

```
In [2]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

import warnings
warnings.filterwarnings('ignore')
```

```
In [3]: df = pd.read_csv(r"C:\\Users\\mdmaz\\Downloads\\amazon.csv")
```

```
In [4]: df
```

Out[4]:

	product_id	product_name	category	discounted_price	actual_price	discount_perc
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories Accessories&Peripherals ...	₹399	₹1,099	
1	B098NS6PVG	Ambrane Unbreakable 60W / 3A Fast Charging 1.5...	Computers&Accessories Accessories&Peripherals ...	₹199	₹349	
2	B096MSW6CT	Sounce Fast Phone Charging Cable & Data Sync U...	Computers&Accessories Accessories&Peripherals ...	₹199	₹1,899	
3	B08HDJ86NZ	boAt Deuce USB 300 2 in 1 Type-C & Micro USB S...	Computers&Accessories Accessories&Peripherals ...	₹329	₹699	
4	B08CF3B7N1	Portronics Konnect L 1.2M Fast Charging 3A 8 P...	Computers&Accessories Accessories&Peripherals ...	₹154	₹399	
...	...	...	...	...	...	...
1460	B08L7J3T31	Noir Aqua - 5pcs PP Spun Filter + 1 Spanner   ...	Home&Kitchen Kitchen&HomeAppliances WaterPurif...	₹379	₹919	
1461	B01M6453MB	Prestige Delight PRWO Electric Rice Cooker (1 ...	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	₹2,280	₹3,045	
1462	B009P2LIL4	Bajaj Majesty RX10 2000 Watts Heat Convector R...	Home&Kitchen Heating,Cooling&AirQuality RoomHe...	₹2,219	₹3,080	
1463	B00J5DYCCA	Havells Ventil Air DSP 230mm Exhaust Fan (Pist...	Home&Kitchen Heating,Cooling&AirQuality Fans E...	₹1,399	₹1,890	
1464	B01486F4G6	Borosil Jumbo 1000-Watt Grill Sandwich Maker (...	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	₹2,863	₹3,690	

1465 rows × 6 columns

```
In [5]: df.columns
```

```
Out[5]: Index(['product_id', 'product_name', 'category', 'discounted_price',
              'actual_price', 'discount_percentage', 'rating', 'rating_count',
              'about_product', 'user_id', 'user_name', 'review_id', 'review_title',
              'review_content', 'img_link', 'product_link'],
              dtype='object')
```

```
In [6]: df.head(5)
```

Out[6]:

	product_id	product_name	category	discounted_price	actual_price	discount_percentag
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories Accessories&Peripherals ...	₹399	₹1,099	64
1	B098NS6PVG	Ambrane Unbreakable 60W / 3A Fast Charging 1.5...	Computers&Accessories Accessories&Peripherals ...	₹199	₹349	43
2	B096MSW6CT	Source Fast Phone Charging Cable & Data Sync U...	Computers&Accessories Accessories&Peripherals ...	₹199	₹1,899	90
3	B08HDJ86NZ	boAt Deuce USB 300 2 in 1 Type-C & Micro USB S...	Computers&Accessories Accessories&Peripherals ...	₹329	₹699	53
4	B08CF3B7N1	Portronics Konnect L 1.2M Fast Charging 3A 8 P...	Computers&Accessories Accessories&Peripherals ...	₹154	₹399	61

In [7]:

df.tail()

Out[7]:

	product_id	product_name	category	discounted_price	actual_price	discount_perce
1460	B08L7J3T31	Noir Aqua - 5pcs PP Spun Filter + 1 Spanner   ...	Home&Kitchen Kitchen&HomeAppliances WaterPurif...	₹379	₹919	
1461	B01M6453MB	Prestige Delight PRWO Electric Rice Cooker (1 ...	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	₹2,280	₹3,045	
1462	B009P2LIL4	Bajaj Majesty RX10 2000 Watts Heat Convector R...	Home&Kitchen Heating,Cooling&AirQuality RoomHe...	₹2,219	₹3,080	
1463	B00J5DYCCA	Havells Ventil Air DSP 230mm Exhaust Fan (Pist...	Home&Kitchen Heating,Cooling&AirQuality Fans E...	₹1,399	₹1,890	
1464	B01486F4G6	Borosil Jumbo 1000-Watt Grill Sandwich Maker (...	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	₹2,863	₹3,690	

In [8]:

df.sample(2)

