

myntra-data-analysis

March 24, 2024

```
[7]: import pandas as pd

df = pd.read_csv('products.csv')
df.head()
```

```
[7]:
```

	product_name	brand_name	rating	rating_count	\
0	Slim Fit Casual Shirt	Greenfibre	0.0	0	
1	Men Printed T-shirt	FREESOUL	0.0	0	
2	Leather Two Fold Wallet	Da Milano	0.0	0	
3	Pure Georgette Bandhani Saree	Mitera	3.6	5	
4	Men Geometric Printed Kurta	Indo Era	0.0	0	

	marked_price	discounted_price	sizes	\
0	1699	1699	38,39,40,42,44	
1	799	799	S,M,L,XL,XXL	
2	4499	4499	Onesize	
3	6727	1681	Onesize	
4	2599	1429	S,M,L,XL,XXL	

	product_link	\
0	shirts/greenfibre/greenfibre-men-yellow-slim-f...	
1	tshirts/freesoul/freesoul-men-white-printed-co...	
2	wallets/da-milano/da-milano-women-blue-texture...	
3	sarees/mitera/mitera-women-sea-green--white-em...	
4	kurtas/indo-era/indo-era-men-lavender-geometri...	

	img_link	product_tag	brand_tag	\
0	https://assets.myntassets.com/dpr_2,q_60,w_210...	shirts	greenfibre	
1	https://assets.myntassets.com/dpr_2,q_60,w_210...	tshirts	freesoul	
2	https://assets.myntassets.com/dpr_2,q_60,w_210...	wallets	da-milano	
3	https://assets.myntassets.com/dpr_2,q_60,w_210...	sarees	mitera	
4	https://assets.myntassets.com/dpr_2,q_60,w_210...	kurtas	indo-era	

	discount_amount	discount_percent	\
0	0	0	
1	0	0	
2	0	0	

3	5046	75
4	1170	45

	tags \
0	clothing,men-clothing,shirts,greenfibre-shirts
1	clothing,men-clothing,tshirts,freesoul-tshirts
2	accessories,women-accessories,wallets,da-milan...
3	clothing,women-clothing,sarees,mitera-sarees
4	clothing,men-clothing,kurtas,indo-era-kurtas

	seller_address \
0	25-28, Pariseema Complex,Nr. Lal Bunglow Cross...
1	21-E1,2ND PHASE, PEENYA INDUSTRIAL AREA,Bengal...
2	85, Kapashera Bijwasan Road, New DelhiNew Delh...
3	7007, Hindustan Textile Centre-2,Dumbhal,Surat...
4	4TH FLOOR, Block No B-4021/4022/4023/4026/4038...

	meta_description \
0	Buy Greenfibre Men Yellow Slim Fit Checked Cas...
1	Buy FREESOUL Men White Printed Cotton T Shirt ...
2	Buy Da Milano Women Blue Textured Leather Two ...
3	Buy Mitera Women Sea Green & White Embroidered...
4	Buy Indo Era Men Lavender Geometric Printed Ku...

	meta_keyword
0	Greenfibre Men Yellow Slim Fit Checked Casual ...
1	FREESOUL Men White Printed Cotton T Shirt, FR...
2	Da Milano Women Blue Textured Leather Two Fold...
3	Mitera Women Sea Green & White Embroidered Pur...
4	Indo Era Men Lavender Geometric Printed Kurta,...

0.0.1 1. Checking Null Values

```
[11]: df.isnull().sum()
```

```
[11]: product_name      0
      brand_name        0
      rating            0
      rating_count      0
      marked_price      0
      discounted_price  0
      sizes             0
      product_link      0
      img_link          0
      product_tag       0
      brand_tag         0
      discount_amount   0
```

```
discount_percent    0
tags                0
seller_address      0
meta_description    0
meta_keyword        0
dtype: int64
```

