assi-eda-2

May 6, 2024

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

```python
[2]: df = pd.read_csv(r"https://drive.usercontent.google.com/download?
     ↪id=1yKlW0x5yyRQijmWbEg5YCsq4wqLl3_oB&export=download&authuser=0")
```

```python
[3]: df
```

```
[3]:        product_id                                        product_name  \
     0      B07JW9H4J1  Wayona Nylon Braided USB to Lightning Fast Cha…
     1      B098NS6PVG  Ambrane Unbreakable 60W / 3A Fast Charging 1.5…
     2      B096MSW6CT  Sounce Fast Phone Charging Cable & Data Sync U…
     3      B08HDJ86NZ  boAt Deuce USB 300 2 in 1 Type-C & Micro USB S…
     4      B08CF3B7N1  Portronics Konnect L 1.2M Fast Charging 3A 8 P…
     …             …                                                   …
     1460   B08L7J3T31  Noir Aqua - 5pcs PP Spun Filter + 1 Spanner | …
     1461   B01M6453MB  Prestige Delight PRWO Electric Rice Cooker (1 …
     1462   B009P2LIL4  Bajaj Majesty RX10 2000 Watts Heat Convector R…
     1463   B00J5DYCCA  Havells Ventil Air DSP 230mm Exhaust Fan (Pist…
     1464   B01486F4G6  Borosil Jumbo 1000-Watt Grill Sandwich Maker (…

                                              category discounted_price  \
     0      Computers&Accessories|Accessories&Peripherals|…              399
     1      Computers&Accessories|Accessories&Peripherals|…              199
     2      Computers&Accessories|Accessories&Peripherals|…              199
     3      Computers&Accessories|Accessories&Peripherals|…              329
     4      Computers&Accessories|Accessories&Peripherals|…              154
     …                                                   …                …
     1460   Home&Kitchen|Kitchen&HomeAppliances|WaterPurif…              379
     1461   Home&Kitchen|Kitchen&HomeAppliances|SmallKitch…            2,280
     1462   Home&Kitchen|Heating,Cooling&AirQuality|RoomHe…            2,219
     1463   Home&Kitchen|Heating,Cooling&AirQuality|Fans|E…            1,399
     1464   Home&Kitchen|Kitchen&HomeAppliances|SmallKitch…            2,863

            actual_price discount_percentage rating rating_count  \
```

0	1,099	64%	4.2	24,269
1	349	43%	4.0	43,994
2	1,899	90%	3.9	7,928
3	699	53%	4.2	94,363
4	399	61%	4.2	16,905
…	…	…	…	…
1460	919	59%	4	1,090
1461	3,045	25%	4.1	4,118
1462	3,080	28%	3.6	468
1463	1,890	26%	4	8,031
1464	3,690	22%	4.3	6,987

```
                                                 about_product  \
0      High Compatibility : Compatible With iPhone 12…
1      Compatible with all Type C enabled devices, be…
2       Fast Charger& Data Sync -With built-in safet…
3      The boAt Deuce USB 300 2 in 1 cable is compati…
4      [CHARGE & SYNC FUNCTION]- This cable comes wit…
…                                                    …
1460   SUPREME QUALITY 90 GRAM 3 LAYER THIK PP SPUN F…
1461                         230 Volts, 400 watts, 1 Year
1462   International design and styling|Two heat sett…
1463   Fan sweep area: 230 MM ; Noise level: (40 – 45…
1464   Brand-Borosil, Specification â€" 23V ~ 5Hz;1 W…


                                                       user_id  \
0      AG3D6O4STAQKAY2UVGEUV46KN35Q,AHMY5CWJMMK5BJRBB…
1      AECPFYFQVRUWC3KGNLJIOREFP5LQ,AGYYVPDD7YG7FYNBX…
2      AGU3BBQ2V2DDAMOAKGFAWDDQ6QHA,AESFLDV2PT363T2AQ…
3      AEWAZDZZJLQUYVOVGBEUKSLXHQ5A,AG5HTSFRRE6NL3M5S…
4      AE3Q6KSUK5P75D5HFYHCRAOLODSA,AFUGIFH5ZAFXRDSZH…
…                                                    …
1460   AHITFY6AHALOFOHOZEOC6XBP4FEA,AFRABBODZJZQB6Z4U…
1461   AFG5FM3NEMOL6BNFRV2NK5FNJCHQ,AGEINTRN6Z563RMLH…
1462   AGVPWCMAHYQWJOQKMUJN4DW3KM5Q,AF4Q3E66MY4SR7YQZ…
1463   AF2JQCLSCY3QJATWUNNHUSVUPNQQ,AFDMLUXC5LS5RXDJS…
1464   AFGW5PT3R6ZAVQR4Y5MWVAKBZAYA,AG7QNJ2SCS5VS5VYY…


                                                     user_name  \
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…

                                              review_id  \
0       R3HXWT0LRP0NMF,R2AJM3LFTLZHFO,R6AQJGUP6P86,R1K…
1       RGIQEG07R9HS2,R1SMWZQ86XIN8U,R2J3Y1WL29GWDE,RY…
2       R3J3EQQ9TZI5ZJ,R3E7WBGK7IDOKV,RWU79XKQ6I1QF,R2…
3       R3EEUZKKK9J36I,R3HJVYCLYOY554,REDECAZ7AMPQC,R1…
4       R1BP4L2HH9TFUP,R16PVJEXKV6QZS,R2UPDB81N66T4P,R…
…                                                     …
1460    R3G3XFHPBFF0E8,R3C0BZCD32EIGW,R2EBVBCN9QPD9R,R…
1461    R3DDL2UPKQ2CK9,R2SYYU1OATVIU5,R1VM993161IYRW,R…
1462    R1TLRJVW4STY5I,R2O455KRN493R1,R3Q5MVGBRIAS2G,R…
1463    R39Q2Y79MM9SWK,R3079BG1NIH6MB,R29A31ZELTZNJM,R…
1464    R2ORBRZOWEUJT9,ROKIFK9R2ISSE,R30EEG2FNJSN5I,R2…

                                           review_title  \
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…

                                         review_content  \
0       Looks durable Charging is fine tooNo complains…
1       I ordered this cable to connect my phone to An…
2       Not quite durable and sturdy,https://m.media-a…
3       Good product,long wire,Charges good,Nice,I bou…
4       Bought this instead of original apple, does th…
…                                                     …
1460    I received product without spanner,Excellent p…
1461    ok,got everything as mentioned but the measuri…
1462    plastic but cool body ,u have to find sturdy s…
1463    I have installed this in my kitchen working fi…
1464    It does it job perfectly..only issue is temp c…

                                               img_link  \
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…


                                         product_link
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-…


[1465 rows x 16 columns]
```

[4]: `df.shape`

[4]: (1465, 16)

[5]: `df.isnull()`

[5]:
```
      product_id  product_name  category  discounted_price  actual_price  \
0         False         False     False             False         False
1         False         False     False             False         False
2         False         False     False             False         False
3         False         False     False             False         False
4         False         False     False             False         False
…             …             …         …                 …             …
1460      False         False     False             False         False
1461      False         False     False             False         False
1462      False         False     False             False         False
1463      False         False     False             False         False
1464      False         False     False             False         False


      discount_percentage  rating  rating_count  about_product  user_id  \
0                   False   False         False          False    False
1                   False   False         False          False    False
2                   False   False         False          False    False
```

3	False	False	False	False	False
4	False	False	False	False	False
…	…	…	…	…	…
1460	False	False	False	False	False
1461	False	False	False	False	False
1462	False	False	False	False	False
1463	False	False	False	False	False
1464	False	False	False	False	False

	user_name	review_id	review_title	review_content	img_link \
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False
…	…	…	…	…	…
1460	False	False	False	False	False
1461	False	False	False	False	False
1462	False	False	False	False	False
1463	False	False	False	False	False
1464	False	False	False	False	False

	product_link
0	False
1	False
2	False
3	False
4	False
…	…
1460	False
1461	False
1462	False
1463	False
1464	False

[1465 rows x 16 columns]

```
[6]: 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
```

[7]: `df['rating'].dtype`

[7]: `dtype('O')`

[8]: `df.dtypes[df.dtypes == 'object'].index`

[8]:
```
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')
```

[9]: `df.dtypes[df.dtypes != 'object'].index`

[9]: `Index([], dtype='object')`

[10]: `pd.Categorical(df['product_name'])`

[10]:
```
['Wayona Nylon Braided USB to Lightning Fast Ch…, 'Ambrane Unbreakable 60W /
 3A Fast Charging 1…, 'Sounce Fast Phone Charging Cable & Data Sync …, 'boAt
 Deuce USB 300 2 in 1 Type-C & Micro USB …, 'Portronics Konnect L 1.2M Fast
 Charging 3A 8 …, …, 'Noir Aqua - 5pcs PP Spun Filter + 1 Spanner |…,
 'Prestige Delight PRWO Electric Rice Cooker (1…, 'Bajaj Majesty RX10 2000
 Watts Heat Convector …, 'Havells Ventil Air DSP 230mm Exhaust Fan (Pis…,
 'Borosil Jumbo 1000-Watt Grill Sandwich Maker …]
Length: 1465
Categories (1337, object): ['!!!1000 Watt/2000-Watt Room Heater!! Fan Heate…,
 '!!!HANEUL!!1000 Watt/2000-Watt Room Heater!! F…, '10WeRun Id-116 Bluetooth
 Smartwatch Wireless …, '10k 8k 4k HDMI Cable, Certified 48Gbps 1ms Ul…, …,
 'realme narzo 50 (Speed Blue, 4GB RAM+64GB Sto…, 'realme narzo 50i (Mint
 Green, 2GB RAM+32GB St…, 'rts [2 Pack] Mini USB C Type C Adapter Plug, …,
```

```
'tizum HDMI to VGA Adapter Cable 1080P for Pro…]
```
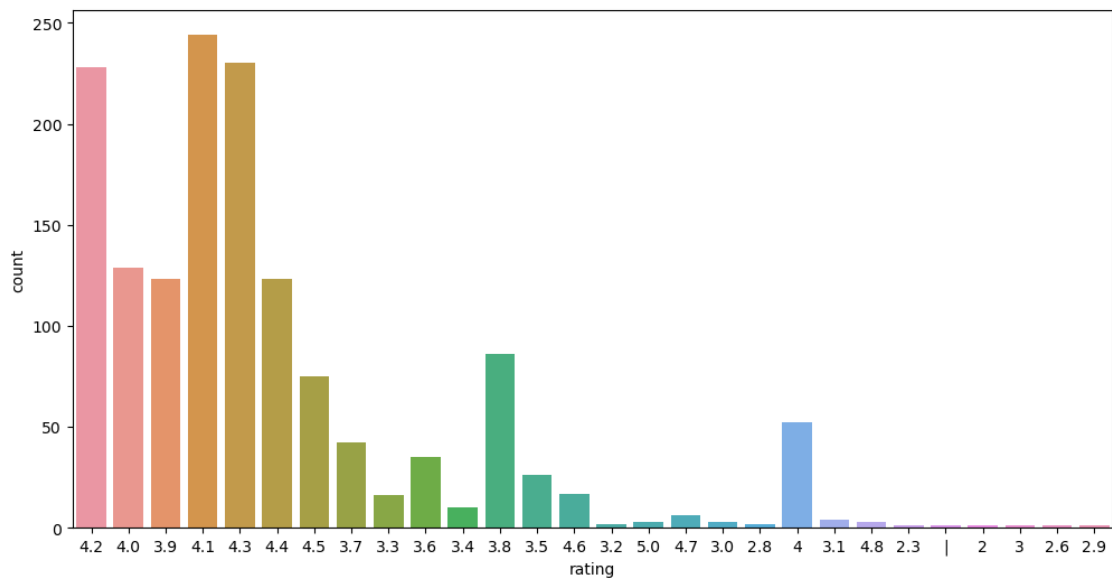
```
[11]: pd.Categorical(df['rating'])
```

```
[11]: ['4.2', '4.0', '3.9', '4.2', '4.2', …, '4', '4.1', '3.6', '4', '4.3']
      Length: 1465
      Categories (28, object): ['2', '2.3', '2.6', '2.8', …, '4.7', '4.8', '5.0',
      '|']
```

0.1 Q. 1 what is the Average rating for each product category

```
[12]: df['rating'].unique()
```

```
[12]: array(['4.2', '4.0', '3.9', '4.1', '4.3', '4.4', '4.5', '3.7', '3.3',
             '3.6', '3.4', '3.8', '3.5', '4.6', '3.2', '5.0', '4.7', '3.0',
             '2.8', '4', '3.1', '4.8', '2.3', '|', '2', '3', '2.6', '2.9'],
             dtype=object)
```

```
[13]: plt.figure(figsize=(12,6))
      sns.countplot(x='rating',data=df)

      plt.show()
```



```
[14]: df.groupby('rating').mean()
```

```
C:\Users\Pooja Rathore\AppData\Local\Temp\ipykernel_4940\1171111788.py:1:
FutureWarning: The default value of numeric_only in DataFrameGroupBy.mean is
deprecated. In a future version, numeric_only will default to False. Either
```

```
     specify numeric_only or select only columns which should be valid for the
     function.
       df.groupby('rating').mean()
```