Out[8]:

	product_id	product_name	category	discounted_price	actual_price	discount_pe
1257	B08HDCWDXD	Eureka Forbes Active Clean 700 Watts Powerful ...	Home&Kitchen Kitchen&HomeAppliances Vacuum,Cle...	₹3,179	₹6,999	
1456	B0B9JZW1SQ	4 in 1 Handheld Electric Vegetable Cutter Set,...	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	₹498	₹1,200	

In [9]:

df.shape

Out[9]:

(1465, 16)

In [10]:

df.describe()

Out[10]:

	product_id	product_name	category	discounted_price	actual_price	discount_perce
count	1465	1465	1465	1465	1465	
unique	1351	1337	211	550	449	
top	B07JW9H4J1	Fire-Boltt Ninja Call Pro Plus 1.83" Smart Wat...	Computers&Accessories Accessories&Peripherals ...	₹199	₹999	
freq	3	5	233	53	120	

In [11]:

df.info()

<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 1465 entries, 0 to 1464  
Data columns (total 16 columns):  
# Column Non-Null Count Dtype  
--- -  
0 product\_id 1465 non-null object  
1 product\_name 1465 non-null object  
2 category 1465 non-null object  
3 discounted\_price 1465 non-null object  
4 actual\_price 1465 non-null object  
5 discount\_percentage 1465 non-null object  
6 rating 1465 non-null object  
7 rating\_count 1463 non-null object  
8 about\_product 1465 non-null object  
9 user\_id 1465 non-null object  
10 user\_name 1465 non-null object  
11 review\_id 1465 non-null object  
12 review\_title 1465 non-null object  
13 review\_content 1465 non-null object  
14 img\_link 1465 non-null object  
15 product\_link 1465 non-null object  
dtypes: object(16)  
memory usage: 183.3+ KB

In [12]:

df.duplicated().sum()

Out[12]:

0

In [13]:

df.drop\_duplicates().shape

Out[13]:

(1465, 16)

In [14]:

df.dtypes

Out[14]:

product\_id object  
product\_name object  
category object  
discounted\_price object  
actual\_price object  
discount\_percentage object  
rating object  
rating\_count object  
about\_product object  
user\_id object  
user\_name object  
review\_id object  
review\_title object  
review\_content object  
img\_link object  
product\_link object  
dtype: object

In [15]:

df['actual\_price'].dtypes

Out[15]:

dtype('O')

In [16]:

df.actual\_price.str

Out[16]:

<pandas.core.strings.accessor.StringMethods at 0x229354f0b30>

In [17]:

df.actual\_price.str.isnumeric()

```
Out[17]: 0      False
          1      False
          2      False
          3      False
          4      False
          ...
        1460    False
        1461    False
        1462    False
        1463    False
        1464    False
        Name: actual_price, Length: 1465, dtype: bool
```

```
In [18]: df[~df.actual_price.str.isnumeric()]
```

Out[18]:

	product_id	product_name	category	discounted_price	actual_price	discount_perc
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories Accessories&Peripherals ...	₹399	₹1,099	
1	B098NS6PVG	Ambrane Unbreakable 60W / 3A Fast Charging 1.5...	Computers&Accessories Accessories&Peripherals ...	₹199	₹349	
2	B096MSW6CT	Source Fast Phone Charging Cable & Data Sync U...	Computers&Accessories Accessories&Peripherals ...	₹199	₹1,899	
3	B08HDJ86NZ	boAt Deuce USB 300 2 in 1 Type-C & Micro USB S...	Computers&Accessories Accessories&Peripherals ...	₹329	₹699	
4	B08CF3B7N1	Portronics Konnect L 1.2M Fast Charging 3A 8 P...	Computers&Accessories Accessories&Peripherals ...	₹154	₹399	
...	...	...	...	...	...	
1460	B08L7J3T31	Noir Aqua - 5pcs PP Spun Filter + 1 Spanner   ...	Home&Kitchen Kitchen&HomeAppliances WaterPurif...	₹379	₹919	
1461	B01M6453MB	Prestige Delight PRWO Electric Rice Cooker (1 ...	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	₹2,280	₹3,045	
1462	B009P2LIL4	Bajaj Majesty RX10 2000 Watts Heat Convector R...	Home&Kitchen Heating,Cooling&AirQuality RoomHe...	₹2,219	₹3,080	
1463	B00J5DYCCA	Havells Ventil Air DSP 230mm Exhaust Fan (Pist...	Home&Kitchen Heating,Cooling&AirQuality Fans E...	₹1,399	₹1,890	
1464	B01486F4G6	Borosil Jumbo 1000-Watt Grill Sandwich Maker (...	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	₹2,863	₹3,690	