0.0.2 2. Remove Columns that are not in use

```
[13]: del df['product_link']
del df['img_link']
del df['meta_keyword']
del df['meta_description']

df.head()
```

```
[13]:
```

	product_name	brand_name	rating	rating_count	\
0	Slim Fit Casual Shirt	Greenfibre	0.0	0	
1	Men Printed T-shirt	FREESOUL	0.0	0	
2	Leather Two Fold Wallet	Da Milano	0.0	0	
3	Pure Georgette Bandhani Saree	Mitera	3.6	5	
4	Men Geometric Printed Kurta	Indo Era	0.0	0	

	marked_price	discounted_price	sizes	product_tag	brand_tag	\
0	1699	1699	38,39,40,42,44	shirts	greenfibre	
1	799	799	S,M,L,XL,XXL	tshirts	freesoul	
2	4499	4499	Onesize	wallets	da-milano	
3	6727	1681	Onesize	sarees	mitera	
4	2599	1429	S,M,L,XL,XXL	kurtas	indo-era	

	discount_amount	discount_percent	\
0	0	0	
1	0	0	
2	0	0	
3	5046	75	
4	1170	45	

	tags	\
0	clothing,men-clothing,shirts,greenfibre-shirts	
1	clothing,men-clothing,tshirts,freesoul-tshirts	
2	accessories,women-accessories,wallets,da-milan...	
3	clothing,women-clothing,sarees,mitera-sarees	
4	clothing,men-clothing,kurtas,indo-era-kurtas	

	seller_address
0	25-28, Pariseema Complex,Nr. Lal Bunglow Cross...
1	21-E1,2ND PHASE, PEENYA INDUSTRIAL AREA,Bengal...

```

2 85, Kapashera Bijwasan Road, New DelhiNew Delh...
3 7007, Hindustan Textile Centre-2,Dumbhal,Surat...
4 4TH FLOOR, Block No B-4021/4022/4023/4026/4038...

```

0.0.3 3. Data Analysis - Numeric Column

Average rating of a product (non zero)

```

[27]: rat = 0
      c = 0

      for i in df['rating']:

          if(i != 0.0):
              rat += i
              c += 1

      print(round(int(rat) / c, 2))

```

4.16

3.2) Most Expensive Product (Marked Price)

```

[41]: df[df['marked_price'] == df['marked_price'].max()]

```

```

[41]:
      product_name brand_name  rating \
4565  Men Leather Straps Automatic Motion Watch  EARNSHAW      0.0

      rating_count  marked_price  discounted_price  sizes product_tag \
4565              0          44950           26970  Onesize    watches

      brand_tag  discount_amount  discount_percent \
4565  earnshaw           17980             40

      tags \
4565  accessories,men-accessories,watches,earnshaw-w...

      seller_address
4565  NO.1, 3RD FLOOR, MARUTHI COMPLEX, R.T.NAGAR MA...

```

3.3) Most Expensive Product (Discounted Price)

```

[43]: df[df['discounted_price'] == df['discounted_price'].max()]

```

```

[43]:
      product_name brand_name  rating  rating_count  marked_price \
4458  Women Aviator Sunglasses  Tom Ford      0.0           0      40900

      discounted_price  sizes product_tag brand_tag  discount_amount \

```

```

4458          40900      L  sunglasses  tom-ford          0

discount_percent                                     tags \
4458          0  accessories,women-accessories,sunglasses,tom-f...

seller_address
4458  Not Mentioned

```

3.4) Most Discounted Product (Percentage)

```
[49]: df[df['discount_percent'] == df['discount_percent'].max()]
```

```

[49]:
      product_name brand_name  rating  rating_count \
4157  Geometric Block Print Saree  HRITIKA    4.2         6
4701      Poly Chiffon Saree  HRITIKA    0.0         0
6106              Saree  HRITIKA    0.0         0
8164      Poly Chiffon Saree  HRITIKA    0.0         0
8212      Poly Chiffon Saree  HRITIKA    0.0         0
12938              Saree  HRITIKA    0.0         0
13005              Saree  HRITIKA    0.0         0
18226  Paisley Block Print Saree  HRITIKA    0.0         0

      marked_price  discounted_price  sizes product_tag brand_tag \
4157          6099           609  Onesize    sarees  hritika
4701          7990           799  Onesize    sarees  hritika
6106          7990           799  Onesize    sarees  hritika
8164          6599           659  Onesize    sarees  hritika
8212          6599           659  Onesize    sarees  hritika
12938          6599           659  Onesize    sarees  hritika
13005          6599           659  Onesize    sarees  hritika
18226          6099           609  Onesize    sarees  hritika

      discount_amount  discount_percent \
4157          5490           90
4701          7191           90
6106          7191           90
8164          5940           90
8212          5940           90
12938          5940           90
13005          5940           90
18226          5490           90

      tags \
4157  clothing,women-clothing,sarees,hritika-sarees
4701  clothing,women-clothing,sarees,hritika-sarees
6106  clothing,women-clothing,sarees,hritika-sarees
8164  clothing,women-clothing,sarees,hritika-sarees

