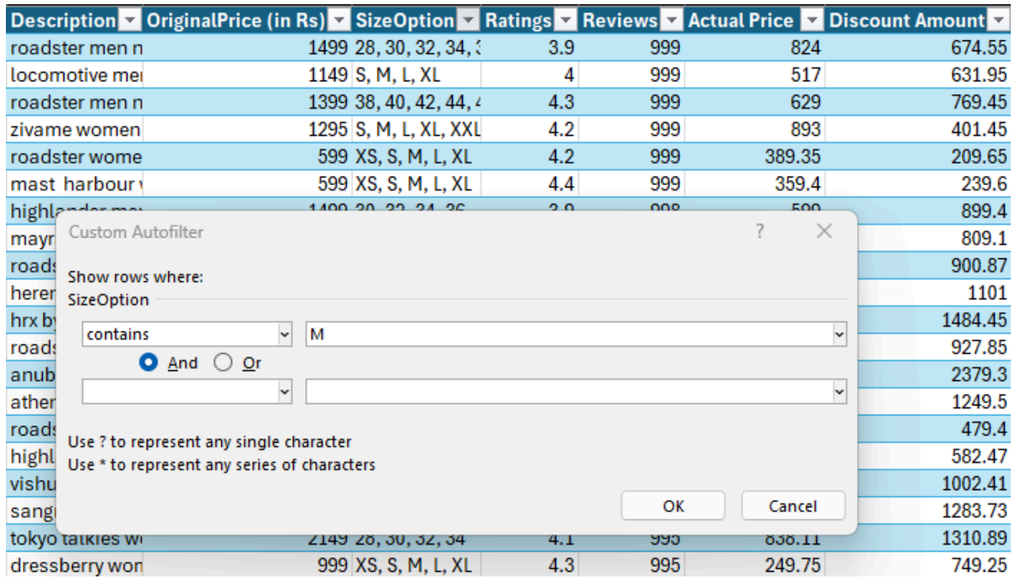
OriginalPrice (in Rs)	SizeOption	Ratings	Reviews	Actual Price	Discount Amount	Discount in %
1499	28, 30, 32, 34, 3	3.9	999	824	674.55	=[@[Discount Amount]]/[@[OriginalPrice (in Rs)]]
1149	S, M, L, XL	4	999	517	631.95	0.55
1399	38, 40, 42, 44, 4	4.3	999	629	769.45	0.55
1295	S, M, L, XL, XXL	4.2	999	893	401.45	0.31
599	XS, S, M, L, XL	4.2	999	389.35	209.65	0.35
599	XS, S, M, L, XL	4.4	999	359.4	239.6	0.4
1499	30, 32, 34, 36	3.9	998	599	899.4	0.6
1395	S, M, L, XL	3.7	998	585.9	809.1	0.58

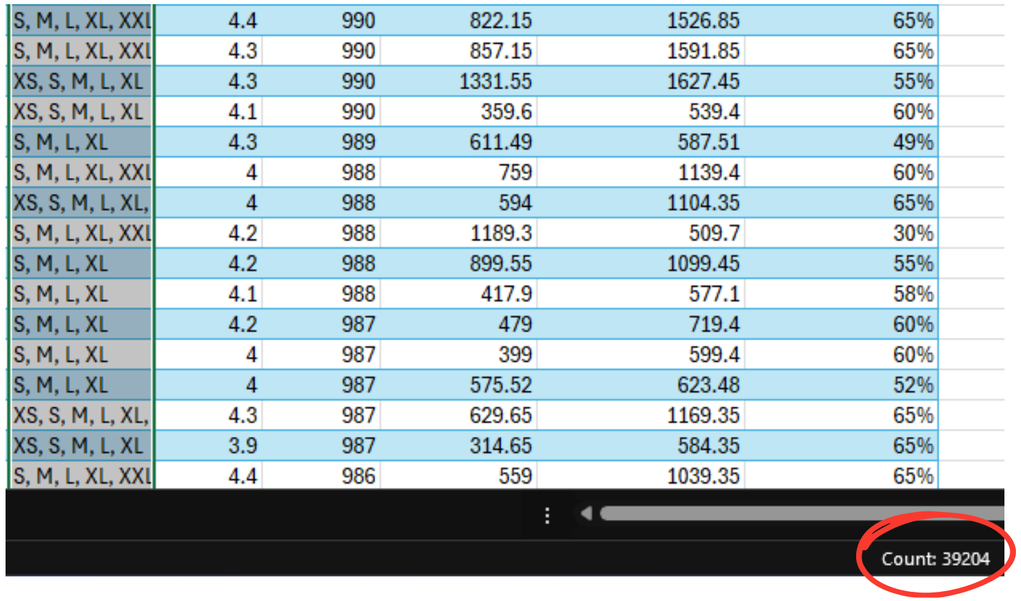
Step 2: To count the number of products with a discount offer greater than 50% off, we need to use the formula: =COUNTIF(Table1[Discount in %],">50%")



