

## 23rd-march-eda-assignment-2

June 15, 2024

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

```
[123]: data=pd.read_csv("C:\\Users\\Malleesh\\Downloads\\amazon.csv")
```

```
[124]: data
```

```
[124]:
```

	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 PRW0 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	AG3D604STAQKAY2UVGEUV46KN35Q,AHMY5CWJMMK5BJRBB...
1	AECPFYFQVRUWC3KGNLJIOREFP5LQ,AGYYVPDD7YG7FYNBX...
2	AGU3BBQ2V2DDAMOAKGFAWDDQ6QHA,AESFLDV2PT363T2AQ...
3	AEWAZDZZJLQUYVOVGBEUKSLXHQA,AG5HTSFRRE6NL3M5S...
4	AE3Q6KSUK5P75D5HFYHCRAOLODSA,AFUGIFH5ZAFXRDSZH...
...	...
1460	AHITFY6AHALOFOHOZEOC6XBP4FEA,AFRABBODZJZQB6Z4U...
1461	AFG5FM3NEMOL6BNFRV2NK5FNJCHQ,AGEINTRN6Z563RMLH...
1462	AGVPWCMAHYQWJOQKMUJN4DW3KM5Q,AF4Q3E66MY4SR7YQZ...
1463	AF2JQCLSCY3QJATWUNNHUSVUPNQ,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      RGIQEG07R9HS2,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,R2SYYU10ATVIU5,R1VM993161IYRW,R...
1462   R1TLRJVV4STY5I,R20455KRN493R1,R3Q5MVBGBRIAS2G,R...
1463   R39Q2Y79MM9SWK,R3079BG1NIH6MB,R29A31ZELTZNMJ,R...
1464   R2ORBRZOWEJUT9,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-Konnnect-POR-1...
...
1460   https://www.amazon.in/Noir-Aqua-Spanner-Purifi...
1461   https://www.amazon.in/Prestige-Delight-PRW0-1-...
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]

```
[125]: data.head()
```

```

[125]:   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...

```

```

                                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

```

```

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

```

```

                                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...

user\_id \

0 AG3D604STAQKAY2UVGEUV46KN35Q,AHMY5CWJMMK5BJRBB...  
 1 AECPPFYFQVRUWC3KGNLJIOREFP5LQ,AGYYVPDD7YG7FYNBX...  
 2 AGU3BBQ2V2DDAMOAKGFAWDDQ6QHA,AESFLDV2PT363T2AQ...  
 3 AEWAZDZZJLQUYVOVGBEUKSLXHQ5A,AG5HTSFRRE6NL3M5S...  
 4 AE3Q6KSUK5P75D5HIFYHCRAOLODSA,AFUGIFH5ZAFXRDSZH...

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...

review\_id \

0 R3HXWTOLRPONMF,R2AJM3LFTLZHFO,R6AQJGUP6P86,R1K...  
 1 RGIQEG07R9HS2,R1SMWZQ86XIN8U,R2J3Y1WL29GWDE,RY...  
 2 R3J3EQQ9TZI5ZJ,R3E7WBGK7IDOKV,RWU79XKQ6I1QF,R2...  
 3 R3EEUZKKK9J36I,R3HJVYCLY0Y554,REDECAZ7AMPQC,R1...  
 4 R1BP4L2HH9TFUP,R16PVJEXKV6QZS,R2UPDB81N66T4P,R...

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...

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...

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...