```

```

8212 clothing,women-clothing,sarees,hritika-sarees
12938 clothing,women-clothing,sarees,hritika-sarees
13005 clothing,women-clothing,sarees,hritika-sarees
18226 clothing,women-clothing,sarees,hritika-sarees

```

```

seller_address
4157 5005, RADHA RAMAN TEXTILE MARKET, NR. RKPL MAR...
4701 5005, RADHA RAMAN TEXTILE MARKET, NR. RKPL MAR...
6106 5005, RADHA RAMAN TEXTILE MARKET, NR. RKPL MAR...
8164 5005, RADHA RAMAN TEXTILE MARKET, NR. RKPL MAR...
8212 5005, RADHA RAMAN TEXTILE MARKET, NR. RKPL MAR...
12938 5005, RADHA RAMAN TEXTILE MARKET, NR. RKPL MAR...
13005 5005, RADHA RAMAN TEXTILE MARKET, NR. RKPL MAR...
18226 5005, RADHA RAMAN TEXTILE MARKET, NR. RKPL MAR...

```

3.5) Most Discounted Product (Disount Amount)

```
[52]: df[df['discount_amount'] == df['discount_amount'].max()]
```

```

[52]:
product_name brand_name rating \
6214 Textured 360-Degree Rotation Hard-Sided Trolle... Safari 0.0

rating_count marked_price discounted_price sizes product_tag \
6214 0 32997 9239 Pack trolley-bag

brand_tag discount_amount discount_percent \
6214 safari 23758 72

tags \
6214 accessories,unisex-accessories,trolley-bag,saf...

seller_address
6214 SAGAR COMPLEX, BULDING NO.F1, GODOWN NO.1&2,OP...

```

3.6) Shortlisting Products based on specific filters

```
[70]: df_pr = df[df['discounted_price'] > 2000]
df_pr = df_pr[df_pr['discounted_price'] < 3000]
df_pr = df_pr[df_pr['product_tag'] == 'handbags']

df_pr.sort_values(by = 'discount_percent', ascending = False).head(10)
```

```

[70]:
product_name brand_name rating rating_count \
3650 Women Textured Handheld Bag Lino Perros 4.2 37
458 Textured Shoulder Bag GENWAYNE 4.5 56
2625 PU Structured Handheld Bag CORSICA 4.5 11

```

10598	PU Oversized Shopper Tote Bag	Caprese	4.6	88
2866	Solid Structured Shoulder Bag	Caprese	0.0	0
10716	Sustainable IBIZA Handheld Bag	MANGO	0.0	0
3256	PU Structured Handheld Bag	Caprese	4.6	124
23823	Leather Textured Sling Bag	Hidesign	4.6	44
6528	Solid Shoulder Bag	Accessorize	5.0	4
18539	Structured Shoulder Bag	Caprese	0.0	0

	marked_price	discounted_price	sizes	product_tag	brand_tag \
3650	4495	2022	Onesize	handbags	lino-perros
458	4500	2160	M	handbags	genwayne
2625	4299	2149	Onesize	handbags	corsica
10598	4599	2299	Onesize	handbags	caprese
2866	5999	2999	Onesize	handbags	caprese
10716	5590	2795	Onesize	handbags	mango
3256	4399	2199	Onesize	handbags	caprese
23823	4195	2097	Onesize	handbags	hidesign
6528	4995	2597	Onesize	handbags	accessorize
18539	5399	2969	Onesize	handbags	caprese