[14]: Empty DataFrame
      Columns: []
      Index: [2, 2.3, 2.6, 2.8, 2.9, 3, 3.0, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8,
      3.9, 4, 4.0, 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 5.0, |]

[15]: ```
df['rating'].mean
```

[15]: <bound method NDFrame._add_numeric_operations.<locals>.mean of 0        4.2
      1        4.0
      2        3.9
      3        4.2
      4        4.2

              …
      1460     4
      1461    4.1
      1462    3.6
      1463     4
      1464    4.3
      Name: rating, Length: 1465, dtype: object>

[16]: ```
df['rating'].value_counts()
```

[16]: 4.1   244
      4.3   230
      4.2   228
      4.0   129
      3.9   123
      4.4   123
      3.8   86
      4.5   75
      4   52
      3.7   42
      3.6   35
      3.5   26
      4.6   17
      3.3   16
      3.4   10
      4.7   6
      3.1   4
      5.0   3
      3.0   3
      4.8   3
      3.2   2

```
2.8      2
2.3      1
|        1
2        1
3        1
2.6      1
2.9      1
Name: rating, dtype: int64
```

[ ]:

[17]:
```python
df.groupby(['rating','category']).mean()
```

C:\Users\Pooja Rathore\AppData\Local\Temp\ipykernel_4940\1365135195.py:1:
FutureWarning: The default value of numeric_only in DataFrameGroupBy.mean is
deprecated. In a future version, numeric_only will default to False. Either
specify numeric_only or select only columns which should be valid for the
function.
  df.groupby(['rating','category']).mean()

[17]: Empty DataFrame
Columns: []
Index: [(2, Home&Kitchen|Heating,Cooling&AirQuality|RoomHeaters|FanHeaters),
(2.3, Home&Kitchen|Kitchen&HomeAppliances|SmallKitchenAppliances|HandBlenders),
(2.6, Home&Kitchen|Kitchen&HomeAppliances|SmallKitchenAppliances|VacuumSealers),
(2.8, Electronics|WearableTechnology|SmartWatches), (2.8,
Home&Kitchen|Kitchen&HomeAppliances|SmallKitchenAppliances|JuicerMixerGrinders),
(2.9, Home&Kitchen|Kitchen&HomeAppliances|WaterPurifiers&Accessories|WaterPurifi
erAccessories), (3,
Home&Kitchen|Kitchen&HomeAppliances|SmallKitchenAppliances|VacuumSealers), (3.0,
Computers&Accessories|Accessories&Peripherals|Cables&Accessories|Cables|USBCable
s), (3.0, Electronics|HomeTheater,TV&Video|Accessories|RemoteControls), (3.0,
Electronics|WearableTechnology|SmartWatches), (3.1,
Home&Kitchen|Heating,Cooling&AirQuality|RoomHeaters|FanHeaters), (3.1, Home&Kitc
hen|Kitchen&HomeAppliances|SewingMachines&Accessories|Sewing&EmbroideryMachines)
, (3.1, Home&Kitchen|Kitchen&HomeAppliances|SmallKitchenAppliances|Juicers),
(3.1, Home&Kitchen|Kitchen&HomeAppliances|Vacuum,Cleaning&Ironing|Irons,Steamers
&Accessories|LintShavers), (3.2,
Electronics|HomeTheater,TV&Video|Accessories|RemoteControls), (3.2, Home&Kitchen
|Kitchen&HomeAppliances|SmallKitchenAppliances|MiniFoodProcessors&Choppers),
(3.3, Computers&Accessories|Accessories&Peripherals|Audio&VideoAccessories|PCMic
rophones), (3.3, Computers&Accessories|Accessories&Peripherals|Cables&Accessorie
s|Cables|USBCables), (3.3, Computers&Accessories|Accessories&Peripherals|Keyboar
ds,Mice&InputDevices|Keyboard&MiceAccessories|DustCovers), (3.3,
Electronics|HomeTheater,TV&Video|Accessories|RemoteControls), (3.3,
Electronics|WearableTechnology|SmartWatches), (3.3,
Home&Kitchen|Heating,Cooling&AirQuality|RoomHeaters|ElectricHeaters), (3.3,
```

Home&Kitchen|Heating,Cooling&AirQuality|RoomHeaters|FanHeaters), (3.3, Home&Kitc
hen|Kitchen&HomeAppliances|Coffee,Tea&Espresso|CoffeeGrinders|ElectricGrinders),
(3.3, Home&Kitchen|Kitchen&HomeAppliances|SmallKitchenAppliances|HandBlenders),
(3.3, Home&Kitchen|Kitchen&HomeAppliances|SmallKitchenAppliances|VacuumSealers),
(3.4, Computers&Accessories|Accessories&Peripherals|USBHubs), (3.4,
Computers&Accessories|NetworkingDevices|NetworkAdapters|WirelessUSBAdapters),
(3.4, Computers&Accessories|Printers,Inks&Accessories|Printers|InkjetPrinters),
(3.4, Electronics|Headphones,Earbuds&Accessories|Headphones|In-Ear), (3.4,
Electronics|HomeTheater,TV&Video|Accessories|RemoteControls), (3.4,
Electronics|HomeTheater,TV&Video|Televisions|SmartTelevisions), (3.4,
Home&Kitchen|Heating,Cooling&AirQuality|RoomHeaters|FanHeaters), (3.4, Home&Kitc
hen|Heating,Cooling&AirQuality|WaterHeaters&Geysers|InstantWaterHeaters), (3.4,
Home&Kitchen|Kitchen&HomeAppliances|SmallKitchenAppliances|Kettles&HotWaterDispe
nsers|Kettle&ToasterSets), (3.4,
Home&Kitchen|Kitchen&HomeAppliances|SmallKitchenAppliances|MixerGrinders), (3.5,
Computers&Accessories|Accessories&Peripherals|Audio&VideoAccessories|PCHeadsets)
, (3.5, Computers&Accessories|Accessories&Peripherals|Cables&Accessories|Cables|
USBCables), (3.5, Computers&Accessories|Accessories&Peripherals|Keyboards,Mice&I
nputDevices|GraphicTablets), (3.5, Computers&Accessories|Accessories&Peripherals
|Keyboards,Mice&InputDevices|Keyboard&MiceAccessories|DustCovers), (3.5, Compute
rs&Accessories|Accessories&Peripherals|Keyboards,Mice&InputDevices|Keyboard&Mous
eSets), (3.5,
Computers&Accessories|Accessories&Peripherals|Keyboards,Mice&InputDevices|Mice),
(3.5, Computers&Accessories|Accessories&Peripherals|LaptopAccessories|Lapdesks),
(3.5, Computers&Accessories|Accessories&Peripherals|LaptopAccessories|LaptopChar
gers&PowerSupplies), (3.5,
Computers&Accessories|Printers,Inks&Accessories|Printers), (3.5,
Electronics|Headphones,Earbuds&Accessories|Headphones|In-Ear), (3.5,
Electronics|HomeTheater,TV&Video|Accessories|3DGlasses), (3.5,
Electronics|HomeTheater,TV&Video|Accessories|RemoteControls), (3.5,
Electronics|Mobiles&Accessories|Smartphones&BasicMobiles|BasicMobiles), (3.5,
Electronics|WearableTechnology|SmartWatches), (3.5,
Home&Kitchen|Heating,Cooling&AirQuality|RoomHeaters|ElectricHeaters), (3.5,
Home&Kitchen|Heating,Cooling&AirQuality|RoomHeaters|FanHeaters), (3.5,
Home&Kitchen|Heating,Cooling&AirQuality|RoomHeaters|HalogenHeaters), (3.5,
Home&Kitchen|Kitchen&HomeAppliances|Coffee,Tea&Espresso|MilkFrothers), (3.5, Hom
e&Kitchen|Kitchen&HomeAppliances|SewingMachines&Accessories|Sewing&EmbroideryMac
hines), (3.5, Home&Kitchen|Kitchen&HomeAppliances|Vacuum,Cleaning&Ironing|Irons,
Steamers&Accessories|LintShavers), (3.6, Computers&Accessories|Accessories&Perip
herals|Keyboards,Mice&InputDevices|Keyboards), (3.6,
Computers&Accessories|Accessories&Peripherals|LaptopAccessories|Lapdesks), (3.6,
Computers&Accessories|Accessories&Peripherals|LaptopAccessories|LaptopChargers&P
owerSupplies), (3.6,
Computers&Accessories|Accessories&Peripherals|USBGadgets|Lamps), (3.6,
Computers&Accessories|NetworkingDevices|NetworkAdapters|WirelessUSBAdapters),
(3.6, Computers&Accessories|Printers,Inks&Accessories|Printers), (3.6,
Electronics|Headphones,Earbuds&Accessories|Headphones|In-Ear), (3.6,

```
Electronics|HomeTheater,TV&Video|Accessories|RemoteControls), (3.6, Electronics|
HomeTheater,TV&Video|Accessories|TVMounts,Stands&Turntables|TVWall&CeilingMounts
), (3.6, Electronics|Mobiles&Accessories|MobileAccessories|Stands), (3.6,
Home&Kitchen|Heating,Cooling&AirQuality|Humidifiers), (3.6,
Home&Kitchen|Heating,Cooling&AirQuality|RoomHeaters|ElectricHeaters), (3.6,
Home&Kitchen|Heating,Cooling&AirQuality|RoomHeaters|FanHeaters), (3.6,
Home&Kitchen|Heating,Cooling&AirQuality|RoomHeaters|HeatConvectors), (3.6, Home&
Kitchen|Heating,Cooling&AirQuality|WaterHeaters&Geysers|InstantWaterHeaters),
(3.6, Home&Kitchen|Kitchen&HomeAppliances|Coffee,Tea&Espresso|MilkFrothers),
(3.6, Home&Kitchen|Kitchen&HomeAppliances|SmallKitchenAppliances|HandBlenders),
(3.6,
Home&Kitchen|Kitchen&HomeAppliances|SmallKitchenAppliances|JuicerMixerGrinders),
(3.6, Home&Kitchen|Kitchen&HomeAppliances|SmallKitchenAppliances|Kettles&HotWate
rDispensers|Kettle&ToasterSets), (3.6, Home&Kitchen|Kitchen&HomeAppliances|Vacuu
m,Cleaning&Ironing|Irons,Steamers&Accessories|Irons|DryIrons), (3.6, Home&Kitche
n|Kitchen&HomeAppliances|Vacuum,Cleaning&Ironing|Irons,Steamers&Accessories|Iron
s|SteamIrons), (3.6, Home&Kitchen|Kitchen&HomeAppliances|Vacuum,Cleaning&Ironing
|Irons,Steamers&Accessories|LintShavers), (3.6, Home&Kitchen|Kitchen&HomeApplian
ces|WaterPurifiers&Accessories|WaterFilters&Purifiers), (3.6, Home&Kitchen|Kitch
en&HomeAppliances|WaterPurifiers&Accessories|WaterPurifierAccessories), (3.7, Co
mputers&Accessories|Accessories&Peripherals|Cables&Accessories|Cables|USBCables)
, (3.7, Computers&Accessories|Accessories&Peripherals|USBHubs), (3.7,
Computers&Accessories|NetworkingDevices|NetworkAdapters|WirelessUSBAdapters),
(3.7, Electronics|Cameras&Photography|VideoCameras), (3.7,
Electronics|Headphones,Earbuds&Accessories|Headphones|In-Ear), (3.7,
Electronics|HomeAudio|Speakers|OutdoorSpeakers), (3.7,
Electronics|HomeTheater,TV&Video|Accessories|Cables|HDMICables), (3.7,
Electronics|HomeTheater,TV&Video|Accessories|RemoteControls), (3.7,
Electronics|HomeTheater,TV&Video|Televisions|SmartTelevisions), (3.7,
Electronics|Mobiles&Accessories|MobileAccessories|Mounts|Bedstand&DeskMounts),
(3.7, Electronics|WearableTechnology|SmartWatches), (3.7,
Home&Kitchen|Heating,Cooling&AirQuality|RoomHeaters|ElectricHeaters), (3.7,
Home&Kitchen|Heating,Cooling&AirQuality|RoomHeaters|HalogenHeaters), (3.7, Home&
Kitchen|Heating,Cooling&AirQuality|WaterHeaters&Geysers|InstantWaterHeaters),
(3.7, Home&Kitchen|HomeStorage&Organization|LaundryOrganization|LaundryBaskets),
(3.7, Home&Kitchen|Kitchen&HomeAppliances|SewingMachines&Accessories|Sewing&Embr
oideryMachines), (3.7, Home&Kitchen|Kitchen&HomeAppliances|SmallKitchenAppliance
s|DigitalKitchenScales), (3.7,
Home&Kitchen|Kitchen&HomeAppliances|SmallKitchenAppliances|HandBlenders), (3.7,
Home&Kitchen|Kitchen&HomeAppliances|SmallKitchenAppliances|JuicerMixerGrinders),
(3.7, Home&Kitchen|Kitchen&HomeAppliances|SmallKitchenAppliances|Kettles&HotWate
rDispensers|Kettle&ToasterSets), …]

[645 rows x 0 columns]
```

[ ]:

0.2 Q.2 what are the top rating __count products by category ?