```

                                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...

```

```
[126]: data.dtypes
```

```

[126]: 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

```

```
[127]: data.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

```

```
[128]: data.columns
```

```
[128]: 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')
```

```
[129]: data.shape
```

```
[129]: (1465, 16)
```

```
[130]: data.isna()
```

```
[130]:
```

	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]

```
[131]: data.isna().sum()
```

```
[131]: 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
```



```
[132]: df=data.dropna()
```

```
[133]: df
```

```
[133]:
```

	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 PRW0 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 AG3D604STAQKAY2UVGEUV46KN35Q,AHMY5CWJMMK5BJRBB...  
 1 AECPFYFQVRUWC3KGNLJIOREFP5LQ,AGYYVPDD7YG7FYNBX...  
 2 AGU3BBQ2V2DDAMOAKGFAWDDQ6QHA,AESFLDV2PT363T2AQ...  
 3 AEWAZDZZJLQUYVOVGBEUKSLXHQ5A,AG5HTSFRRE6NL3M5S...  
 4 AE3Q6KSUK5P75D5HFYHCRAOLODSA,AFUGIFH5ZAFXRDSZH...  
 ...  
 1460 AHITFY6AHALOFOHOZEOC6XBP4FEA,AFRABBODZJZQB6Z4U...  
 1461 AFG5FM3NEMOL6BNFRV2NK5FNJCHQ,AGEINTRN6Z563RMLH...  
 1462 AGVPWCMahyQWJQKMUN4DW3KM5Q,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 R3HXWTO LRPNMF,R2AJM3LFTLZHF0,R6AQJGUP6P86,R1K...  
 1 RGIQEG07R9HS2,R1SMWZQ86XIN8U,R2J3Y1WL29GWDE,RY...  
 2 R3J3EQQ9TZI5ZJ,R3E7WBGK7IDOKV,RWU79XKQ6I1QF,R2...  
 3 R3EEUZKKK9J36I,R3HJVYCLYOY554,REDECAZ7AMPQC,R1...  
 4 R1BP4L2HH9TFUP,R16PVJEXKV6QZS,R2UPDB81N66T4P,R...  
 ...  
 1460 R3G3XFHPBF0E8,R3COBZCD32EIGW,R2EBVBCN9QPD9R,R...  
 1461 R3DDL2UPKQ2CK9,R2SYYU10ATVIU5,R1VM993161IYRW,R...  
 1462 R1TLRJVV4STY5I,R2O455KRN493R1,R3Q5MVBGRIAS2G,R...  
 1463 R39Q2Y79MM9SWK,R3O79BG1NIH6MB,R29A31ZELTZNJM,R...  
 1464 R2ORBRZOWEIJT9,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	<a href="https://m.media-amazon.com/images/W/WEBP_40237...">https://m.media-amazon.com/images/W/WEBP_40237...</a>
1	<a href="https://m.media-amazon.com/images/W/WEBP_40237...">https://m.media-amazon.com/images/W/WEBP_40237...</a>
2	<a href="https://m.media-amazon.com/images/W/WEBP_40237...">https://m.media-amazon.com/images/W/WEBP_40237...</a>
3	<a href="https://m.media-amazon.com/images/I/41V5FtEWPk...">https://m.media-amazon.com/images/I/41V5FtEWPk...</a>
4	<a href="https://m.media-amazon.com/images/W/WEBP_40237...">https://m.media-amazon.com/images/W/WEBP_40237...</a>
...	...
1460	<a href="https://m.media-amazon.com/images/I/41fDdRtjfx...">https://m.media-amazon.com/images/I/41fDdRtjfx...</a>
1461	<a href="https://m.media-amazon.com/images/I/41gzDxk4+k...">https://m.media-amazon.com/images/I/41gzDxk4+k...</a>
1462	<a href="https://m.media-amazon.com/images/W/WEBP_40237...">https://m.media-amazon.com/images/W/WEBP_40237...</a>
1463	<a href="https://m.media-amazon.com/images/W/WEBP_40237...">https://m.media-amazon.com/images/W/WEBP_40237...</a>
1464	<a href="https://m.media-amazon.com/images/W/WEBP_40237...">https://m.media-amazon.com/images/W/WEBP_40237...</a>

	product_link
0	<a href="https://www.amazon.in/Wayona-Braided-WN3LG1-Sy...">https://www.amazon.in/Wayona-Braided-WN3LG1-Sy...</a>
1	<a href="https://www.amazon.in/Ambrane-Unbreakable-Char...">https://www.amazon.in/Ambrane-Unbreakable-Char...</a>
2	<a href="https://www.amazon.in/Sounce-iPhone-Charging-C...">https://www.amazon.in/Sounce-iPhone-Charging-C...</a>
3	<a href="https://www.amazon.in/Deuce-300-Resistant-Tang...">https://www.amazon.in/Deuce-300-Resistant-Tang...</a>
4	<a href="https://www.amazon.in/Portronics-Konnect-POR-1...">https://www.amazon.in/Portronics-Konnect-POR-1...</a>
...	...