1465 rows × 16 columns

```
In [19]: df_copy = df.copy()
df_copy
```

Out[19]:

	product_id	product_name	category	discounted_price	actual_price	discount_perc
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories Accessories&Peripherals ...	₹399	₹1,099	
1	B098NS6PVG	Ambrane Unbreakable 60W / 3A Fast Charging 1.5...	Computers&Accessories Accessories&Peripherals ...	₹199	₹349	
2	B096MSW6CT	Sounce Fast Phone Charging Cable & Data Sync U...	Computers&Accessories Accessories&Peripherals ...	₹199	₹1,899	
3	B08HDJ86NZ	boAt Deuce USB 300 2 in 1 Type-C & Micro USB S...	Computers&Accessories Accessories&Peripherals ...	₹329	₹699	
4	B08CF3B7N1	Portronics Konnect L 1.2M Fast Charging 3A 8 P...	Computers&Accessories Accessories&Peripherals ...	₹154	₹399	
...	...	...	...	...	...	
1460	B08L7J3T31	Noir Aqua - 5pcs PP Spun Filter + 1 Spanner   ...	Home&Kitchen Kitchen&HomeAppliances WaterPurif...	₹379	₹919	
1461	B01M6453MB	Prestige Delight PRWO Electric Rice Cooker (1 ...	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	₹2,280	₹3,045	
1462	B009P2LIL4	Bajaj Majesty RX10 2000 Watts Heat Convector R...	Home&Kitchen Heating,Cooling&AirQuality RoomHe...	₹2,219	₹3,080	
1463	B00J5DYCCA	Havells Ventil Air DSP 230mm Exhaust Fan (Pist...	Home&Kitchen Heating,Cooling&AirQuality Fans E...	₹1,399	₹1,890	
1464	B01486F4G6	Borosil Jumbo 1000-Watt Grill Sandwich Maker (...	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	₹2,863	₹3,690	

1465 rows × 16 columns

```
In [20]: df_copy.reset_index(drop = True,inplace = True)
```

```
In [21]: df_copy
```

Out [21]:

	product_id	product_name	category	discounted_price	actual_price	discount_perc
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories Accessories&Peripherals ...	₹399	₹1,099	
1	B098NS6PVG	Ambrane Unbreakable 60W / 3A Fast Charging 1.5...	Computers&Accessories Accessories&Peripherals ...	₹199	₹349	
2	B096MSW6CT	Sounce Fast Phone Charging Cable & Data Sync U...	Computers&Accessories Accessories&Peripherals ...	₹199	₹1,899	
3	B08HDJ86NZ	boAt Deuce USB 300 2 in 1 Type-C & Micro USB S...	Computers&Accessories Accessories&Peripherals ...	₹329	₹699	
4	B08CF3B7N1	Portronics Konnect L 1.2M Fast Charging 3A 8 P...	Computers&Accessories Accessories&Peripherals ...	₹154	₹399	
...	...	...	...	...	...	...
1460	B08L7J3T31	Noir Aqua - 5pcs PP Spun Filter + 1 Spanner   ...	Home&Kitchen Kitchen&HomeAppliances WaterPurif...	₹379	₹919	
1461	B01M6453MB	Prestige Delight PRWO Electric Rice Cooker (1 ...	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	₹2,280	₹3,045	
1462	B009P2LIL4	Bajaj Majesty RX10 2000 Watts Heat Convector R...	Home&Kitchen Heating,Cooling&AirQuality RoomHe...	₹2,219	₹3,080	
1463	B00J5DYCCA	Havells Ventil Air DSP 230mm Exhaust Fan (Pist...	Home&Kitchen Heating,Cooling&AirQuality Fans E...	₹1,399	₹1,890	
1464	B01486F4G6	Borosil Jumbo 1000-Watt Grill Sandwich Maker (...	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	₹2,863	₹3,690	