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

Click on filter option in Size Option column > Go to Text Filter > Select Contains > Type “M” and then click OK. Select the column and it will reflect the count at the bottom of the Screen.





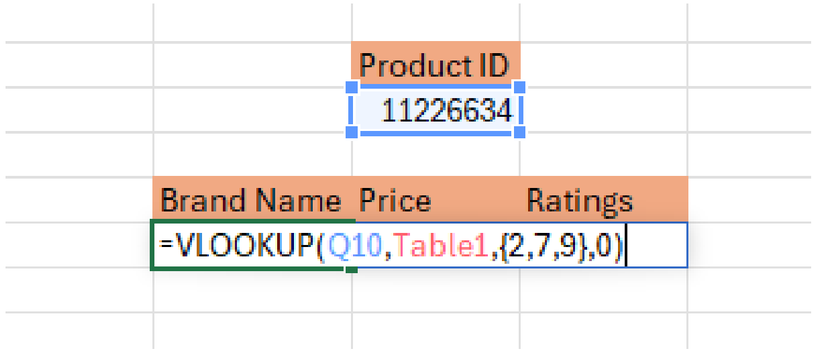
Create a new column to label the products as "High Discount" if the discount offer is greater than 50% OFF, otherwise label them as "Low Discount."

L	M	N	O	P	Q	R
Discount Amount	Discount in %	Discount Category				
674.55	45%	=IF([@[Discount in %]]>50%,"High Discount","Low Discount")				
631.95	55%	High Discount				
769.45	55%	High Discount				
401.45	31%	Low Discount				
209.65	35%	Low Discount				
239.6	40%	Low Discount				
899.4	60%	High Discount				
809.1	58%	High Discount				
900.87	82%	High Discount				
1101	40%	Low Discount				
1484.45	55%	High Discount				
927.85	133%	High Discount				

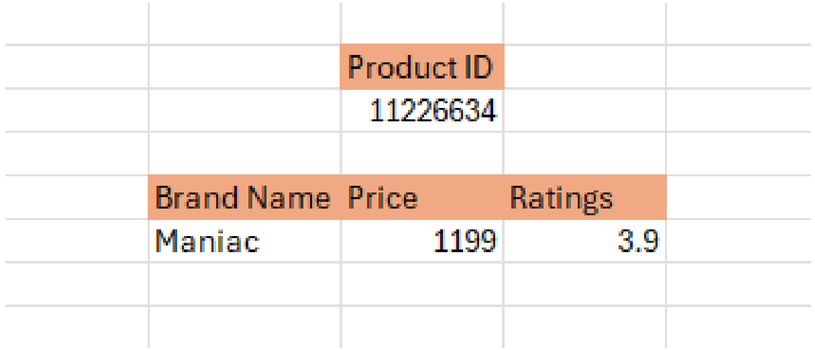
Create a column labeling offers as “High Discount” for over 50% and “Low Discount” for under 50% using the “IF” function.