	discount_amount	discount_percent \
3650	2473	55
458	2340	52
2625	2150	50
10598	2300	50
2866	3000	50
10716	2795	50
3256	2200	50
23823	2098	50
6528	2398	48
18539	2430	45

	tags \
3650	accessories,women-accessories,handbags,lino-pe...
458	accessories,women-accessories,handbags,genwayn...
2625	accessories,women-accessories,handbags,corsica...
10598	accessories,women-accessories,handbags,caprese...
2866	accessories,women-accessories,handbags,caprese...
10716	accessories,women-accessories,handbags,mango-h...
3256	accessories,women-accessories,handbags,caprese...
23823	accessories,women-accessories,handbags,hidesig...
6528	accessories,women-accessories,handbags,accessor...
18539	accessories,women-accessories,handbags,caprese...

	seller_address
3650	PLOT NO 324 UDYOG VIHAR PHASE 4 GURGAONGurugra...
458	679 Madhyamgram Bazar,Malamba Road, Madhyamgra...

```

2625    Second Floor, Plot No. 82, Okhla Industrial Es...
10598                                     Not Mentioned
2866    5A/19,Ground Floor ( Front Portion),Ansari Rao...
10716    406, B2 and C, 4th Floor,Centrum Plaza, Golf C...
3256                                     Not Mentioned
23823    Awfis, 7th floor, HM Vibha Tower, 25 Laskar Ho...
6528                                     Not Mentioned
18539    5A/19,Ground Floor ( Front Portion),Ansari Rao...

```

3.7) Finding all the brand names

```
[76]: df['brand_name'].unique()
```

```
[76]: array(['Greenfibre', 'FREESOUL', 'Da Milano', ..., 'FEVER', 'Oxyglow',
          'RARE ROOTS'], dtype=object)
```

3.8) Finding the top 10 Brands with most number of products

```
[89]: data = []

for brand_name in df['brand_name'].unique():

    c = 0
    for i in df['brand_name']:
        if (i == brand_name):
            c += 1

    data.append([brand_name, c])

data = pd.DataFrame(data, columns = ['name', 'products'])

data.sort_values(by = 'products', ascending = False).head(10)
```

```
[89]:
```

	name	products
69	Roadster	633
16	Mast & Harbour	369
11	DressBerry	311
136	HERE&NOW	298
32	max	289
9	HRX by Hrithik Roshan	288
19	Anouk	280
78	H&M	278
88	URBANIC	234
74	MANGO	230