1465 rows × 16 columns

⏪		⏩
---	--	---

In [22]:

```
pd.isnull(df)
```

Out [22]:

	product_id	product_name	category	discounted_price	actual_price	discount_percentage	rating	rating_count	about_produc
0	False	False	False	False	False	False	False	False	Fals
1	False	False	False	False	False	False	False	False	Fals
2	False	False	False	False	False	False	False	False	Fals
3	False	False	False	False	False	False	False	False	Fals
4	False	False	False	False	False	False	False	False	Fals
...	...	...	...	...	...	...	...	...	.
1460	False	False	False	False	False	False	False	False	Fals
1461	False	False	False	False	False	False	False	False	Fals
1462	False	False	False	False	False	False	False	False	Fals
1463	False	False	False	False	False	False	False	False	Fals
1464	False	False	False	False	False	False	False	False	Fals

1465 rows × 16 columns

⏪		⏩
---	--	---

In [23]:

```
pd.isnull(df).sum()
```

```
Out[23]: product_id      0
product_name    0
category        0
discounted_price 0
actual_price    0
discount_percentage 0
rating          0
rating_count    2
about_product   0
user_id         0
user_name       0
review_id       0
review_title    0
review_content  0
img_link        0
product_link    0
dtype: int64

In [24]: df.describe(include = 'object')

Out[24]:
```