```

1460 https://www.amazon.in/Noir-Aqua-Spanner-Purifi...
1461 https://www.amazon.in/Prestige-Delight-PRW0-1-...
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-...

```

```
[1463 rows x 16 columns]
```

```
[134]: df.isna().sum()
```

```

[134]: product_id      0
       product_name    0
       category        0
       discounted_price 0
       actual_price     0
       discount_percentage 0
       rating          0
       rating_count     0
       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

```

```

[135]: #1 Check for non-numeric values
print(df['rating'].unique())

# Convert the 'rating' column to numeric, coercing errors
df['rating'] = pd.to_numeric(df['rating'], errors='coerce')

# Drop rows with NaN ratings (or handle them as needed)
df= df.dropna(subset=['rating'])

# Group by 'category' and calculate the mean of 'rating'
average_ratings = df.groupby('category')['rating'].mean().reset_index()

# Rename columns for clarity
average_ratings.columns = ['category', 'average_rating']

print(average_ratings)

```

```

['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']
```

	category	average_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]
```

```
C:\Users\Mallesha\AppData\Local\Temp\ipykernel_2172\102299873.py:5:
```

```
SettingWithCopyWarning:
```

```
A value is trying to be set on a copy of a slice from a DataFrame.
```

```
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
df['rating'] = pd.to_numeric(df['rating'], errors='coerce')
```

```
[136]: #2 What are the top rating_count products by category?
# Ensure 'rating_count' is numeric
df['rating_count'] = pd.to_numeric(df['rating_count'], errors='coerce')

# Drop rows where 'rating_count' is NaN
df = df.dropna(subset=['rating_count'])

# Get the index of the top products by rating_count for each category
idx = df.groupby('category')['rating_count'].idxmax()

# Use these indices to get the top products
top_products_by_category = df.loc[idx]

# Display the results
top_products_by_category
```

```
C:\Users\Mallesha\AppData\Local\Temp\ipykernel_2172\3035907703.py:3:
```

```
SettingWithCopyWarning:
```

```
A value is trying to be set on a copy of a slice from a DataFrame.
```

```
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
df['rating_count'] = pd.to_numeric(df['rating_count'], errors='coerce')
```

```
[136]:
```

	product_id	product_name \
900	B09F3PDDRF	Lapster USB 3.0 sata Cable for 2.5 inch SSD an...
32	B08QSC1XY8	Zoul USB C 60W Fast Charging 3A 6ft/2M Long Ty...
743	B0BBMPH39N	Amazon Basics Magic Slate 8.5-inch LCD Writing...
653	B0B9LDCX89	STRIFF Mpad Mouse Mat 230X190X3mm Gaming Mouse...
940	B09GBBJV72	HP 330 Wireless Black Keyboard and Mouse Set w...
...	...	...
1257	B08HDCWDXD	Eureka Forbes Active Clean 700 Watts Powerful ...
1317	B0B7FJNSZR	Proven® Copper + Mineral RO+UV+UF 10 to 12 Lit...
1424	B09BL2KHQW	KENT POWP-Sediment Filter 10'' Thread WCAP
808	B095X38CJS	BRUSTRO Copytinta Coloured Craft Paper A4 Size...
950	B099S26HWG	Classmate Pulse 1 Subject Notebook - 240mm x 1...

