

# Analysis of Myntra Apparel

---

Presented By: SHLOK VASHISHTH

# 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

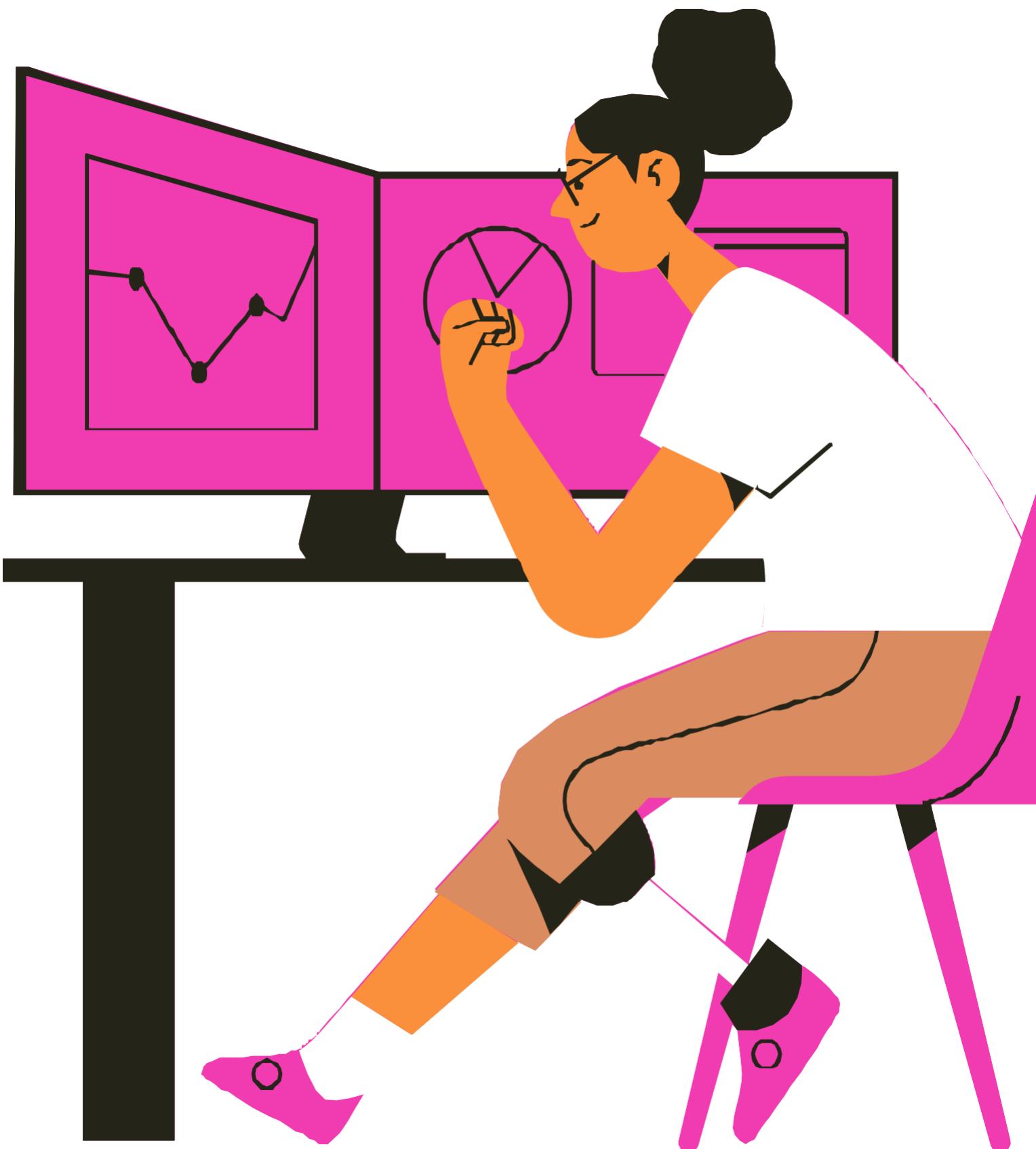
1. Calculate the overall average original price for products with ratings greater than 4.
2. Count the number of products with a discount offer greater than 50% OFF.
3. Count the number of products available in size "M."
4. Create a new column to label the products as "High Discount" if the discount 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.