	product_id	product_name	category	discounted_price	actual_price	discount_perce
count	1465	1465	1465	1465	1465	
unique	1351	1337	211	550	449	
top	B07JW9H4J1	Fire-Boltt Ninja Call Pro Plus 1.83" Smart Wat...	Computers&Accessories Accessories&Peripherals ...	₹199	₹999	
freq	3	5	233	53	120	

```


In [25]: df['actual_price'].dtype

Out[25]: dtype('O')

In [26]: df['actual_price'] = df['actual_price'].astype("str")

In [27]: df['actual_price']

Out[27]: 0      ₹1,099
1      ₹349
2      ₹1,899
3      ₹699
4      ₹399
...
1460   ₹919
1461   ₹3,045
1462   ₹3,080
1463   ₹1,890
1464   ₹3,690
Name: actual_price, Length: 1465, dtype: object

In [28]: df["actual_price"] = df["actual_price"].replace({"₹": "", ",": "", " ": ""}, regex=True)
df["actual_price"] = pd.to_numeric(df["actual_price"], errors="coerce")

In [29]: df['actual_price']

Out[29]: 0      1099.0
1      349.0
2      1899.0
3      699.0
4      399.0
...
1460    919.0
1461   3045.0
1462   3080.0
1463   1890.0
1464   3690.0
Name: actual_price, Length: 1465, dtype: float64

In [30]: df['discount_percentage'] = df['discount_percentage'].replace("%", " ")

In [31]: df['discount_percentage']= pd.to_numeric(df['discount_percentage'], errors="coerce")

In [32]: df['discount_percentage']
```

```
Out[32]: 0      NaN
1      NaN
2      NaN
3      NaN
4      NaN
...
1460   NaN
1461   NaN
1462   NaN
1463   NaN
1464   NaN
Name: discount_percentage, Length: 1465, dtype: float64
```

```
In [37]: grouped_data = df.groupby('actual_price')['discount_percentage'].sum()
result = grouped_data.reset_index()

plt.plot(result['actual_price'], result['discount_percentage'], color='red', marker="o")
plt.xlabel("Actual Price", color="blue", fontsize=20)
plt.ylabel("Total Discount Percentage", color="green", fontsize=20)
plt.title("Discount vs Actual Price")
plt.grid(True)
plt.show()
```



```
In [38]: df['rating'] = df['rating'].replace(".", " ")
df['rating'] = pd.to_numeric(df['rating'], errors="coerce")
```

```
In [39]: df['rating']
```

```
Out[39]: 0      4.2
1      4.0
2      3.9
3      4.2
4      4.2
...
1460   4.0
1461   4.1
1462   3.6
1463   4.0
1464   4.3
Name: rating, Length: 1465, dtype: float64
```

```
In [40]: df['rating_count'] = df['rating_count'].replace(".", " ")
df['rating_count'] = pd.to_numeric(df['rating_count'], errors="coerce")
```

```
In [41]: df['rating_count']
```



```
Out[41]: 0      NaN
         1      NaN
         2      NaN
         3      NaN
         4      NaN
         ...
        1460    NaN
        1461    NaN
        1462    468.0
        1463    NaN
        1464    NaN
        Name: rating_count, Length: 1465, dtype: float64
```

```
In [42]: df['about_product'] = df['about_product'].replace("int", " ")
df['about_product'].astype("str")
df['about_product'] = pd.to_numeric(df['about_product'], errors="coerce")
```

```
In [43]: df['about_product']
```

```
Out[43]: 0      NaN
         1      NaN
         2      NaN
         3      NaN
         4      NaN
         ..
        1460    NaN
        1461    NaN
        1462    NaN
        1463    NaN
        1464    NaN
        Name: about_product, Length: 1465, dtype: float64
```

```
In [44]: df.dtypes
```

```
Out[44]: product_id      object
product_name      object
category          object
discounted_price   object
actual_price      float64
discount_percentage float64
rating            float64
rating_count      float64
about_product     float64
user_id          object
user_name        object
review_id        object
review_title     object
review_content   object
img_link         object
product_link     object
dtype: object
```

```
In [45]: df['product_id']
```

```
Out[45]: 0      B07JW9H4J1
         1      B098NS6PVG
         2      B096MSW6CT
         3      B08HDJ86NZ
         4      B08CF3B7N1
         ...
        1460    B08L7J3T31
        1461    B01M6453MB
        1462    B009P2LIL4
        1463    B00J5DYCCA
        1464    B01486F4G6
        Name: product_id, Length: 1465, dtype: object
```

```
In [46]: df['product_id'].str.strip().astype("str")
df['product_id'].apply(str)
df["product_id"] = df["product_id"].astype("string").fillna("")
```

```
In [47]: df['product_name']
```

```
Out[47]: 0      Wayona Nylon Braided USB to Lightning Fast Cha...
        1      Ambrane Unbreakable 60W / 3A Fast Charging 1.5...
        2      Sounce Fast Phone Charging Cable & Data Sync U...
        3      boAt Deuce USB 300 2 in 1 Type-C & Micro USB S...
        4      Portronics Konnect L 1.2M Fast Charging 3A 8 P...
        ...
        1460   Noir Aqua - 5pcs PP Spun Filter + 1 Spanner | ...
        1461   Prestige Delight PRW0 Electric Rice Cooker (1 ...