3.9) Finding the top 10 Most Expensive Brands (Marked Price)


```
[98]: df.groupby('brand_name').mean()['marked_price'].sort_values(ascending = False).  
      ↪head(10)
```

```
[98]: brand_name  
      EARNSHAW      38950.0  
      Tom Ford      35900.0  
      dyson         33900.0  
      Cerruti 1881   29999.0  
      TISSOT        29250.0  
      SEIKO         24250.0  
      Emporio Armani 23495.0  
      ROMEING       22140.0  
      Safari        21218.5  
      VRSALES       20803.0  
      Name: marked_price, dtype: float64
```

```
[101]: df.groupby('brand_name').mean()['discount_percent'].sort_values(ascending =  
      ↪False).head(10)
```

```
[101]: brand_name  
      HRITIKA      89.454545  
      THE AAB STUDIO 89.000000  
      OOMPH         87.500000  
      Brandsoon     86.000000  
      HAMT          86.000000  
      VAIVIDHYAM    86.000000  
      Kord Store    85.250000  
      VAIRAGEE      85.000000  
      Dhrohar       84.000000  
      awesome       84.000000  
      Name: discount_percent, dtype: float64
```

```
[102]: df.groupby('brand_name').mean()['discount_percent'].sort_values(ascending =  
      ↪True).head(10)
```

```
[102]: brand_name  
      M.A.C         0.0  
      METTLE        0.0  
      Born Babies   0.0  
      Bonjour       0.0  
      MICHAEL ANGELO 0.0  
      MISS NAILS    0.0  
      Bodycare      0.0  
      MOCEMSA       0.0  
      METAL         0.0  
      Bobbi Brown   0.0  
      Name: discount_percent, dtype: float64
```

```
[110]: df_pr = df[df['rating'] != 0]

df_pr.groupby('brand_name').mean()['rating'].sort_values(ascending = False)
```

```
[110]: brand_name
London Rag      5.0
KREDO           5.0
UCLA            5.0
SHIRT THEORY    5.0
Apratim         5.0
...
DECHEN          2.0
Kaifoo          1.5
honeybud        1.5
ZEBRONICS       1.0
Lovely Lady     1.0
Name: rating, Length: 1458, dtype: float64
```

```
[111]: df_pr = df[df['rating_count'] != 0]

df_pr.groupby('brand_name').mean()['rating_count'].sort_values(ascending =
↪False)
```

```
[111]: brand_name
Envy            9000.000000
JAGUAR          6953.666667
Plum            5106.866667
Denver         5084.500000
Maybelline     5058.545455
...
ABROS           2.000000
UCLA            2.000000
Lovely Lady     1.000000
Apratim         1.000000
Hugo Boss       1.000000
Name: rating_count, Length: 1458, dtype: float64
```

```
[142]: import numpy as np

pincode = []
state    = []

for i in df['seller_address']:

    if (i != 'Not Mentioned'):

        pincode.append(i.split(' ')[-1])
```

```

        state.append(i.split(',')[ -1].split('-')[0].strip())

    else:

        pincode.append(np.nan)
        state.append(np.nan)

df['pincode'] = pincode
df['state'] = state

```

```
[143]: df.head()
```

```
[143]:
```

	product_name	brand_name	rating	rating_count	\
0	Slim Fit Casual Shirt	Greenfibre	0.0	0	
1	Men Printed T-shirt	FREESOUL	0.0	0	
2	Leather Two Fold Wallet	Da Milano	0.0	0	
3	Pure Georgette Bandhani Saree	Mitera	3.6	5	
4	Men Geometric Printed Kurta	Indo Era	0.0	0	

	marked_price	discounted_price	sizes	product_tag	brand_tag	\
0	1699	1699	38,39,40,42,44	shirts	greenfibre	
1	799	799	S,M,L,XL,XXL	tshirts	freesoul	
2	4499	4499	Onesize	wallets	da-milano	
3	6727	1681	Onesize	sarees	mitera	
4	2599	1429	S,M,L,XL,XXL	kurtas	indo-era	

	discount_amount	discount_percent	\
0	0	0	
1	0	0	
2	0	0	
3	5046	75	
4	1170	45	

	tags	\
0	clothing,men-clothing,shirts,greenfibre-shirts	
1	clothing,men-clothing,tshirts,freesoul-tshirts	
2	accessories,women-accessories,wallets,da-milan...	
3	clothing,women-clothing,sarees,mitera-sarees	
4	clothing,men-clothing,kurtas,indo-era-kurtas	

	seller_address	pincode	state
0	25-28, Pariseema Complex,Nr. Lal Bunglow Cross...	380006	Gujarat
1	21-E1,2ND PHASE, PEENYA INDUSTRIAL AREA,Bengal...	560058	Karnataka
2	85, Kapashera Bijwasan Road, New DelhiNew Delh...	110037	Delhi
3	7007, Hindustan Textile Centre-2,Dumbhal,Surat...	395010	Gujarat
4	4TH FLOOR, Block No B-4021/4022/4023/4026/4038...	395010	Gujarat

```
[144]: df.groupby('pincode').size().sort_values(ascending = False).head(5)
```

```
[144]: pincode
      110020      2025
      110001      1445
      560044      1078
      110002       966
      122102       889
      dtype: int64
```

```
[145]: df.groupby('state').size().sort_values(ascending = False).head(5)
```

```
[145]: state
      Delhi      8415
      Karnataka  3921
      Haryana    2609
      Maharashtra 2336
      Gujarat    1103
      dtype: int64
```

```
[ ]:
```