```
[18]: df['rating_count']
```

```
[18]: 0          24,269
      1          43,994
      2           7,928
      3          94,363
      4          16,905
                  …
      1460        1,090
      1461        4,118
      1462          468
      1463        8,031
      1464        6,987
      Name: rating_count, Length: 1465, dtype: object
```

```
[29]: df['rating_count'].value_counts()[:5]
```

```
[29]: 9,378      9
      18,998     8
      24,269     7
      43,994     6
      32,840     5
      Name: rating_count, dtype: int64
```

```
[20]: category_rating_counts = df.groupby('category')['rating'].count()
```

```
[21]: category_rating_counts
```

```
[21]: category
      Car&Motorbike|CarAccessories|InteriorAccessories|AirPurifiers&Ionizers
      1
      Computers&Accessories|Accessories&Peripherals|Adapters|USBtoUSBAdapters
      2
      Computers&Accessories|Accessories&Peripherals|Audio&VideoAccessories|PCHeadsets
      1
      Computers&Accessories|Accessories&Peripherals|Audio&VideoAccessories|PCMicrophon
      es                                                 2
      Computers&Accessories|Accessories&Peripherals|Audio&VideoAccessories|PCSpeakers
      2
                                                     ..
      OfficeProducts|OfficePaperProducts|Paper|Stationery|Pens,Pencils&WritingSupplies
      |Pens&Refills|GelInkRollerballPens        2
      OfficeProducts|OfficePaperProducts|Paper|Stationery|Pens,Pencils&WritingSupplies
      |Pens&Refills|LiquidInkRollerballPens      2
      OfficeProducts|OfficePaperProducts|Paper|Stationery|Pens,Pencils&WritingSupplies
```

```
       |Pens&Refills|RetractableBallpointPens    2
OfficeProducts|OfficePaperProducts|Paper|Stationery|Pens,Pencils&WritingSupplies
       |Pens&Refills|StickBallpointPens          3
Toys&Games|Arts&Crafts|Drawing&PaintingSupplies|ColouringPens&Markers
1
Name: rating, Length: 211, dtype: int64
```

[22]: ```python
rating_count_by_catagory = category_rating_counts.sort_values(ascending=False)
```

[23]: ```python
rating_count_by_catagory
```

[23]: ```
category
Computers&Accessories|Accessories&Peripherals|Cables&Accessories|Cables|USBCable
s    233
Electronics|WearableTechnology|SmartWatches
76
Electronics|Mobiles&Accessories|Smartphones&BasicMobiles|Smartphones
68
Electronics|HomeTheater,TV&Video|Televisions|SmartTelevisions
63
Electronics|Headphones,Earbuds&Accessories|Headphones|In-Ear
52
      …
Electronics|HomeAudio|Speakers|SoundbarSpeakers
1
Electronics|HomeAudio|Speakers|MultimediaSpeakerSystems
1
Electronics|HomeAudio|MediaStreamingDevices|StreamingClients
1
Electronics|HomeAudio|Accessories|SpeakerAccessories|Mounts
1
Toys&Games|Arts&Crafts|Drawing&PaintingSupplies|ColouringPens&Markers
1
Name: rating, Length: 211, dtype: int64
```

[24]: *##insights>>>>Computers&Accessories is top rating catagory*

[25]: ```python
df
```

[25]: ```
         product_id                                    product_name  \
0        B07JW9H4J1  Wayona Nylon Braided USB to Lightning Fast Cha…
1        B098NS6PVG  Ambrane Unbreakable 60W / 3A Fast Charging 1.5…
2        B096MSW6CT  Sounce Fast Phone Charging Cable & Data Sync U…
3        B08HDJ86NZ  boAt Deuce USB 300 2 in 1 Type-C & Micro USB S…
4        B08CF3B7N1  Portronics Konnect L 1.2M Fast Charging 3A 8 P…
…               …                                              …
1460     B08L7J3T31  Noir Aqua - 5pcs PP Spun Filter + 1 Spanner | …
```

```
1461  B01M6453MB  Prestige Delight PRWO Electric Rice Cooker (1 …
1462  B009P2LIL4  Bajaj Majesty RX10 2000 Watts Heat Convector R…
1463  B00J5DYCCA  Havells Ventil Air DSP 230mm Exhaust Fan (Pist…
1464  B01486F4G6  Borosil Jumbo 1000-Watt Grill Sandwich Maker (…


                                        category discounted_price  \
0     Computers&Accessories|Accessories&Peripherals|…              399
1     Computers&Accessories|Accessories&Peripherals|…              199
2     Computers&Accessories|Accessories&Peripherals|…              199
3     Computers&Accessories|Accessories&Peripherals|…              329
4     Computers&Accessories|Accessories&Peripherals|…              154
…                                            …                …
1460  Home&Kitchen|Kitchen&HomeAppliances|WaterPurif…              379
1461  Home&Kitchen|Kitchen&HomeAppliances|SmallKitch…            2,280
1462  Home&Kitchen|Heating,Cooling&AirQuality|RoomHe…            2,219
1463  Home&Kitchen|Heating,Cooling&AirQuality|Fans|E…            1,399
1464  Home&Kitchen|Kitchen&HomeAppliances|SmallKitch…            2,863


     actual_price discount_percentage rating rating_count  \
0           1,099                 64%    4.2       24,269
1             349                 43%    4.0       43,994
2           1,899                 90%    3.9        7,928
3             699                 53%    4.2       94,363
4             399                 61%    4.2       16,905
…               …                   …      …            …
1460          919                 59%      4        1,090
1461        3,045                 25%    4.1        4,118
1462        3,080                 28%    3.6          468
1463        1,890                 26%      4        8,031
1464        3,690                 22%    4.3        6,987


                                     about_product  \
0     High Compatibility : Compatible With iPhone 12…
1     Compatible with all Type C enabled devices, be…
2      Fast Charger& Data Sync -With built-in safet…
3     The boAt Deuce USB 300 2 in 1 cable is compati…
4     [CHARGE & SYNC FUNCTION]- This cable comes wit…
…                                                  …
1460  SUPREME QUALITY 90 GRAM 3 LAYER THIK PP SPUN F…
1461                      230 Volts, 400 watts, 1 Year
1462  International design and styling|Two heat sett…
1463  Fan sweep area: 230 MM ; Noise level: (40 – 45…
1464  Brand-Borosil, Specification â€" 23V ~ 5Hz;1 W…


                                           user_id  \
0     AG3D6O4STAQKAY2UVGEUV46KN35Q,AHMY5CWJMMK5BJRBB…
1     AECPFYFQVRUWC3KGNLJIOREFP5LQ,AGYYVPDD7YG7FYNBX…
```

```
2      AGU3BBQ2V2DDAMOAKGFAWDDQ6QHA,AESFLDV2PT363T2AQ…
3      AEWAZDZZJLQUYVOVGBEUKSLXHQ5A,AG5HTSFRRE6NL3M5S…
4      AE3Q6KSUK5P75D5HFYHCRAOLODSA,AFUGIFH5ZAFXRDSZH…
…                                                    …
1460   AHITFY6AHALOFOHOZEOC6XBP4FEA,AFRABBODZJZQB6Z4U…
1461   AFG5FM3NEMOL6BNFRV2NK5FNJCHQ,AGEINTRN6Z563RMLH…
1462   AGVPWCMAHYQWJOQKMUJN4DW3KM5Q,AF4Q3E66MY4SR7YQZ…
1463   AF2JQCLSCY3QJATWUNNHUSVUPNQQ,AFDMLUXC5LS5RXDJS…
1464   AFGW5PT3R6ZAVQR4Y5MWVAKBZAYA,AG7QNJ2SCS5VS5VYY…


                                         user_name  \
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…


                                         review_id  \
0      R3HXWT0LRP0NMF,R2AJM3LFTLZHF0,R6AQJGUP6P86,R1K…
1      RGIQEG07R9HS2,R1SMWZQ86XIN8U,R2J3Y1WL29GWDE,RY…
2      R3J3EQQ9TZI5ZJ,R3E7WBGK7ID0KV,RWU79XKQ6I1QF,R2…
3      R3EEUZKKK9J36I,R3HJVYCLYOY554,REDECAZ7AMPQC,R1…
4      R1BP4L2HH9TFUP,R16PVJEXKV6QZS,R2UPDB81N66T4P,R…
…                                                    …
1460   R3G3XFHPBFF0E8,R3C0BZCD32EIGW,R2EBVBCN9QPD9R,R…
1461   R3DDL2UPKQ2CK9,R2SYYU1OATVIU5,R1VM993161IYRW,R…
1462   R1TLRJVW4STY5I,R2O455KRN493R1,R3Q5MVGBRIAS2G,R…
1463   R39Q2Y79MM9SWK,R3O79BG1NIH6MB,R29A31ZELTZNJM,R…
1464   R2ORBRZOWEUJT9,ROKIFK9R2ISSE,R3OEEG2FNJSN5I,R2…


                                      review_title  \
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…

                                    review_content  \
0      Looks durable Charging is fine tooNo complains…
1      I ordered this cable to connect my phone to An…
2      Not quite durable and sturdy,https://m.media-a…
3      Good product,long wire,Charges good,Nice,I bou…
4      Bought this instead of original apple, does th…
…                                                   …
1460   I received product without spanner,Excellent p…
1461   ok,got everything as mentioned but the measuri…
1462   plastic but cool body ,u have to find sturdy s…
1463   I have installed this in my kitchen working fi…
1464   It does it job perfectly..only issue is temp c…

                                        img_link  \
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…

                                      product_link
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-…

[1465 rows x 16 columns]
```