        1462   Bajaj Majesty RX10 2000 Watts Heat Convector R...
        1463   Havells Ventil Air DSP 230mm Exhaust Fan (Pist...
        1464   Borosil Jumbo 1000-Watt Grill Sandwich Maker (...
Name: product_name, Length: 1465, dtype: object
```

```
In [48]: df["product_name"] = df["product_name"].astype("string")
```

```
In [49]: df['product_name']
```

```
Out[49]: 0      Wayona Nylon Braided USB to Lightning Fast Cha...
        1      Ambrane Unbreakable 60W / 3A Fast Charging 1.5...
        2      Sounce Fast Phone Charging Cable & Data Sync U...
        3      boAt Deuce USB 300 2 in 1 Type-C & Micro USB S...
        4      Portronics Konnect L 1.2M Fast Charging 3A 8 P...
        ...
        1460   Noir Aqua - 5pcs PP Spun Filter + 1 Spanner | ...
        1461   Prestige Delight PRW0 Electric Rice Cooker (1 ...
        1462   Bajaj Majesty RX10 2000 Watts Heat Convector R...
        1463   Havells Ventil Air DSP 230mm Exhaust Fan (Pist...
        1464   Borosil Jumbo 1000-Watt Grill Sandwich Maker (...
Name: product_name, Length: 1465, dtype: string
```

```
In [50]: df["review_title"] = df["review_title"].astype(str).fillna("").str.strip()
```

```
In [51]: df['category']
```

```
Out[51]: 0      Computers&Accessories|Accessories&Peripherals|...
        1      Computers&Accessories|Accessories&Peripherals|...
        2      Computers&Accessories|Accessories&Peripherals|...
        3      Computers&Accessories|Accessories&Peripherals|...
        4      Computers&Accessories|Accessories&Peripherals|...
        ...
        1460   Home&Kitchen|Kitchen&HomeAppliances|WaterPurif...
        1461   Home&Kitchen|Kitchen&HomeAppliances|SmallKitch...
        1462   Home&Kitchen|Heating,Cooling&AirQuality|RoomHe...
        1463   Home&Kitchen|Heating,Cooling&AirQuality|Fans|E...
        1464   Home&Kitchen|Kitchen&HomeAppliances|SmallKitch...
Name: category, Length: 1465, dtype: object
```

```
In [50]: df["category"] = df["category"].str.split("|")
```

```
In [52]: df["category"] = df["category"].fillna("").astype(str).str.strip()
df["category"] = df["category"] .replace(",","")
```

```
In [53]: df.dtypes
```

```
Out[53]: product_id      string[python]
product_name      string[python]
category           object
discounted_price   object
actual_price       float64
discount_percentage float64
rating             float64
rating_count       float64
about_product      float64
user_id            object
user_name          object
review_id          object
review_title       object
review_content     object
img_link           object
product_link       object
dtype: object
```

```
In [54]: df['discounted_price']
```

```
Out[54]: 0      ₹399
          1      ₹199
          2      ₹199
          3      ₹329
          4      ₹154
          ...
          1460    ₹379
          1461    ₹2,280
          1462    ₹2,219
          1463    ₹1,399
          1464    ₹2,863
          Name: discounted_price, Length: 1465, dtype: object
```

```
In [71]: df['discounted_price'] = df['discounted_price'].str.replace("₹", " ", regex = True)
          df['discounted_price'] = df['discounted_price'].str.replace(",", " ", regex = True)
          df['discounted_price'] = df['discounted_price'].astype("str")
          df['discounted_price'] = pd.to_numeric(df['discounted_price'], errors='coerce')
```

```
In [72]: df['discounted_price']
```

```
Out[72]: 0      399.0
          1      199.0
          2      199.0
          3      329.0
          4      154.0
          ...
          1460    379.0
          1461      NaN
          1462      NaN
          1463      NaN
          1464      NaN
          Name: discounted_price, Length: 1465, dtype: float64
```

```
In [74]: df.dtypes
```

```
Out[74]: product_id      string[python]
          product_name    string[python]
          category        object
          discounted_price float64
          actual_price     float64
          discount_percentage float64
          rating           float64
          rating_count     float64
          about_product    float64
          user_id          object
          user_name        object
          review_id        object
          review_title     object
          review_content   object
          img_link         object
          product_link     object
          dtype: object
```

```
In [75]: df['user_id']
```

```
Out[75]: 0      AG3D604STAQKAY2UVGEUV46KN35Q,AHMY5CWJMMK5BJRBB...
          1      AECPFYFQVRUWC3KGNLJI0REFP5LQ,AGYYVPDD7YG7FYNBX...
          2      AGU3BBQ2V2DDAMOAKGFAWDDQ6QHA,AESFLDV2PT363T2AQ...
          3      AEWAZDZZJLQUYVOVGBEUKSLXHQ5A,AG5HTSFRRE6NL3M5S...
          4      AE3Q6KSUK5P75D5HIFYHCRAOL0DSA,AFUGIFH5ZAFXRDSZH...