Data Retrieval and Lookup

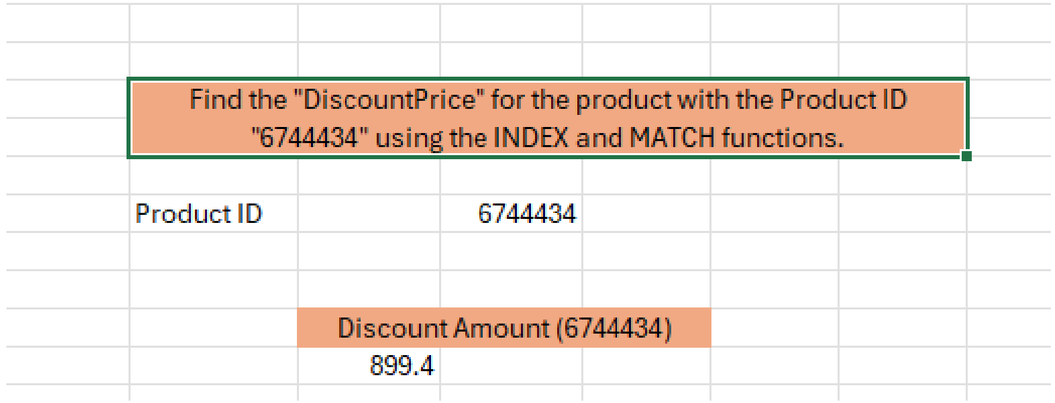
Use VLOOKUP/XLOOKUP to find the product brand, price, and rating of the product with Product_id "11226634".



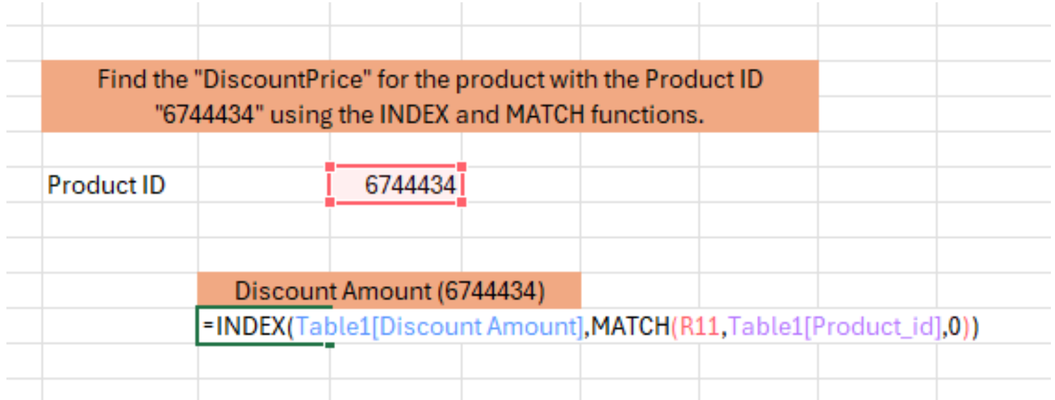
To find the product brand, price and rating of the product with Product_id “11226634”, we can use the VLOOKUP Formula : =VLOOKUP(Q10,Table1,{2,7,9},0)



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



To find the "DiscountPrice" for the product with the Product ID "6744434" using the INDEX and MATCH functions, we can use the formula: =INDEX(Table1[Discount Amount],MATCH(R11,Table1[Product_id],0))



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

PRODUCTID
10307375

Column Name	Columns Details using PRODUCT ID
Brand Name:	=XLOOKUP(R8,Table1[Product_id],Table1[BrandName], "Not Found")
Category:	Topwear
Description:	roadster men navy blue white striped polo collar pure cotton t shirt
Actual Price:	0
Size Options:	XS, S, M, L, XL, XXL, 3XL, 4XL
Ratings	4.1
Discount Amount	927.85
Discount in %	133%
Discount Category	High Discount

In order to get brand name using product key, we need to use the formula: =XLOOKUP(R8,Table1[Product_id],Table1[BrandName], "Not Found").
Similarly, we need to add this formula to each column and replace the column name in the formula. Now, if you change the product ID, it will display the column's details on the bsis of the product ID.

THANKYOU