[26]: `df.columns`

[26]: 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')
```

[ ]:

0.3   Q.3 What is the distribution of discounted price vs actual price

[27]:
```
plt.figure(figsize=(10, 6))
plt.hist(df['actual_price'], bins=10, alpha=0.5, label='Actual_Price')
```

[27]:
```
(array([635., 164., 155., 139.,  83.,  70.,  62.,  59.,  52.,  46.]),
 array([  0. ,  44.8,  89.6, 134.4, 179.2, 224. , 268.8, 313.6, 358.4,
        403.2, 448. ]),
 <BarContainer object of 10 artists>)
```
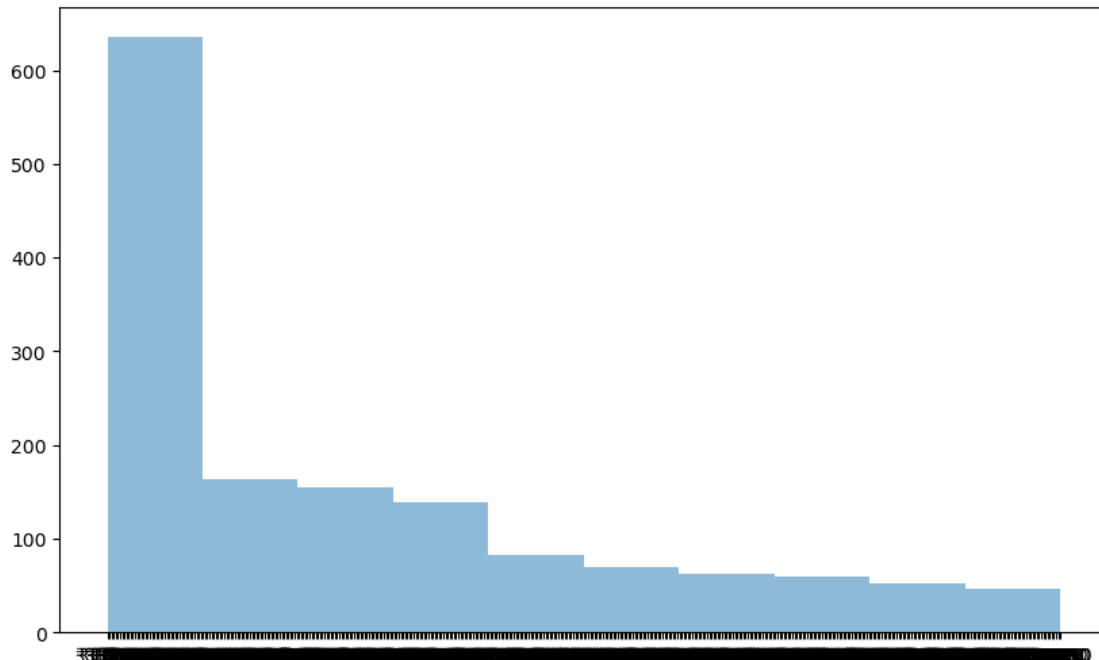

[32]:
```
plt.figure(figsize=(12, 9))
df.plot.scatter(x = 'actual_price', y = "discounted_price")
```

[32]: <Axes: xlabel='actual_price', ylabel='discounted_price'>

<Figure size 1200x900 with 0 Axes>
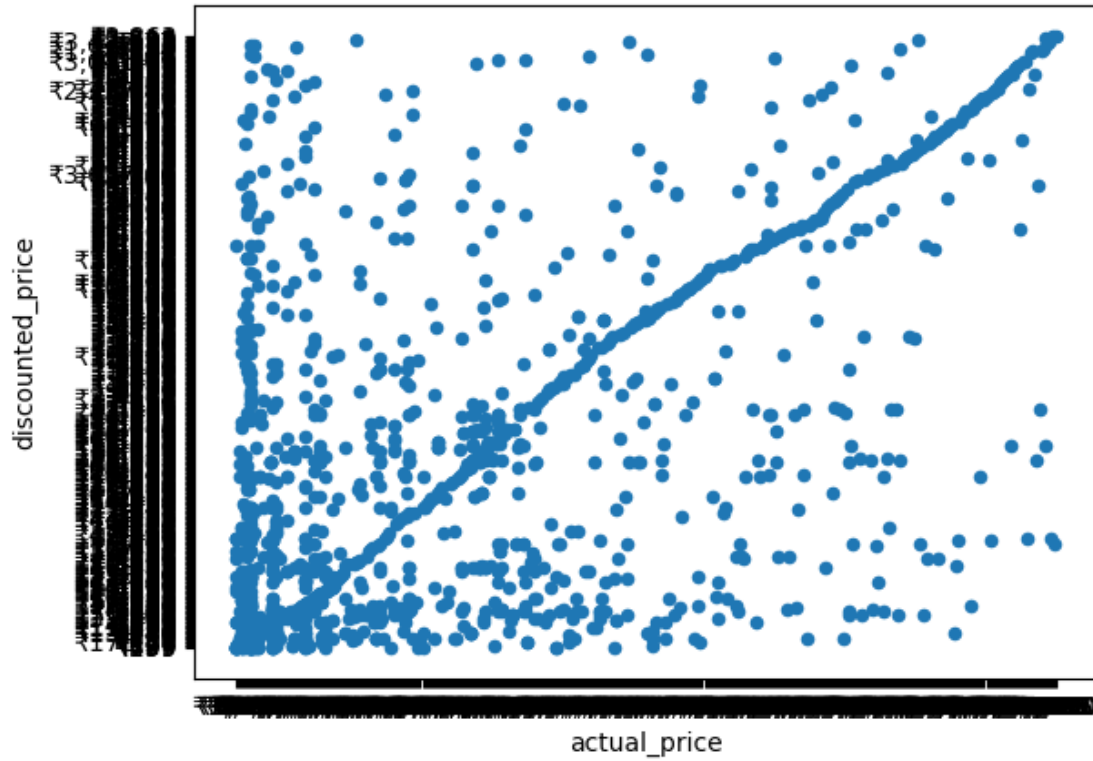
```
[33]: plt.figure(figsize=(10, 6))
      plt.hist(df['actual_price'], bins=20, alpha=0.5, label='Actual_Price')
      plt.hist(df['discounted_price'], bins=20, alpha=0.5, label='Discounted_Price')
      plt.xlabel('Price')
      plt.ylabel('Frequency')
      plt.title('Distribution of Actual Price vs. Discounted Price')
      plt.legend()
      plt.show()
```

Distribution of Actual Price vs. Discounted Price
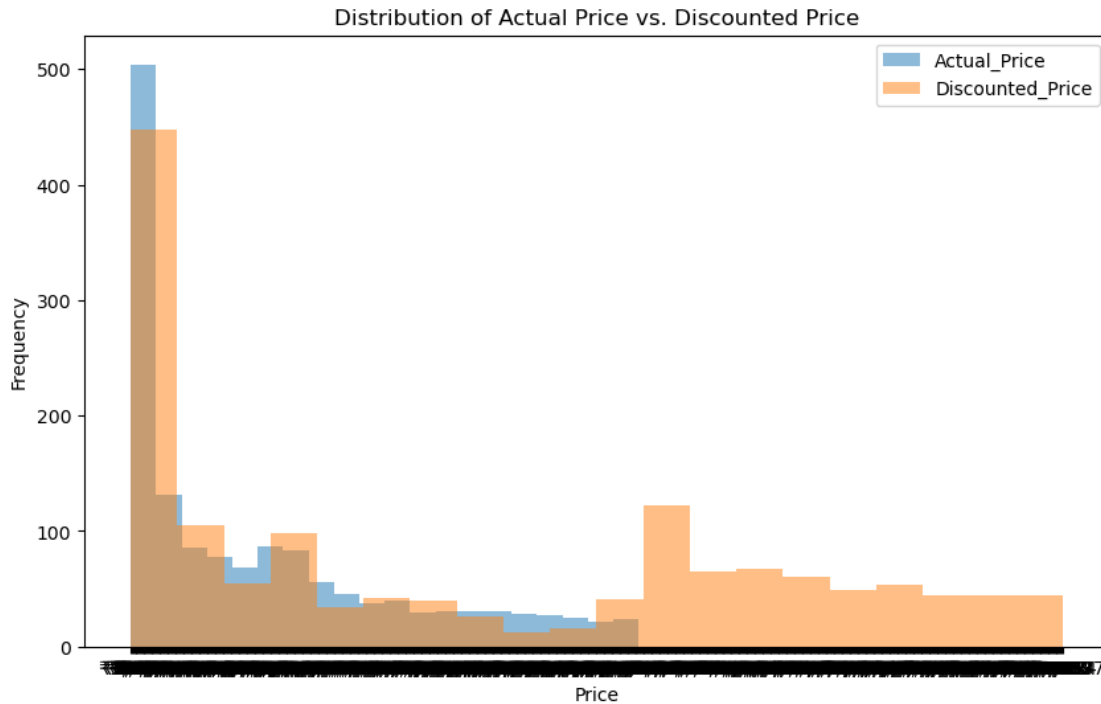
```
[34]:  # Plot density plots for actual price and discounted price
       plt.figure(figsize=(10, 6))
       sns.kdeplot(df['actual_price'], shade=True, label='Actual_Price')
       sns.kdeplot(df['discounted_price'], shade=True, label='Discounted_Price')
       plt.xlabel('Price')
       plt.ylabel('Density')
       plt.title('Distribution of Actual Price vs. Discounted Price (Density Plot)')
       plt.legend()
       plt.show()
```

C:\Users\Pooja Rathore\AppData\Local\Temp\ipykernel_4940\3022347711.py:3:
FutureWarning:

`shade` is now deprecated in favor of `fill`; setting `fill=True`.
This will become an error in seaborn v0.14.0; please update your code.

  sns.kdeplot(df['actual_price'], shade=True, label='Actual_Price')

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
Cell In[34], line 3
      1 # Plot density plots for actual price and discounted price
      2 plt.figure(figsize=(10, 6))
----> 3 sns.kdeplot(df['actual_price'], shade=True, label='Actual_Price')
```

```
  4 sns.kdeplot(df['discounted_price'], shade=True, label='Discounted_Price')
  5 plt.xlabel('Price')

File ~\New folder\A\Lib\site-packages\seaborn\distributions.py:1695, in↵
  ↪kdeplot(data, x, y, hue, weights, palette, hue_order, hue_norm, color, fill,↵
  ↪multiple, common_norm, common_grid, cumulative, bw_method, bw_adjust,↵
  ↪warn_singular, log_scale, levels, thresh, gridsize, cut, clip, legend, cbar,↵
  ↪cbar_ax, cbar_kws, ax, **kwargs)
  1692 if ax is None:
  1693     ax = plt.gca()
-> 1695 p._attach(ax, allowed_types=["numeric", "datetime"], log_scale=log_scale)
  1697 method = ax.fill_between if fill else ax.plot
  1698 color = _default_color(method, hue, color, kwargs)

File ~\New folder\A\Lib\site-packages\seaborn\_oldcore.py:1211, in VectorPlotter.
  ↪_attach(self, obj, allowed_types, log_scale)
  1206     if var_type not in allowed_types:
  1207         err = (
  1208             f"The {var} variable is {var_type}, but one of "
  1209             f"{allowed_types} is required"
  1210         )
-> 1211         raise TypeError(err)
  1213 # -- Get axis objects for each row in plot_data for type conversions and↵
  ↪scaling
  1215 facet_dim = {"x": "col", "y": "row"}

TypeError: The x variable is categorical, but one of ['numeric', 'datetime'] is↵
  ↪required
```
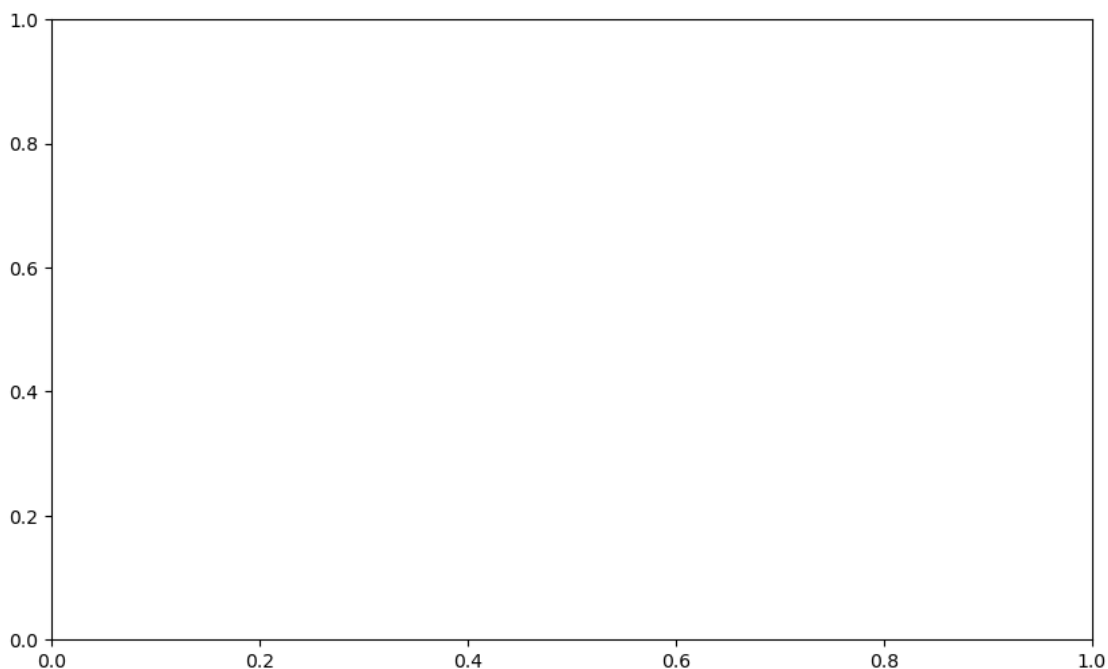
0.4 4. How does the average discount percentage vary across categories?

```
[35]: df
```

```
[35]:        product_id                                        product_name  \
      0      B07JW9H4J1  Wayona Nylon Braided USB to Lightning Fast Cha…
      1      B098NS6PVG  Ambrane Unbreakable 60W / 3A Fast Charging 1.5…
      2      B096MSW6CT  Sounce Fast Phone Charging Cable & Data Sync U…
      3      B08HDJ86NZ  boAt Deuce USB 300 2 in 1 Type-C & Micro USB S…
      4      B08CF3B7N1  Portronics Konnect L 1.2M Fast Charging 3A 8 P…
      …            …                                                 …
      1460   B08L7J3T31  Noir Aqua - 5pcs PP Spun Filter + 1 Spanner | …
      1461   B01M6453MB  Prestige Delight PRWO Electric Rice Cooker (1 …
      1462   B009P2LIL4  Bajaj Majesty RX10 2000 Watts Heat Convector R…
      1463   B00J5DYCCA  Havells Ventil Air DSP 230mm Exhaust Fan (Pist…
      1464   B01486F4G6  Borosil Jumbo 1000-Watt Grill Sandwich Maker (…


                                               category discounted_price  \
      0       Computers&Accessories|Accessories&Peripherals|…              399
      1       Computers&Accessories|Accessories&Peripherals|…              199
      2       Computers&Accessories|Accessories&Peripherals|…              199
      3       Computers&Accessories|Accessories&Peripherals|…              329
      4       Computers&Accessories|Accessories&Peripherals|…              154
      …                                                   …                …
      1460    Home&Kitchen|Kitchen&HomeAppliances|WaterPurif…              379
      1461    Home&Kitchen|Kitchen&HomeAppliances|SmallKitch…            2,280
      1462    Home&Kitchen|Heating,Cooling&AirQuality|RoomHe…            2,219
      1463    Home&Kitchen|Heating,Cooling&AirQuality|Fans|E…            1,399
      1464    Home&Kitchen|Kitchen&HomeAppliances|SmallKitch…            2,863


            actual_price discount_percentage rating rating_count  \
      0            1,099                 64%    4.2       24,269
      1              349                 43%    4.0       43,994
      2            1,899                 90%    3.9        7,928
      3              699                 53%    4.2       94,363
      4              399                 61%    4.2       16,905
      …                …                   …      …            …
      1460           919                 59%      4        1,090
      1461         3,045                 25%    4.1        4,118
      1462         3,080                 28%    3.6          468
      1463         1,890                 26%      4        8,031
      1464         3,690                 22%    4.3        6,987


                                             about_product  \
```

```
0      High Compatibility : Compatible With iPhone 12…
1      Compatible with all Type C enabled devices, be…
2       Fast Charger& Data Sync -With built-in safet…
3      The boAt Deuce USB 300 2 in 1 cable is compati…
4      [CHARGE & SYNC FUNCTION]- This cable comes wit…
…                                                    …
1460   SUPREME QUALITY 90 GRAM 3 LAYER THIK PP SPUN F…
1461                         230 Volts, 400 watts, 1 Year
1462   International design and styling|Two heat sett…
1463   Fan sweep area: 230 MM ; Noise level: (40 - 45…
1464   Brand-Borosil, Specification â€" 23V ~ 5Hz;1 W…


                                              user_id  \
0      AG3D6O4STAQKAY2UVGEUV46KN35Q,AHMY5CWJMMK5BJRBB…
1      AECPFYFQVRUWC3KGNLJIOREFP5LQ,AGYYVPDD7YG7FYNBX…
2      AGU3BBQ2V2DDAMOAKGFAWDDQ6QHA,AESFLDV2PT363T2AQ…
3      AEWAZDZZJLQUYVOVGBEUKSLXHQ5A,AG5HTSFRRE6NL3M5S…
4      AE3Q6KSUK5P75D5HFYHCRAOLODSA,AFUGIFH5ZAFXRDSZH…
…                                                    …
1460   AHITFY6AHALOFOHOZEOC6XBP4FEA,AFRABBODZJZQB6Z4U…
1461   AFG5FM3NEMOL6BNFRV2NK5FNJCHQ,AGEINTRN6Z563RMLH…
1462   AGVPWCMAHYQWJOQKMUJN4DW3KM5Q,AF4Q3E66MY4SR7YQZ…
1463   AF2JQCLSCY3QJATWUNNHUSVUPNQQ,AFDMLUXC5LS5RXDJS…
1464   AFGW5PT3R6ZAVQR4Y5MWVAKBZAYA,AG7QNJ2SCS5VS5VYY…


                                            user_name  \
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…


                                            review_id  \
0      R3HXWTOLRPONMF,R2AJM3LFTLZHFO,R6AQJGUP6P86,R1K…
1      RGIQEGO7R9HS2,R1SMWZQ86XIN8U,R2J3Y1WL29GWDE,RY…
2      R3J3EQQ9TZI5ZJ,R3E7WBGK7IDOKV,RWU79XKQ6I1QF,R2…
3      R3EEUZKKK9J36I,R3HJVYCLYOY554,REDECAZ7AMPQC,R1…
4      R1BP4L2HH9TFUP,R16PVJEXKV6QZS,R2UPDB81N66T4P,R…
…                                                    …
1460   R3G3XFHPBFFOE8,R3COBZCD32EIGW,R2EBVBCN9QPD9R,R…
1461   R3DDL2UPKQ2CK9,R2SYYU1OATVIU5,R1VM993161IYRW,R…
```

```
1462   R1TLRJVW4STY5I,R2O455KRN493R1,R3Q5MVGBRIAS2G,R…
1463   R39Q2Y79MM9SWK,R3079BG1NIH6MB,R29A31ZELTZNJM,R…
1464   R2ORBRZ0WEUJT9,R0KIFK9R2ISSE,R30EEG2FNJSN5I,R2…


                                             review_title  \
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…


                                           review_content  \
0      Looks durable Charging is fine tooNo complains…
1      I ordered this cable to connect my phone to An…
2      Not quite durable and sturdy,https://m.media-a…
3      Good product,long wire,Charges good,Nice,I bou…
4      Bought this instead of original apple, does th…
…                                                       …
1460   I received product without spanner,Excellent p…
1461   ok,got everything as mentioned but the measuri…
1462   plastic but cool body ,u have to find sturdy s…
1463   I have installed this in my kitchen working fi…
1464   It does it job perfectly..only issue is temp c…


                                                 img_link  \
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…


                                             product_link
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-…

[1465 rows x 16 columns]
```