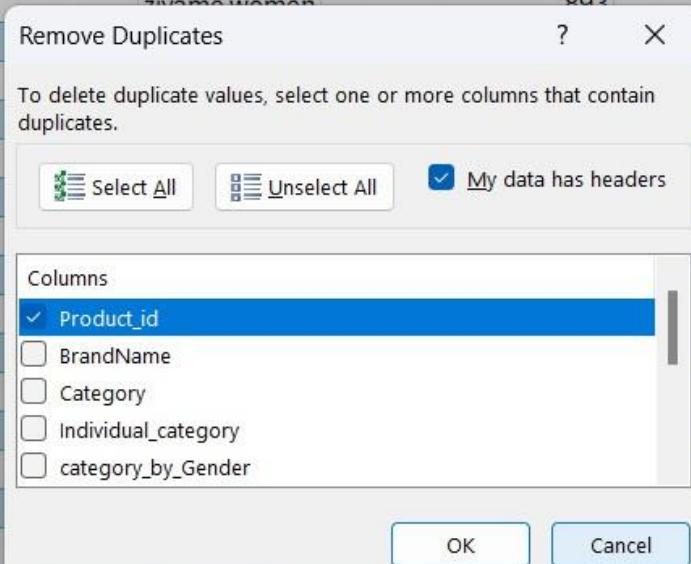
# Data Cleaning and Preparation



# 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.

Category	Individual_category	category_by_Gender	Description	DiscountPrice (in Rs)	OriginalPrice (in Rs)	DiscountOffer
Bottom Wea jeans	Men	roadster men n	824	1499	45% OFF	
Bottom Wea track-pants	Men	locomotive mer	517	1149	55% OFF	
Topwear shirts	Men	roadster men n	629	1399	55% OFF	
Lingerie & Sl shapewear	Women	zivame women	802	1295	31% OFF	
Western tshirts	Women		599	35% OFF		
Western tops	Women		599	40% OFF		
Bottom Wea trousers	Men		1499	60% OFF		
Western tops	Women		1395	58% OFF		
Western tshirts	Women		1098			
Bottom Wea jeans	Men		2749			
Sports Wear tights	Men		2699	55% OFF		
Topwear tshirts	Men		699			
Indian Wear kurta-sets	Women		3399	70% OFF		
Western jumpsuit	Women		2499	50% OFF		
Western tshirts	Women		799	60% OFF		
Topwear shirts	Men		1099	53% OFF		
Indian Wear kurtas	Women		1699	59% OFF		
Plus Size kurta-sets	Women		3999			
Western trousers	Women	tokyo talkies women olive green regular fit	2149	61% OFF		
Western tshirts	Women	dressberry women lovely olive typography	999	75% OFF		
Western tshirts	Women	roadster women pack of 2 printed cotton	1298			
Indian Wear kurtas	Women	anouk women r	719	1599	55% OFF	
Lingerie & Sl bra	Women	enamor black non wired removable pads	849			
Western shirts	Women	all about you women mauve stretchable c	1499	50% OFF		
Western tops	Women	kassually blue floral cinched waist top wil	1299	62% OFF		
Western tops	Women	rare women maroon striped top	1299	57% OFF		
Western tops	Women		2699	70% OFF		



Category	Individual_category	category_by_Gender	Description	DiscountPrice (in Rs)	OriginalPrice (in Rs)	DiscountOffer
Bottom Wea jeans	Men	roadster men n	824	1499	45% OFF	
Bottom Wea track-pants	Men	locomotive mer	517	1149	55% OFF	
Topwear shirts	Men	roadster men n	629	1399	55% OFF	
Lingerie & Sl shapewear	Women	zivame women	893	1295	31% OFF	
Western tshirts	Women	roadster women white solid v neck pure c	599	1399	35% OFF	
Western tops	Women	mast harbour women yellow solid tank to	599	1499	40% OFF	
Bottom Wea trousers	Men	highlander mer	599	1499	60% OFF	
Western tops	Women	mayra pink embroidered a line pure cotto	1395	2699	55% OFF	
Western tshirts	Women	mid rise clean	1098	2699	58% OFF	
Bottom Wea jeans	Men	mid rise clean	2749	599	50% OFF	
Sports Wear tights	Men	1214	2699	55% OFF		
Topwear tshirts	Men	white striped polo	699	1399	50% OFF	
Indian Wear kurta-sets	Women	1019	3399	70% OFF		
Western jumpsuit	Women	basic jumpsuit	2499	599	50% OFF	
Western tshirts	Women	roadster women maroon solid round nec	799	1099	55% OFF	
Topwear shirts	Men	highlander mer	516	1099	50% OFF	
Indian Wear kurtas	Women	vishudh women	696	1699	50% OFF	
Plus Size kurta-sets	Women	sangria women green off white printed ku	3999	799	50% OFF	
Western trousers	Women	tokyo talkies women olive green regular fi	2149	499	50% OFF	
Western tshirts	Women	dressberry women lovely olive typograph	999	1999	75% OFF	
Western tshirts	Women	roadster women pack of 2 printed cotton	1298	2499	50% OFF	
Indian Wear kurtas	Women	anouk women r	719	1599	50% OFF	
Lingerie & Sl bra	Women	enamor black non wired removable pads	849			
Western shirts	Women	all about you women mauve stretchable c	1499			
Western tops	Women	kassually blue floral cinched waist top wil	1299			
Western tops	Women	rare women maroon striped top	1299			
Western tops	Women		2699			