	category	discounted_price \
900	Computers&Accessories Accessories&Peripherals ...	349
32	Computers&Accessories Accessories&Peripherals ...	389
743	Computers&Accessories Accessories&Peripherals ...	289
653	Computers&Accessories Accessories&Peripherals ...	129
940	Computers&Accessories Accessories&Peripherals ...	1,409
...	...	...
1257	Home&Kitchen Kitchen&HomeAppliances Vacuum,Cle...	3,179
1317	Home&Kitchen Kitchen&HomeAppliances WaterPurif...	5,395
1424	Home&Kitchen Kitchen&HomeAppliances WaterPurif...	231
808	OfficeProducts OfficePaperProducts Paper Copy&...	99
950	OfficeProducts OfficePaperProducts Paper Stati...	300

	actual_price	discount_percentage	rating	rating_count \
900	999	65%	3.9	817.0
32	1,099	65%	4.3	974.0
743	999	71%	4.1	401.0
653	999	87%	4.2	491.0
940	2,199	36%	3.9	427.0
...	...	...	...	...
1257	6,999	55%	4.0	743.0
1317	19,990	73%	4.4	535.0
1424	260	11%	4.1	490.0
808	99	0%	4.3	388.0
950	300	0%	4.2	419.0

	about_product \
900	Sata cable is Powered by ASM Chipset bridge wh...
32	{3A/QC 3.0 FAST CHARGING and DATA SYNC} : This...
743	LCD screen without glare, radiation, and blue ...
653	9.4 Inches X 7.9 Inches) 240mm x 200mm x 2mm S...
940	Num pad and common controls are integrated wit...

...  
1257 Note-This vacuum cleaner comes with an auto sh...  
1317 Color: Black and Copper, Capacity: 10-12 liter...  
1424 Sediment filter 10 inch Kent  
808 Made at the renowned Fabriano paper mill in It...  
950 The cover design of the notebook is subject to...

user\_id \  
900 AG726NQT4GKLFNXFOAQBFA6JQGQ,AHAI3QJ4Q5GQAQRQT...  
32 AHMKXORT3VNMB75C3EUBYMFYELFQ,AEKJRELVNMICYPOYT...  
743 AEREPHGDEP7EUMBCNLBQ67GNJQMA,AHLOJPA3DIPNKLDZV...  
653 AGASWLGAJEYSNHPWSR74GSDXU5JQ,AGFT22PQSW5ZDJLFL...  
940 AETHN2CGVNPVX5Y6SAW06IO7QOE,AHWRHOUQWASZTLTB4...

...  
1257 AHFILHSL3P3VABTMFUYKAWTNUWVQ,AHLGS6FCOG5AUP23U...  
1317 AG6ST6L57J4B7UHNXKEV55ZP3NPQ,AHJYH6BZ2SXIWIEUP...  
1424 AFJLDRIDWU5X34BNJZSWOG3FHLRA,AHXKIRJTURRXQ7DQD...  
808 AG7MI6MZP3GMUTO65QNUR25VP7VA,AHBZCVGYU7B07DCSN...  
950 AFXZNVON4LZKKL23DAL7IPT5ZJUA,AFSZKLP705V7SHFKR...

user\_name \  
900 Vignesh Prabu VP,Sachine,Shobhit Chaurasiya,Ku...  
32 Pratyush Pahuja,Thiyagarajan,yeses,Amazon Cust...  
743 Amazon Customer,Prince kumar,SK,Nishikanta Sah...  
653 muhamad s.,Aruna V,Ankush Jindal,SJ,Ronak,Deve...  
940 Selva sibi,TZ,c d johnson rajakumar,Credit Mav...