[38]: `df.groupby('discounted_price').mean()`

```
C:\Users\Pooja Rathore\AppData\Local\Temp\ipykernel_4940\844214054.py:1:
FutureWarning: The default value of numeric_only in DataFrameGroupBy.mean is
deprecated. In a future version, numeric_only will default to False. Either
specify numeric_only or select only columns which should be valid for the
function.
  df.groupby('discounted_price').mean()
```

[38]:
```
Empty DataFrame
Columns: []
Index: [ 1,043,  1,049,  1,052,  1,055,  1,059,  1,069,  1,075,  1,089,  1,090,
 1,099,  1,109,  1,110,  1,130,  1,149,  1,180,  1,182,  1,187,  1,189,  1,190,
 1,199,  1,219,  1,220,  1,234,  1,235,  1,249,  1,260,  1,289,  1,290,  1,295,
 1,299,  1,321,  1,324,  1,329,  1,345,  1,349,  1,369,  1,399,  1,400,  1,409,
 1,414,  1,434,  1,439,  1,448,  1,449,  1,456,  1,464,  1,469,  1,474,  1,484,
 1,490,  1,495,  1,498,  1,499,  1,500,  1,519,  1,529,  1,547,  1,549,  1,563,
 1,565,  1,595,  1,598,  1,599,  1,601,  1,614,  1,624,  1,625,  1,649,  1,656,
 1,665,  1,679,  1,695,  1,699,  1,709,  1,745,  1,792,  1,799,  1,804,  1,815,
 1,819,  1,849,  1,850,  1,889,  1,890,  1,898,  1,899,  1,928,  1,959,  1,969,
 1,982.84,  1,989,  1,990,  1,995,  1,998,  1,999,  10,099,  10,389,  10,499,
 10,901,  10,990, …]

[550 rows x 0 columns]
```

[44]: `sns.catplot(x = 'discounted_price', y = 'category', data = df)`

```
C:\Users\Pooja Rathore\New folder\A\Lib\site-packages\seaborn\axisgrid.py:118:
UserWarning: Tight layout not applied. The left and right margins cannot be made
large enough to accommodate all axes decorations.
  self._figure.tight_layout(*args, **kwargs)
```

[44]: `<seaborn.axisgrid.FacetGrid at 0x255f93f0690>`
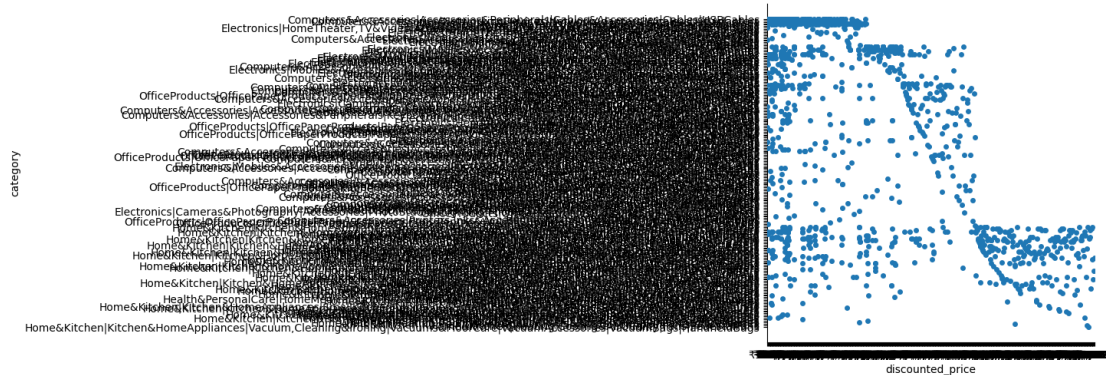
0.5 5. What are the most popular product names?

```
[46]: df.groupby(['product_name'])['rating'].sum()[:1]
```

```
[46]: product_name
      !!1000 Watt/2000-Watt Room Heater!! Fan Heater!!Pure White!!HN-2500!!Made in
      India!!    4.5
      Name: rating, dtype: object
```

0.6 6. What are the most popular product keywords?

```
[ ]: #Fan Heater!!Pure White!!HN-2500!!Made in India
```

0.7 7. What are the most popular product reviews?

```
[53]: df.groupby(['review_title'])['rating'].sum()[:1]
```

```
[53]: review_title
       (after nearly one year use) Value for money TV, except for the sound,Spectacular
      Specification, Build Quality, and Performance    3.9
      Name: rating, dtype: object
```

```
[56]: sns.histplot(df['rating'])
```

```
[56]: <Axes: xlabel='rating', ylabel='Count'>
```
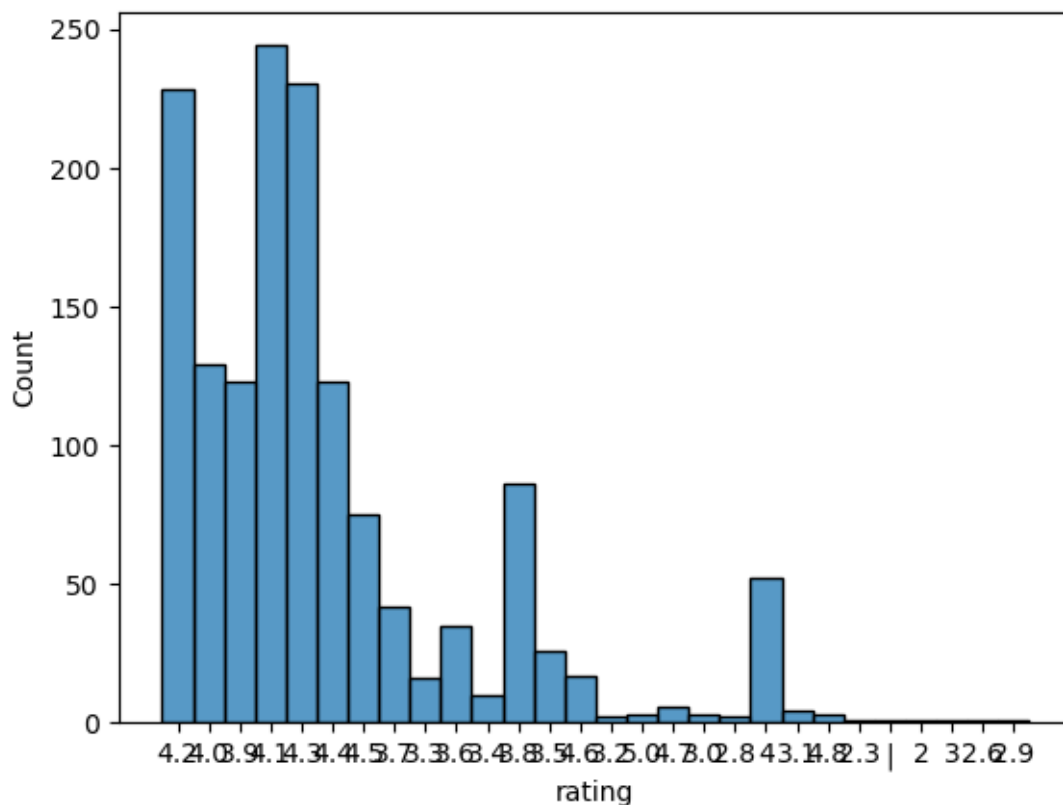
0.8 8. What is the correlation between discounted_price and rating?

```
[61]: 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
```

[67]:
```python
df['discounted_price'] = pd.to_numeric(df['discounted_price'], errors='coerce')
df['rating'] = pd.to_numeric(df['rating'], errors='coerce')
```

[68]:
```python
data=df[['discounted_price','rating']]
data.corr()
```

[68]:
```
                  discounted_price  rating
discounted_price               NaN     NaN
rating                         NaN     1.0
```