          ...
          1460    AHITFY6AHAL0F0H0ZE0C6XBP4FEA,AFRABB0DZJZQB6Z4U...
          1461    AFG5FM3NEMOL6BNFRV2NK5FNJCHQ,AGEINTRN6Z563RMLH...
          1462    AGVPWCMAHYQWJ0QKMUJN4DW3KM5Q,AF4Q3E66MY4SR7YQZ...
          1463    AF2JQCLSCY3QJATWUNNHUSVUPNQ,AFDMLUXC5LS5RXDJS...
          1464    AFGW5PT3R6ZAVQR4Y5MMVAKBZAYA,AG7QNJ2SCS5VS5VYY...
          Name: user_id, Length: 1465, dtype: object
```

```
In [77]: df['user_id'] = df['user_id'].str.replace(","," ", regex = True)
          df['user_id'] = df['user_id'].astype("str")
          df['user_id'] = pd.to_numeric(df['user_id'], errors='coerce')
```

```
In [78]: df['user_id']
```

```
Out[78]: 0      NaN
          1      NaN
          2      NaN
          3      NaN
          4      NaN
          ..
        1460    NaN
        1461    NaN
        1462    NaN
        1463    NaN
        1464    NaN
        Name: user_id, Length: 1465, dtype: float64
```

```
In [79]: df.dtypes
```

```
Out[79]: product_id      string[python]
product_name    string[python]
category                object
discounted_price    float64
actual_price        float64
discount_percentage  float64
rating              float64
rating_count        float64
about_product       float64
user_id             float64
user_name            object
review_id            object
review_title         object
review_content       object
img_link             object
product_link         object
dtype: object
```

```
In [80]: df['user_name']
```

```
Out[80]: 0      Manav,Adarsh gupta,Sundeep,S.Sayeed Ahmed,jasp...
          1      ArdKn,Nirbhay kumar,Sagar Viswanathan,Asp,Plac...
          2      Kunal,Himanshu,viswanath,sai niharka,saqib mal...
          3      Omkar dhale,JD,HEMALATHA,Ajwadh a.,amar singh ...
          4      rahuls6099,Swasat Borah,Ajay Wadke,Pranali,RVK...
          ...
        1460    Prabha ds,Raghuram bk,Real Deal,Amazon Custome...
        1461    Manu Bhai,Naveenpittu,Evatira Sangma,JAGANNADH...
        1462    Nehal Desai,Danish Parwez,Amazon Customer,Amaz...
        1463    Shubham Dubey,E.GURUBARAN,Mayank S.,eusuf khan...
        1464    Rajib,Ajay B,Vikas Kahol,PARDEEP,Anindya Prama...
        Name: user_name, Length: 1465, dtype: object
```

```
In [83]: df['user_name'] = df['user_name'].str.replace(","," ",regex = True)
df['user_name'] = df['user_name'].str.replace(".", "", regex = True)
df['user_name'] = df['user_name'].astype("str")
df['user_name'] = pd.to_numeric(df['user_name'], errors = 'coerce')
```

```
In [84]: df['user_name']
```

```
Out[84]: 0      NaN
          1      NaN
          2      NaN
          3      NaN
          4      NaN
          ..
        1460    NaN
        1461    NaN
        1462    NaN
        1463    NaN
        1464    NaN
        Name: user_name, Length: 1465, dtype: float64
```

```
In [85]: df.dtypes
```

```
Out[85]: product_id      string[python]
product_name    string[python]
category        object
discounted_price float64
actual_price     float64
discount_percentage float64
rating           float64
rating_count     float64
about_product   float64
user_id         float64
user_name       float64
review_id       object
review_title    object
review_content  object
img_link        object
product_link    object
dtype: object
```

```
In [86]: df['review_id']
```

```
Out[86]: 0      R3HXWT0LRP0NMF,R2AJM3LFTLZHF0,R6AQJGUP6P86,R1K...
1      RGIQEG07R9HS2,R1SMWZQ86XIN8U,R2J3Y1WL29GWDE,RY...
2      R3J3EQQ9TZI5ZJ,R3E7WBGK7ID0KV,RWU79XKQ6I1QF,R2...
3      R3EEUZKKK9J36I,R3HJVYCLY0Y554,REDECAZ7AMPQC,R1...
4      R1BP4L2HH9TFUP,R16PVJEXKV6QZS,R2UPDB81N66T4P,R...
...
1460    R3G3XFHPBFF0E8,R3C0BZCD32EIGW,R2EBVBCN9QPD9R,R...
1461    R3DDL2UPKQ2CK9,R2SYYU10ATVIU5,R1VM993161IYRW,R...
1462    R1TLRJW4STY5I,R20455KRN493R1,R3Q5MVBGBRIAS2G,R...
1463    R39Q2Y79MM9SWK,R3079BG1NIH6MB,R29A31ZELTZNJM,R...
1464    R20RBRZ0WEUJT9,R0KIFK9R2ISSE,R30EEG2FNJSN5I,R2...
Name: review_id, Length: 1465, dtype: object
```

```
In [89]: df['review_id'] = df['review_id'].str.replace(","," ", regex = True)
df['review_id'] = df['review_id'].astype("str")
df['review_id'] = pd.to_numeric(df['review_id'], errors = 'coerce')
```

```
In [91]: df.dtypes
```

```
Out[91]: product_id      string[python]
product_name    string[python]
category        object
discounted_price float64
actual_price     float64
discount_percentage float64
rating           float64
rating_count     float64
about_product   float64
user_id         float64
user_name       float64
review_id       float64
review_title    object
review_content  object
img_link        object
product_link    object
dtype: object
```

```
In [92]: df['review_title']
```

```