...  
1257 shiny,Placeholder,Karunakar,Veena rastogi,Chet...  
1317 Bobby,Arunkumar,aditya,sanket,srinivasa m,Sand...  
1424 Vasu P,KRISHNA MOHAN LANKA,BISWAJIT KUILA,Sai ...  
808 kashif,sezhian,Dinesh K.,Jenifer,Harsh,suresh,...  
950 Ayush pandey,Puneet Sharma,Lakshay,Vala 18,Jel...

review\_id \  
900 R1CJOMB11B1FIY,RIDJYDQN13E73,R34VA5BFT3PL9D,R1...  
32 R2S0AYWUV349HP,R35OW9CYQNAYHY,R3B3DDF1D5NULK,R...  
743 R37B1CGX8LWLNS,RUFFDMZAXAVOE,R1FUPYVQBR42LV,R2...  
653 R3ET8JTEIDTNUO,R1FAH4M3BSL55F,R3I8GGSZJCEUGV,R...  
940 R2RDB07DGL4GM9,R3H2WY92CQUJMX,R2LDUGW3VRNHAB,R...

...  
1257 R1EOXYGHB00B9,R2MQLUR661FKOA,R235YDZ5Q9LII7,R...  
1317 R1BRNGXN1P2SNY,R1MLFH XV5FZHKJ,R3JDJSYI7QMMXW,R...  
1424 R2MP3ZHMZJIHP0,RMTBPDSRHU000,R1ZJ2RU3C1TION,R3...  
808 R2ETD6AVA4AFF1,R3CUOW1DUP8N92,R3NYD5UBRVJDWP,R...  
950 R3I568NPPF5187,R19KS9NAHZME09,R384JBLG7VAYNP,R...

review\_title \

900 Compatible with laptop SSD,Good to connect say...  
 32 Great Cable, Charging Speeds Could Be Better,G...  
 743 I like to draw on it,Best writing pad,Great f...  
 653 Decent quality,Good for the price,Value buy,It...  
 940 Good,Excellent quality but mouse is small for ...  
 ...  
 1257 Ok but not for deep cleaning,Just ok. Trips fr...  
 1317 Gud product and gud service,Good product in th...  
 1424 Good product,Delivery at door step,Price is gr...  
 808 Good,Heads up,Good product,Bright in color go...  
 950 Pages size is small but good quality,Okay,Qual...

#### review\_content \

900 It works great for my laptop,This is a good ac...  
 32 Not charging as fast as I'd have expected. May...  
 743 I use it for writing messages or drawing.. the...  
 653 Mousepad is sturdy enough doesn't move around ...  
 940 Mouse open cover was little less quality,Excel...  
 ...  
 1257 Handy but no carpet cleaning. Stops in 10-15 m...  
 1317 Like the product with decent look,Newely insta...  
 1424 Value for money.,Good product,Printed price is...  
 808 Good value of money if u have students at home...  
 950 Pages are small,Okay okay,Best product but siz...

#### img\_link \

900 [https://m.media-amazon.com/images/W/WEBP\\_40237...](https://m.media-amazon.com/images/W/WEBP_40237...)  
 32 <https://m.media-amazon.com/images/I/419QKVTxaS...>  
 743 [https://m.media-amazon.com/images/W/WEBP\\_40237...](https://m.media-amazon.com/images/W/WEBP_40237...)  
 653 [https://m.media-amazon.com/images/W/WEBP\\_40237...](https://m.media-amazon.com/images/W/WEBP_40237...)  
 940 [https://m.media-amazon.com/images/W/WEBP\\_40237...](https://m.media-amazon.com/images/W/WEBP_40237...)  
 ...  
 1257 <https://m.media-amazon.com/images/I/41zqeckaQt...>  
 1317 <https://m.media-amazon.com/images/I/41EzVyKoA0...>  
 1424 <https://m.media-amazon.com/images/I/41ugz3c3G1...>  
 808 <https://m.media-amazon.com/images/I/41WggyozHQ...>  
 950 <https://m.media-amazon.com/images/I/51h6eqwfeP...>