[69]:
```python
sns.heatmap(data.corr())
```

[69]: <Axes: >
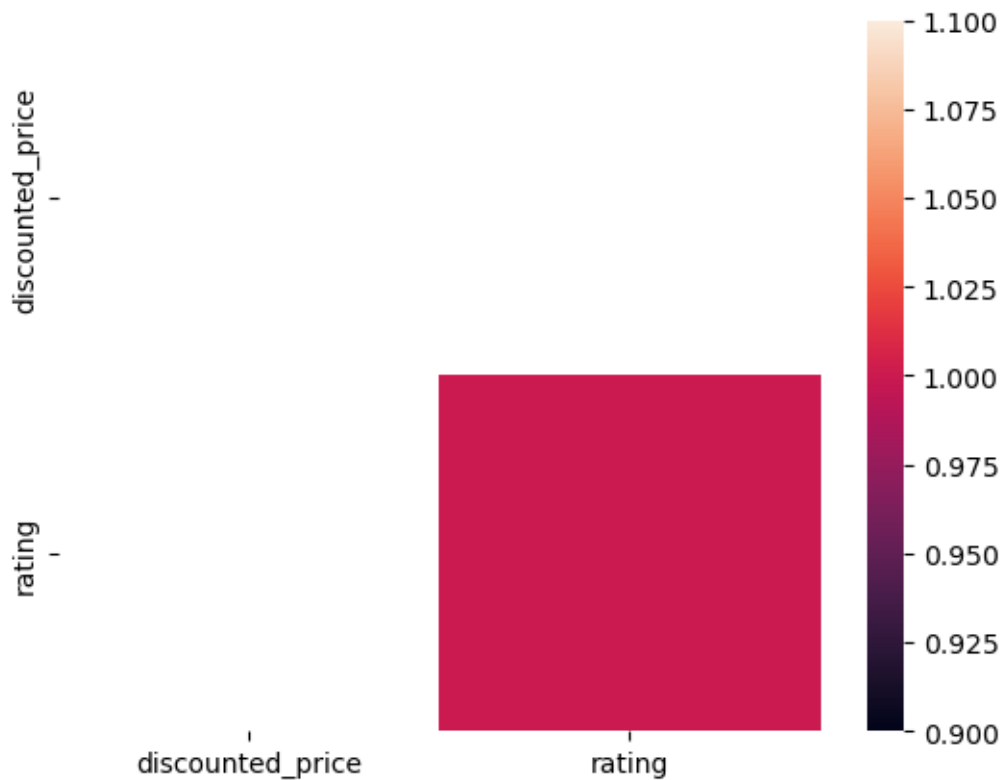
0.9 9. What are the Top 5 categories based on the highest ratings?

```
[85]: df['rating'].value_counts()
```

```
[85]: 4.1    244
      4.3    230
      4.2    228
      4.0    181
      3.9    123
      4.4    123
      3.8     86
      4.5     75
      3.7     42
      3.6     35
      3.5     26
      4.6     17
      3.3     16
      3.4     10
      4.7      6
      3.0      4
      3.1      4
      5.0      3
      4.8      3
      3.2      2
      2.8      2
      2.3      1
      2.0      1
      2.6      1
      2.9      1
      Name: rating, dtype: int64
```

```
[86]: category_ratings = df.groupby('category')['rating'].mean().reset_index()
```

```
[87]: category_ratings
```

```
[87]:                                      category     rating
      0    Car&Motorbike|CarAccessories|InteriorAccessori…  3.800000
      1    Computers&Accessories|Accessories&Peripherals|…  4.150000
      2    Computers&Accessories|Accessories&Peripherals|…  3.500000
      3    Computers&Accessories|Accessories&Peripherals|…  3.600000
      4    Computers&Accessories|Accessories&Peripherals|…  4.050000
      ..                                           …         …
      206  OfficeProducts|OfficePaperProducts|Paper|Stati…  4.250000
      207  OfficeProducts|OfficePaperProducts|Paper|Stati…  4.150000
      208  OfficeProducts|OfficePaperProducts|Paper|Stati…  4.300000
      209  OfficeProducts|OfficePaperProducts|Paper|Stati…  4.133333
      210  Toys&Games|Arts&Crafts|Drawing&PaintingSupplie…  4.300000
```

```
    [211 rows x 2 columns]
```

[88]: `# Sort categories based on average ratings`
`sorted_categories = category_ratings.sort_values(by='rating', ascending=False)`

[89]: `sorted_categories`

[89]:
```
                                        category  rating
57                  Computers&Accessories|Tablets     4.6
48    Computers&Accessories|NetworkingDevices|Networ…     4.5
62     Electronics|Cameras&Photography|Accessories|Film     4.5
81    Electronics|HomeAudio|MediaStreamingDevices|St…     4.5
196   OfficeProducts|OfficeElectronics|Calculators|B…     4.5
..                                           …       …
3     Computers&Accessories|Accessories&Peripherals|…     3.6
88    Electronics|HomeTheater,TV&Video|Accessories|3…     3.5
2     Computers&Accessories|Accessories&Peripherals|…     3.5
14    Computers&Accessories|Accessories&Peripherals|…     3.4
146   Home&Kitchen|Kitchen&HomeAppliances|Coffee,Tea…     3.3

[211 rows x 2 columns]
```

[92]: `sns.histplot(x='sorted_categories')`

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
Cell In[92], line 1
----> 1 sns.histplot(x='sorted_categories')

File ~\New folder\A\Lib\site-packages\seaborn\distributions.py:1395, in
 ↪histplot(data, x, y, hue, weights, stat, bins, binwidth, binrange, discrete,
 ↪cumulative, common_bins, common_norm, multiple, element, fill, shrink, kde,
 ↪kde_kws, line_kws, thresh, pthresh, pmax, cbar, cbar_ax, cbar_kws, palette,
 ↪hue_order, hue_norm, color, log_scale, legend, ax, **kwargs)
   1374 def histplot(
   1375     data=None, *,
   1376     # Vector variables
   (…)
   1392     **kwargs,
   1393 ):
-> 1395     p = _DistributionPlotter(
   1396         data=data,
   1397         variables=_DistributionPlotter.get_semantics(locals())
   1398     )
   1400     p.map_hue(palette=palette, order=hue_order, norm=hue_norm)
   1402     if ax is None:
```

```
File ~\New folder\A\Lib\site-packages\seaborn\distributions.py:113, in␣
 ↪_DistributionPlotter.__init__(self, data, variables)
   107 def __init__(
   108     self,
   109     data=None,
   110     variables={},
   111 ):
--> 113     super().__init__(data=data, variables=variables)


File ~\New folder\A\Lib\site-packages\seaborn\_oldcore.py:640, in VectorPlotter
 ↪.__init__(self, data, variables)
   635 # var_ordered is relevant only for categorical axis variables, and may
   636 # be better handled by an internal axis information object that tracks
   637 # such information and is set up by the scale_* methods. The analogous
   638 # information for numeric axes would be information about log scales.
   639 self._var_ordered = {"x": False, "y": False}  # alt., used DefaultDict
--> 640 self.assign_variables(data, variables)
   642 for var, cls in self._semantic_mappings.items():
   643
   644     # Create the mapping function
   645     map_func = partial(cls.map, plotter=self)


File ~\New folder\A\Lib\site-packages\seaborn\_oldcore.py:701, in VectorPlotter
 ↪assign_variables(self, data, variables)
   699 else:
   700     self.input_format = "long"
--> 701     plot_data, variables = self._assign_variables_longform(
   702         data, **variables,
   703     )
   705 self.plot_data = plot_data
   706 self.variables = variables


File ~\New folder\A\Lib\site-packages\seaborn\_oldcore.py:938, in VectorPlotter
 ↪_assign_variables_longform(self, data, **kwargs)
   933 elif isinstance(val, (str, bytes)):
   934
   935     # This looks like a column name but we don't know what it means!
   937     err = f"Could not interpret value `{val}` for parameter `{key}`"
--> 938     raise ValueError(err)
   940 else:
   941
   942     # Otherwise, assume the value is itself data
   943
   944     # Raise when data object is present and a vector can't matched
   945     if isinstance(data, pd.DataFrame) and not isinstance(val, pd.Series :

ValueError: Could not interpret value `sorted_categories` for parameter `x`
```

```
[90]: # Select top 5 categories
      top_5_categories = sorted_categories.head(5)
```

```
[91]: top_5_categories
```

```
[91]:                                              category  rating
      57                    Computers&Accessories|Tablets     4.6
      48   Computers&Accessories|NetworkingDevices|Networ…     4.5
      62     Electronics|Cameras&Photography|Accessories|Film     4.5
      81     Electronics|HomeAudio|MediaStreamingDevices|St…     4.5
      196  OfficeProducts|OfficeElectronics|Calculators|B…     4.5
```

0.10 10. Identify any potential areas for improvement or optimization based on the data analysis.

```
[1]: # some Some products are very low in a rating so improve rating for that␣
     ↪product .
```

1 assi 2—— spotify data

2 Q 1 1.Load the dataframe and ensure data quality by checking for missing values and duplicate rows.

Handle missing values and remove duplicate rows if necessary

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

```
[3]: df= pd.read_csv(r"C:\Users\Pooja Rathore\Downloads\spotify.csv")
```

```
[4]: df
```

```
[4]:             Artist                              Track Name  \
     0            Drake          Rich Baby Daddy (feat. Sexyy Red & SZA)
     1            Drake                                     One Dance
     2            Drake                           IDGAF (feat. Yeat)
     3            Drake            First Person Shooter (feat. J. Cole)
     4            Drake                 Jimmy Cooks (feat. 21 Savage)
     ..             …                                             …
     435  French Montana                           Splash Brothers
     436        Fat Joe                  All The Way Up (feat. Infared)
     437      A$AP Ferg   Work REMIX (feat. A$AP Rocky, French Montana, …
     438          Diddy                 Another One Of Me (feat. 21 Savage)
     439      Rick Ross                                    Stay Schemin
```

```
     Popularity  Duration (ms)                Track ID
0           92           319191   1yeB8MUNeLo9Ek1UEpsyz6
1           91           173986   1zi7xx7UVEFkmKfv06H8x0
2           90           260111   2YSzYUF3jWqb9YP9VXmpjE
3           88           247444   7aqfrAY2p9BUSiupwk3svU
4           88           218364   3F5CgOj3wFlRv51JsHbxhe
..         ...              ...                      ...
435         44           221863   3fBsEOnzwtlkpSOLxXAZhN
436         64           191900   7Ezwtgfw7khBrpvaNPtMoT
437         69           283693   7xVLFuuYdAvcTfcP3IG3dS
438         65           220408   4hGmQboiou09EwhcTWa0H6
439         68           267720   0nq6sfr8z1R5KJ4XUk396e

[440 rows x 5 columns]
```

[5]: `df.shape`

[5]: `(440, 5)`

[6]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 440 entries, 0 to 439
Data columns (total 5 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Artist         440 non-null    object
 1   Track Name     440 non-null    object
 2   Popularity     440 non-null    int64
 3   Duration (ms)  440 non-null    int64
 4   Track ID       440 non-null    object
dtypes: int64(2), object(3)
memory usage: 17.3+ KB
```

[7]: `df.isnull()`

[7]:
	Artist	Track Name	Popularity	Duration (ms)	Track ID
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False
..	...	...	...	...	...
435	False	False	False	False	False
436	False	False	False	False	False
437	False	False	False	False	False

```
438    False       False       False       False       False
439    False       False       False       False       False

[440 rows x 5 columns]
```

[8]: `df.duplicated()`

```
[8]: 0      False
     1      False
     2      False
     3      False
     4      False
            …
     435     True
     436    False
     437    False
     438    False
     439     True
     Length: 440, dtype: bool
```