# 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([@Column1]),"",[@Column1]*[@[OriginalPrice (in Rs)]]))			
55% OFF	S, M, L, XL	4	999	55% OFF				
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		

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]])
```

I	J	K	L	M	N	O
Ratings	Reviews	Standardize Discount Amount	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	SizeOption2
SizeOption	Ratings	Reviews	Actual Price	Discount Amount	
28, 30, 32, 34, 36	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, 46, 48	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, 36	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.

Average of original price for products with ratings greater than 4.
=ROUND(AVERAGEIF(Table1[Ratings],">4",Table1[OriginalPrice (in Rs)]),2)

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, 36	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, 46	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 with a discount offer greater than 50% OFF.
=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.

Description	OriginalPrice (in Rs)	SizeOption	Ratings	Reviews	Actual Price	Discount Amount
roadster men n	1499	28, 30, 32, 34, 36	3.9	999	824	674.55
locomotive men	1149	S, M, L, XL	4	999	517	631.95
roadster men n	1399	38, 40, 42, 44, 46	4.3	999	629	769.45
zivame women	1295	S, M, L, XL, XXL	4.2	999	893	401.45
roadster wome	599	XS, S, M, L, XL	4.2	999	389.35	209.65
mast harbour	599	XS, S, M, L, XL	4.4	999	359.4	239.6
highlander men	1499	28, 30, 32, 34, 36	2.9	999	599	899.4
mayra					809.1	
road					900.87	
herer					1101	
hrx b					1484.45	
road					927.85	
anub					2379.3	
ather					1249.5	
road					479.4	
highl					582.47	
vishu					1002.41	
sang					1283.73	
tokyo talkies w	2149	28, 30, 32, 34	4.1	999	636.11	1310.89
dressberry won	999	XS, S, M, L, XL	4.3	995	249.75	749.25



S, M, L, XL, XXL	4.4	990	822.15	1526.85	65%
S, M, L, XL, XXL	4.3	990	857.15	1591.85	65%
XS, S, M, L, XL	4.3	990	1331.55	1627.45	55%
XS, S, M, L, XL	4.1	990	359.6	539.4	60%
S, M, L, XL	4.3	989	611.49	587.51	49%
S, M, L, XL, XXL	4	988	759	1139.4	60%
XS, S, M, L, XL,	4	988	594	1104.35	65%
S, M, L, XL, XXL	4.2	988	1189.3	509.7	30%
S, M, L, XL	4.2	988	899.55	1099.45	55%
S, M, L, XL	4.1	988	417.9	577.1	58%
S, M, L, XL	4.2	987	479	719.4	60%
S, M, L, XL	4	987	399	599.4	60%
S, M, L, XL	4	987	575.52	623.48	52%
XS, S, M, L, XL,	4.3	987	629.65	1169.35	65%
XS, S, M, L, XL	3.9	987	314.65	584.35	65%
S, M, L, XL, XXL	4.4	986	559	1039.35	65%
:					Count: 39204

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".

Product ID		
11226634		
Brand Name	Price	Ratings
=VLOOKUP(Q10,Table1,{2,7,9},0)		

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)

Product ID		
11226634		
Brand Name	Price	Ratings
Maniac	1199	3.9

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

Product ID	6744434
Discount Amount (6744434)	899.4

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))

Product ID	6744434
Discount Amount (6744434)	=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.

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 basis of the product ID.

<https://www.linkedin.com/in/shlok-vashisht-307b9a15b/>



THANKYOU