Out[92]: 0      Satisfied,Charging is really fast,Value for mo...
1      A Good Braided Cable for Your Type C Device,Go...
2      Good speed for earlier versions,Good Product,W...
3      Good product,Good one,Nice,Really nice product...
4      As good as original,Decent,Good one for second...
...
1460    Received the product without spanner,Excellent...
1461    ok,everything was good couldn't return bcoz I ...
1462    very good,Work but front melt after 2 month,Go...
1463    Fan Speed is slow,Good quality,Good product,go...
1464    Works perfect,Ok good product,Nice Product. Re...
Name: review_title, Length: 1465, dtype: object
```

```
In [94]: df['review_title'] = df['review_title'].str.replace(","," ", regex = True)
df['review_title'] = df['review_title'].astype("str")
df['review_title'] = pd.to_numeric(df['review_title'], errors = 'coerce')
```

```
In [96]: df.dtypes
```

```
Out[96]: product_id      string[python]
product_name    string[python]
category        object
discounted_price float64
actual_price     float64
discount_percentage float64
rating           float64
rating_count     float64
about_product   float64
user_id         float64
user_name       float64
review_id       float64
review_title    float64
review_content  object
img_link        object
product_link    object
dtype: object
```

```
In [97]: df['img_link']
```

```
Out[97]: 0      https://m.media-amazon.com/images/W/WEBP_40237...
1      https://m.media-amazon.com/images/W/WEBP_40237...
2      https://m.media-amazon.com/images/W/WEBP_40237...
3      https://m.media-amazon.com/images/I/41V5FtEWPk...
4      https://m.media-amazon.com/images/W/WEBP_40237...
...
1460   https://m.media-amazon.com/images/I/41fDdRtjfx...
1461   https://m.media-amazon.com/images/I/41gzDxk4+k...
1462   https://m.media-amazon.com/images/W/WEBP_40237...
1463   https://m.media-amazon.com/images/W/WEBP_40237...
1464   https://m.media-amazon.com/images/W/WEBP_40237...
Name: img_link, Length: 1465, dtype: object
```

```
In [98]: df['img_link'] = df['img_link'].str.replace(",", "", regex = True)
df['img_link'] = df['img_link'].str.replace("//", "", regex = True)
df['img_link'] = df['img_link'].str.replace("_", "", regex = True)
df['img_link'] = df['img_link'].astype("str")
df['img_link'] = pd.to_numeric(df['img_link'], errors = 'coerce')
```

```
In [100]: df.dtypes
```

```
Out[100]: product_id      string[python]
product_name    string[python]
category        object
discounted_price float64
actual_price     float64
discount_percentage float64
rating           float64
rating_count     float64
about_product   float64
user_id         float64
user_name       float64
review_id       float64
review_title    float64
review_content  object
img_link        float64
product_link    object
dtype: object
```

```
In [101]: df['product_link']
```

```
Out[101]: 0      https://www.amazon.in/Wayona-Braided-WN3LG1-Sy...
1      https://www.amazon.in/Ambrane-Unbreakable-Char...
2      https://www.amazon.in/Sounce-iPhone-Charging-C...
3      https://www.amazon.in/Deuce-300-Resistant-Tang...
4      https://www.amazon.in/Portronics-Konnect-POR-1...
...
1460   https://www.amazon.in/Noir-Aqua-Spanner-Purifi...
1461   https://www.amazon.in/Prestige-Delight-PRWO-1-...
1462   https://www.amazon.in/Bajaj-RX-10-2000-Watt-Co...
1463   https://www.amazon.in/Havells-Ventilair-230mm-...
1464   https://www.amazon.in/Borosil-Jumbo-1000-Watt-...
Name: product_link, Length: 1465, dtype: object
```

```
In [102]: df['product_link'] = df['product_link'].str.replace(",", "", regex = True)
df['product_link'] = df['product_link'].str.replace("//", "", regex = True)
df['product_link'] = df['product_link'].str.replace("_", "", regex = True)
df['product_link'] = df['product_link'].astype("str")
df['product_link'] = pd.to_numeric(df['product_link'], errors = 'coerce')
```

```
In [103]: df['product_link']
```

```
Out[103.. 0      NaN
          1      NaN
          2      NaN
          3      NaN
          4      NaN
          ..
        1460     NaN
        1461     NaN
        1462     NaN
        1463     NaN
        1464     NaN
        Name: product_link, Length: 1465, dtype: float64
```

```
In [104.. df.dtypes
```

```
Out[104.. product_id      string[python]
product_name      string[python]
category          object
discounted_price   float64
actual_price       float64
discount_percentage float64
rating             float64
rating_count       float64
about_product      float64
user_id            float64
user_name          float64
review_id          float64
review_title       float64
review_content     object
img_link           float64
product_link       float64
dtype: object
```

```
In [ ]:
```