[9]: `df[df.duplicated()]`

```
[9]:                 Artist                                 Track Name  \
     39        Travis Scott                     MELTDOWN (feat. Drake)
     52        Travis Scott             TELEKINESIS (feat. SZA & Future)
     72           21 Savage                                    née-nah
     73               Drake              Jimmy Cooks (feat. 21 Savage)
     76               Drake                                  Rich Flex
     131              Drake        First Person Shooter (feat. J. Cole)
     170       Metro Boomin      Trance (with Travis Scott & Young Thug)
     182      D-Block Europe                                   Overseas
     210        Post Malone     Sunflower - Spider-Man: Into the Spider-Verse
     212       Metro Boomin   Annihilate (Spider-Man: Across the Spider-Vers…
     222             Cardi B             WAP (feat. Megan Thee Stallion)
     224             Cardi B          Bongos (feat. Megan Thee Stallion)
     242            Bizarrap        Quevedo: Bzrp Music Sessions, Vol. 52
     270            Lil Durk              All My Life (feat. J. Cole)
     280                 ¥$                                    CARNIVAL
     282       Travis Scott               FE!N (feat. Playboi Carti)
     290            Lil Baby            Drip Too Hard (Lil Baby & Gunna)
     297     Quality Control             Baby (Lil Baby feat. DaBaby)
     310            Lil Nas X          INDUSTRY BABY (feat. Jack Harlow)
     331         Nicki Minaj            Everybody (feat. Lil Uzi Vert)
     341          Snoop Dogg        Young, Wild & Free (feat. Bruno Mars)
     343           Lil Wayne   Sucker for Pain (with Wiz Khalifa, Imagine Dra…
     352         Nicki Minaj   Barbie World (with Aqua) [From Barbie The Album]
     422       Metro Boomin   Annihilate (Spider-Man: Across the Spider-Vers…
```

```
430     French Montana                              Unforgettable
435     French Montana                              Splash Brothers
439          Rick Ross                                Stay Schemin
```

```
     Popularity   Duration (ms)              Track ID
39           86          246133   67nepsnrcZkowTxMWigSbb
52           86          353754   1i9lZvlaDdWDPyXEE95aiq
72           88          220584   2yUzr8Sr6ldG8vmHhZwTnz
73           88          218364   3F5CgOj3wFlRv51JsHbxhe
76           85          239359   1bDbXMyjaUIooNwFE9wn0N
131          88          247444   7aqfrAY2p9BUSiupwk3svU
170          89          194786   5wG3HvLhF6Y5KTGlK0IW3J
182          74          222154   337kcYVjYXdLBItCw9ry3b
210          87          157560   0RiRZpuVRbi7oqRdSMwhQY
212          79          231746   39MK3d3fonIP8Mz9oHCTBB
222          80          187541   4Oun2ylbjFKMPTiaSbbCih
224          78          175099   4YQImHflXSiIMXntcwPkx8
242          87          198937   2tTmW7RDtMQtBk7m2rYeSw
270          74          223878   6T7FXSuXykeGktMLGp8WgE
280          96          264324   3w0w2T288dec0mgeZZqoNN
282          93          191700   42VsgItocQwOQC3XWZ8JNA
290          85          145542   78QR3Wp35dqAhFEc2qAGjE
297          77          142417   5MPPttjfGap2C6j6eKcO6J
310          78          212352   5Z9KJZvQzH6PFmb8SNkxuk
331          84          180869   5ZJGv7aGdIr9IGpxzSG18T
341          70          207346   6Ybhspu0ar1D9WSSnfe7ds
343          77          243490   4dASQiO1Eoo3RJvt74FtXB
352          83          109750   741UUVE2kuITl0c6zuqqbO
422          79          231746   39MK3d3fonIP8Mz9oHCTBB
430          87          233901   3B54sVLJ402zGa6Xm4YGNe
435          44          221863   3fBsEOnzwtlkpSOLxXAZhN
439          68          267720   0nq6sfr8z1R5KJ4XUk396e
```

[10]: `df.drop_duplicates(inplace=True)`

[11]: `df.duplicated()`

[11]:
```
0      False
1      False
2      False
3      False
4      False
       …
433    False
434    False
436    False
437    False
```

```
438    False
Length: 413, dtype: bool
```

2.1 .What is the distribution of popularity among the tracks in the dataset? Visualize it using a histogram.

[12]: `df`

[12]:
```
                Artist                                        Track Name  \
0               Drake           Rich Baby Daddy (feat. Sexyy Red & SZA)
1               Drake                                          One Dance
2               Drake                                  IDGAF (feat. Yeat)
3               Drake                 First Person Shooter (feat. J. Cole)
4               Drake                     Jimmy Cooks (feat. 21 Savage)
..                 …                                                  …
433   French Montana                                        Stand United
434    Jason Derulo                   Tip Toe (feat. French Montana)
436          Fat Joe                  All The Way Up (feat. Infared)
437        A$AP Ferg   Work REMIX (feat. A$AP Rocky, French Montana, …
438            Diddy                 Another One Of Me (feat. 21 Savage)

     Popularity  Duration (ms)                Track ID
0            92         319191  1yeB8MUNeLo9Ek1UEpsyz6
1            91         173986  1zi7xx7UVEFkmKfv06H8x0
2            90         260111  2YSzYUF3jWqb9YP9VXmpjE
3            88         247444  7aqfrAY2p9BUSiupwk3svU
4            88         218364  3F5CgOj3wFlRv51JsHbxhe
..           …              …                       …
433          54         163971  01CHrTerCzyRpMI1MzQ4fz
434          65         187521  0TY3jVGwGDwDabLyQLVRQQ
436          64         191900  7Ezwtgfw7khBrpvaNPtMoT
437          69         283693  7xVLFuuYdAvcTfcP3IG3dS
438          65         220408  4hGmQboiou09EwhcTWa0H6

[413 rows x 5 columns]
```

[13]: `df.isnull()`

[13]:
```
      Artist  Track Name  Popularity  Duration (ms)  Track ID
0      False       False       False          False     False
1      False       False       False          False     False
2      False       False       False          False     False
3      False       False       False          False     False
4      False       False       False          False     False
..         …           …           …              …         …
433    False       False       False          False     False
434    False       False       False          False     False
```

```
436   False        False        False        False        False
437   False        False        False        False        False
438   False        False        False        False        False

[413 rows x 5 columns]
```
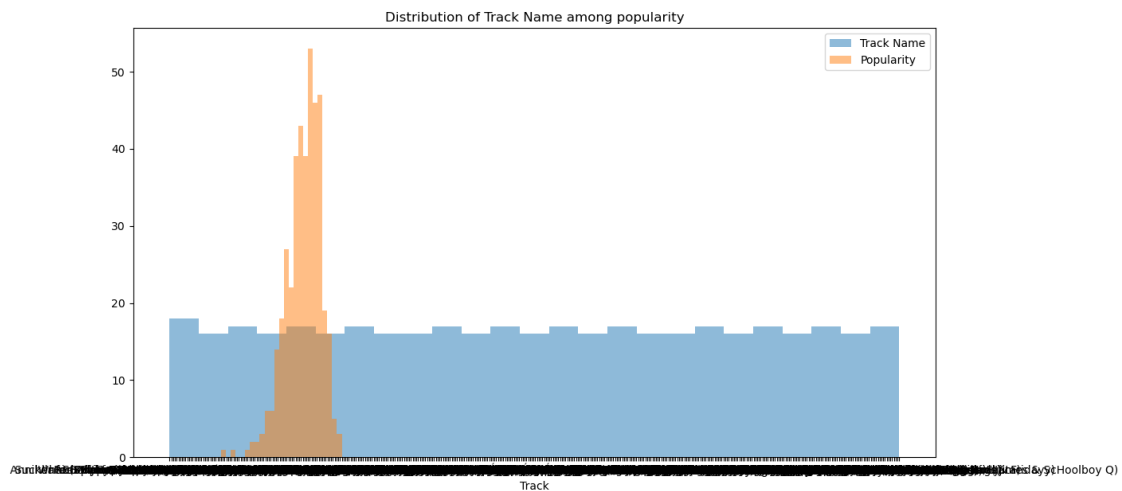
```
[14]: plt.figure(figsize=(13, 7))
      plt.hist(df['Track Name'], bins=25, alpha=0.5, label='Track Name')
      plt.hist(df['Popularity'], bins=25, alpha=0.5, label='Popularity')
      plt.xlabel('Track')
      plt.title('Distribution of Track Name among popularity')
      plt.legend()
      plt.show()
```


2.2 Q.3 Is there any relationship between the popularity and the duration of tracks? Explore this using a scatter plot.

```
[15]: df
```

```
[15]:            Artist                                    Track Name  \
      0           Drake          Rich Baby Daddy (feat. Sexyy Red & SZA)
      1           Drake                                       One Dance
      2           Drake                               IDGAF (feat. Yeat)
      3           Drake              First Person Shooter (feat. J. Cole)
      4           Drake                     Jimmy Cooks (feat. 21 Savage)
      ..            …                                                 …
      433  French Montana                                   Stand United
      434   Jason Derulo               Tip Toe (feat. French Montana)
      436        Fat Joe               All The Way Up (feat. Infared)
      437      A$AP Ferg  Work REMIX (feat. A$AP Rocky, French Montana, …
```

```
438              Diddy             Another One Of Me (feat. 21 Savage)
```

```
     Popularity  Duration (ms)                 Track ID
0            92         319191   1yeB8MUNeLo9Ek1UEpsyz6
1            91         173986   1zi7xx7UVEFkmKfv06H8x0
2            90         260111   2YSzYUF3jWqb9YP9VXmpjE
3            88         247444   7aqfrAY2p9BUSiupwk3svU
4            88         218364   3F5CgOj3wFlRv51JsHbxhe
..          ...            ...                      ...
433          54         163971   01CHrTerCzyRpMI1MzQ4fz
434          65         187521   0TY3jVGwGDwDabLyQLVRQQ
436          64         191900   7Ezwtgfw7khBrpvaNPtMoT
437          69         283693   7xVLFuuYdAvcTfcP3IG3dS
438          65         220408   4hGmQboiou09EwhcTWaOH6

[413 rows x 5 columns]
```
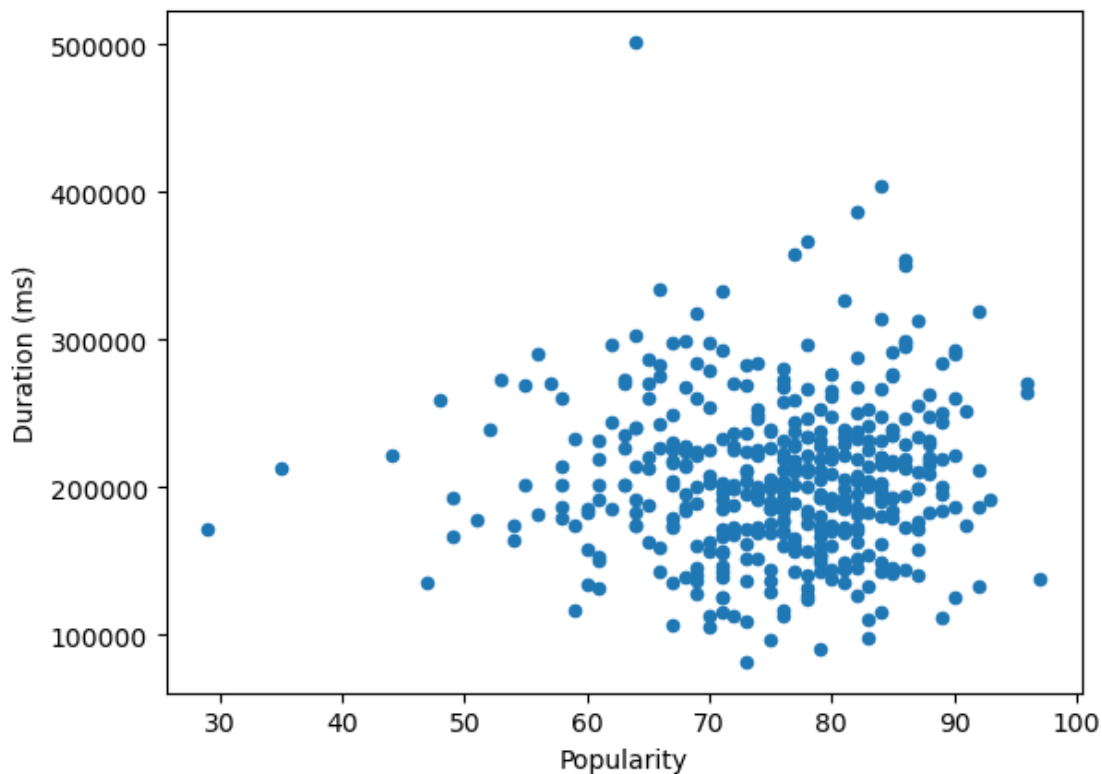
[ ]:

[23]:
```python
plt.figure(figsize=(8,8))
df.plot.scatter(x='Popularity',y='Duration (ms)')
```

[23]: <Axes: xlabel='Popularity', ylabel='Duration (ms)'>

<Figure size 800x800 with 0 Axes>

2.3 Q 4.Which artist has the highest number of tracks in the dataset? Display the count of tracks for each artist

2.4 using a countplot.

```
[24]: df
```

```
[24]:              Artist                                 Track Name  \
      0             Drake          Rich Baby Daddy (feat. Sexyy Red & SZA)
      1             Drake                                       One Dance
      2             Drake                              IDGAF (feat. Yeat)
      3             Drake             First Person Shooter (feat. J. Cole)
      4             Drake                    Jimmy Cooks (feat. 21 Savage)
      ..              …                                               …
      433  French Montana                                     Stand United
      434    Jason Derulo                   Tip Toe (feat. French Montana)
      436         Fat Joe                   All The Way Up (feat. Infared)
      437      A$AP Ferg  Work REMIX (feat. A$AP Rocky, French Montana, …
      438           Diddy                 Another One Of Me (feat. 21 Savage)

           Popularity  Duration (ms)               Track ID
      0            92         319191  1yeB8MUNeLo9Ek1UEpsyz6
```

38

```
1        91        173986   1zi7xx7UVEFkmKfv06H8x0
2        90        260111   2YSzYUF3jWqb9YP9VXmpjE
3        88        247444   7aqfrAY2p9BUSiupwk3svU
4        88        218364   3F5CgOj3wFlRv51JsHbxhe
..       …         …                    …
433      54        163971   01CHrTerCzyRpMI1MzQ4fz
434      65        187521   0TY3jVGwGDwDabLyQLVRQQ
436      64        191900   7Ezwtgfw7khBrpvaNPtMoT
437      69        283693   7xVLFuuYdAvcTfcP3IG3dS
438      65        220408   4hGmQboiou09EwhcTWa0H6

[413 rows x 5 columns]
```