#### product\_link

900 <https://www.amazon.in/LAPSTER-SATA-CABLE-LAPTO...>  
 32 <https://www.amazon.in/Charging-Braided-Charger...>  
 743 <https://www.amazon.in/Amazon-Basics-8-5-inch-W...>  
 653 <https://www.amazon.in/STRIFF-230X190X3mm-Water...>  
 940 <https://www.amazon.in/HP-330-Wireless-Keyboar...>  
 ...  
 1257 <https://www.amazon.in/Eureka-Forbes-Active-Cle...>  
 1317 <https://www.amazon.in/Proven%C2%AE-Copper-ADJU...>



```
1424 https://www.amazon.in/KENT-POWP-Sediment-Filte...
808   https://www.amazon.in/BRUSTRO-Copytinta-Colour...
950   https://www.amazon.in/Classmate-Pulse-Subject-...
```

```
[83 rows x 16 columns]
```

```
[137]: #3 What is the distribution of discounted prices vs. actual prices?

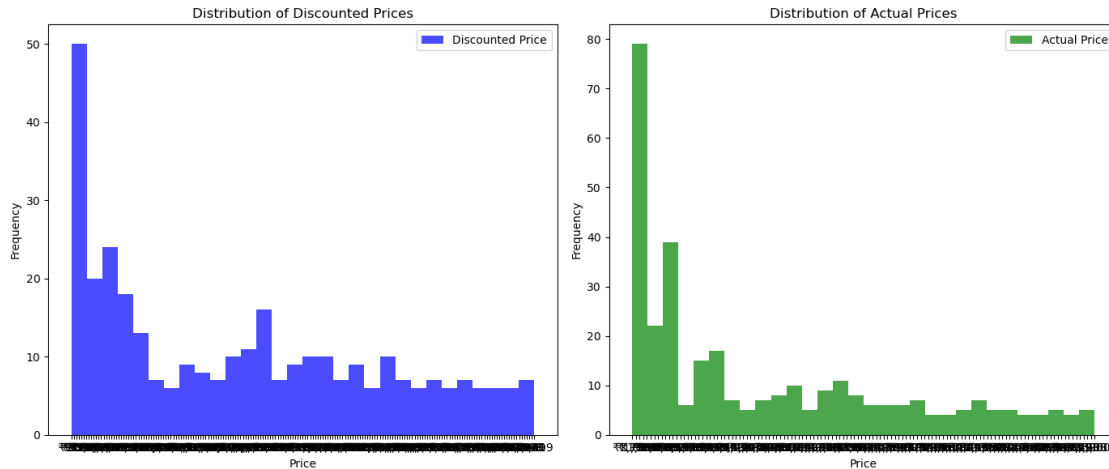
# Check for missing values in the price columns
df[['discounted_price', 'actual_price']].isnull().sum()

# Plot the distributions
plt.figure(figsize=(14, 6))

# Histogram for discounted prices
plt.subplot(1, 2, 1)
plt.hist(df['discounted_price'].dropna(), bins=30, alpha=0.7, color='blue',
        label='Discounted Price')
plt.xlabel('Price')
plt.ylabel('Frequency')
plt.title('Distribution of Discounted Prices')
plt.legend()

# Histogram for actual prices
plt.subplot(1, 2, 2)
plt.hist(df['actual_price'].dropna(), bins=30, alpha=0.7, color='green',
        label='Actual Price')
plt.xlabel('Price')
plt.ylabel('Frequency')
plt.title('Distribution of Actual Prices')
plt.legend()

plt.tight_layout()
plt.show()
```



[138]: #4. How does the average discount percentage vary across categories?

```
df['discount_percentage'] = df['discount_percentage'].str.rstrip('%').
    ↳astype(float)

# Group by 'category' and calculate average discount percentage
avg_discount_by_category = df.groupby('category')['discount_percentage'].mean().
    ↳reset_index()

# Sort by average discount percentage (descending)
avg_discount_by_category = avg_discount_by_category.
    ↳sort_values(by='discount_percentage', ascending=False)

print(avg_discount_by_category)
```

	category	discount_percentage
18	Electronics Headphones,Earbuds&Accessories Ear...	90.0
33	Electronics Mobiles&Accessories MobileAccessor...	90.0
3	Computers&Accessories Accessories&Peripherals ...	87.0
8	Computers&Accessories Accessories&Peripherals ...	80.0
12	Computers&Accessories Components InternalHardD...	80.0
..	...	...
11	Computers&Accessories Accessories&Peripherals ...	10.0
81	OfficeProducts OfficePaperProducts Paper Copy&...	0.0
57	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	0.0
20	Electronics HomeAudio MediaStreamingDevices St...	0.0
82	OfficeProducts OfficePaperProducts Paper Stati...	0.0