[27]:
```python
artist_counts=df['Artist'].value_counts()
```

[29]:
```python
artist_counts
```

[29]:
```
Drake             17
Eminem            11
Flo Rida          10
Ludacris          10
Timbaland         10
                  ..
Arizona Zervas     1
Fivio Foreign      1
Pressa             1
David Guetta       1
Diddy              1
Name: Artist, Length: 115, dtype: int64
```
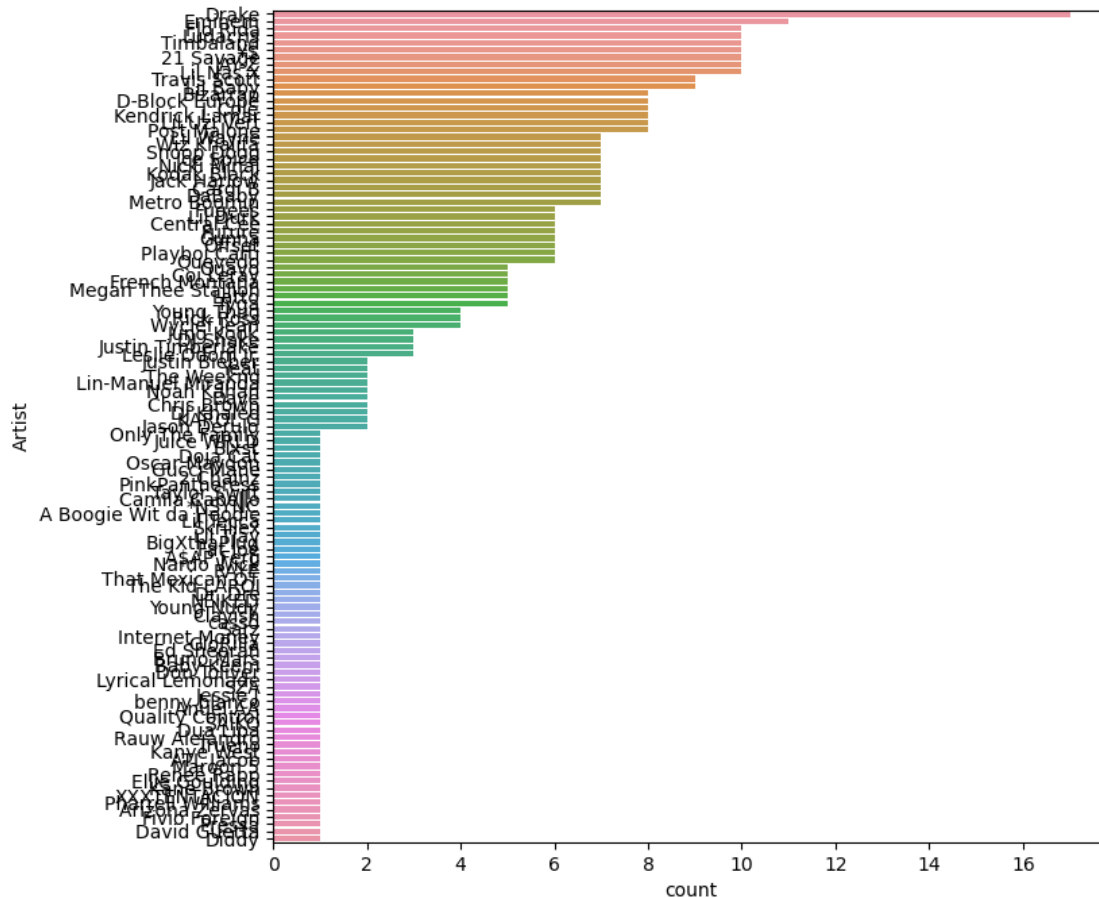
[28]:
```python
artist_counts.max()
```

[28]: 17

[32]:
```python
plt.figure(figsize=(8,8))
sns.countplot(y='Artist',data=df,order=artist_counts.index)
plt.title('Count of Tracks for Each Artist')
plt.xlabel('Count of Tracks')
plt.ylabel('Artist')
plt.show()
```

[32]: <Axes: xlabel='count', ylabel='Artist'>

2.5 5.What are the top 5 least popular tracks in the dataset? Provide the artist name and track name for each.

```
[33]: df
```

```
[33]:            Artist                              Track Name  \
      0           Drake        Rich Baby Daddy (feat. Sexyy Red & SZA)
      1           Drake                                     One Dance
      2           Drake                             IDGAF (feat. Yeat)
      3           Drake          First Person Shooter (feat. J. Cole)
      4           Drake               Jimmy Cooks (feat. 21 Savage)
      ..            …                                            …
      433  French Montana                                  Stand United
      434    Jason Derulo            Tip Toe (feat. French Montana)
      436         Fat Joe             All The Way Up (feat. Infared)
      437        A$AP Ferg  Work REMIX (feat. A$AP Rocky, French Montana, …
      438           Diddy           Another One Of Me (feat. 21 Savage)
```

```
      Popularity  Duration (ms)                    Track ID
0            92         319191    1yeB8MUNeLo9Ek1UEpsyz6
1            91         173986    1zi7xx7UVEFkmKfv06H8x0
2            90         260111    2YSzYUF3jWqb9YP9VXmpjE
3            88         247444    7aqfrAY2p9BUSiupwk3svU
4            88         218364    3F5CgOj3wFlRv51JsHbxhe
..          ...            ...                       ...
433          54         163971    01CHrTerCzyRpMI1MzQ4fz
434          65         187521    0TY3jVGwGDwDabLyQLVRQQ
436          64         191900    7Ezwtgfw7khBrpvaNPtMoT
437          69         283693    7xVLFuuYdAvcTfcP3IG3dS
438          65         220408    4hGmQboiou09EwhcTWa0H6

[413 rows x 5 columns]
```

[37]:
```python
top_artist=df.sort_values(by='Popularity').head(5)
```

[38]:
```python
top_artist
```

[38]:
```
            Artist                   Track Name  Popularity  Duration (ms)  \
207         Pressa  Attachments (feat. Coi Leray)          29         171000
231  Justin Bieber                     Intentions          35         212853
413  French Montana                 Splash Brothers          44         221863
225        Lil Baby                   On Me - Remix          47         135444
407     Wyclef Jean      911 (feat. Mary J. Blige)          48         259333

                   Track ID
207  6EfJ8Ct3GHbBz2YlyPnMYb
231  7jAVt70Xdg8EwOtsFB1ZqK
413  3fBsEOnzwtlkpSOLxXAZhN
225  2uUFVnVFERNxUdcmvEs7LB
407  28hgx2bWXcaBJeC9zVwrBq
```

2.6 6. Among the top 5 most popular artists, which artist has the highest popularity on average?

Calculate and display the average popularity for each artist.

[54]:
```python
top_artist=df.sort_values(by='Popularity')[:5]
```

[55]:
```python
top_artist
```

[55]:
```
            Artist                   Track Name  Popularity  Duration (ms)  \
207         Pressa  Attachments (feat. Coi Leray)          29         171000
231  Justin Bieber                     Intentions          35         212853
413  French Montana                 Splash Brothers          44         221863
225        Lil Baby                   On Me - Remix          47         135444
```

```
407     Wyclef Jean        911 (feat. Mary J. Blige)          48          259333
```

```
                          Track ID
207   6EfJ8Ct3GHbBz2YlyPnMYb
231   7jAVt70Xdg8EwOtsFB1ZqK
413   3fBsEOnzwtlkpSOLxXAZhN
225   2uUFVnVFERNxUdcmvEs7LB
407   28hgx2bWXcaBJeC9zVwrBq
```

[57]: `pop_avg=df.groupby('Artist')['Popularity'].mean()`

[67]: `pop_avg`

[67]:
```
Artist
*NSYNC                      67.00
2 Chainz                    72.00
21 Savage                   83.80
A Boogie Wit da Hoodie      80.00
A$AP Ferg                   69.00
                            …
Young Nudy                  67.00
Young Thug                  73.75
benny blanco                72.00
cassö                       92.00
¥$                          85.10
Name: Popularity, Length: 115, dtype: float64
```
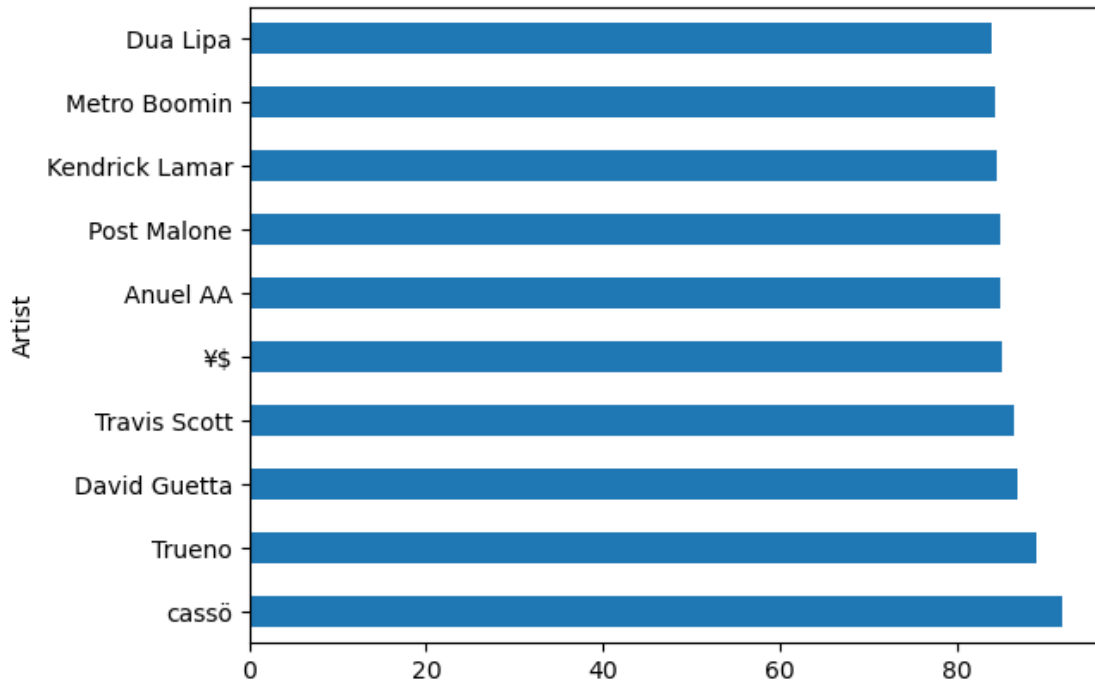
[69]: `top_5=pop_avg.nlargest(10)`

[70]: `top_5`

[70]:
```
Artist
cassö            92.000000
Trueno           89.000000
David Guetta     87.000000
Travis Scott     86.555556
¥$               85.100000
Anuel AA         85.000000
Post Malone      85.000000
Kendrick Lamar   84.500000
Metro Boomin     84.428571
Dua Lipa         84.000000
Name: Popularity, dtype: float64
```

[80]:
```
top_5.plot.barh()
plt.show()
```

2.7 7.For the top 5 most popular artists, what are their most popular tracks? List the track name for each artist

```
[75]: top_artist=df.sort_values(by='Popularity')[:5]
```

```
[76]: top_artist
```

```
[76]:            Artist                     Track Name  Popularity  Duration (ms)  \
      207         Pressa  Attachments (feat. Coi Leray)          29         171000
      231  Justin Bieber                      Intentions          35         212853
      413  French Montana                  Splash Brothers          44         221863
      225        Lil Baby                     On Me - Remix          47         135444
      407     Wyclef Jean          911 (feat. Mary J. Blige)       48         259333

                        Track ID
      207  6EfJ8Ct3GHbBz2YlyPnMYb
      231  7jAVt7OXdg8EwOtsFB1ZqK
      413  3fBsEOnzwt1kpSOLxXAZhN
      225  2uUFVnVFERNxUdcmvEs7LB
      407  28hgx2bWXcaBJeC9zVwrBq
```

```
[82]: df.groupby('Artist').count().
      ↪sort_values(by="Popularity",ascending=False)['Popularity'][:10]
```

43

```
[82]: Artist
      Drake           17
      Eminem          11
      ¥$              10
      21 Savage       10
      Timbaland       10
      Ludacris        10
      Lil Nas X       10
      JAY-Z           10
      Flo Rida        10
      Travis Scott     9
      Name: Popularity, dtype: int64
```

```python
[85]: df[['Artist','Popularity']].sort_values(by='Popularity',ascending=True)[:5]
```

```
[85]:            Artist  Popularity
      207         Pressa          29
      231  Justin Bieber          35
      413  French Montana          44
      225       Lil Baby          47
      407    Wyclef Jean          48
```

```python
[ ]:
```