[83 rows x 2 columns]

[139]: #5. What are the most popular product names?

```
# Count the frequency of each product name
popular_products = df['product_name'].value_counts()

# Display the top 10 most popular product names (you can adjust the number as
needed)
top_product_names = popular_products.head(10).index.tolist()

print("Top 10 Most Popular Product Names:")
for idx, product_name in enumerate(top_product_names, start=1):
    print(f"{idx}. {product_name}")
```

Top 10 Most Popular Product Names:

1. Duracell USB Lightning Apple Certified (Mfi) Braided Sync & Charge Cable For Iphone, Ipad And Ipod. Fast Charging Lightning Cable, 3.9 Feet (1.2M) - Black
2. Duracell USB C To Lightning Apple Certified (Mfi) Braided Sync & Charge Cable For Iphone, Ipad And Ipod. Fast Charging Lightning Cable, 3.9 Feet (1.2M) - Black
3. Ambrane 2 in 1 Type-C & Micro USB Cable with 60W / 3A Fast Charging, 480 mbps High Data, PD Technology & Quick Charge 3.0, Compatible with All Type-C & Micro USB Devices (ABDC-10, Black)
4. Portronics Konnect L 20W PD Quick Charge Type-C to 8-Pin USB Mobile Charging Cable, 1.2M, Tangle Resistant, Fast Data Sync(Grey)
5. Lapster 65W compatible for OnePlus Dash Warp Charge Cable , type c to c cable fast charging Data Sync Cable Compatible with One Plus 10R / 9RT/ 9 pro/ 9R/ 8T/ 9/ Nord & for All Type C Devices - Red, 1 Meter
6. SWAPKART Fast Charging Cable and Data Sync USB Cable Compatible for iPhone 6/6S/7/7+/8/8+/10/11, 12, 13 Pro max iPad Air/Mini, iPod and iOS Devices (White)
7. WeCool S5 Long Selfie Stick, with Large Reinforced Tripod Stand up to 61 Inch / 156 Cms, Ultra Long Multi Function Bluetooth Selfie Stick with 1/4 Screw Compatible with Gopro, Camera, and Ring Light
8. CEDO 65W OnePlus Dash Warp Charge Cable, USB A to Type C Data Sync Fast Charging Cable Compatible with One Plus 3 /3T /5 /5T /6 /6T /7 /7T /7 pro & for All Type C Devices - 1 Meter, Red
9. oraimo 65W Type C to C Fast Charging Cable USB C to USB C Cable High Speed Syncing, Nylon Braided 1M length with LED Indicator Compatible For Laptop, Macbook, Samsung Galaxy S22 S20 S10 S20Fe S21 S21 Ultra A70 A51 A71 A50S M31 M51 M31S M53 5G
10. Firestick Remote

[140]: #6. What are the most popular product keywords?

```
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from collections import Counter
import string
```

```

# Assuming 'df' is your DataFrame containing the product data
# Example code snippet for keyword extraction

# Concatenate relevant fields into a single text field
df['text_combined'] = df['product_name'] + ' ' + df['review_title'] + ' ' +
    df['review_content'] + ' ' + df['about_product']

# Function to clean and tokenize text
def clean_and_tokenize(text):
    # Remove punctuation
    text = text.translate(str.maketrans('', '', string.punctuation))
    # Tokenize
    tokens = word_tokenize(text.lower())
    # Remove stopwords
    stop_words = set(stopwords.words('english'))
    tokens = [token for token in tokens if token not in stop_words]
    return tokens

# Apply cleaning and tokenization
df['tokens'] = df['text_combined'].apply(clean_and_tokenize)

# Count word frequencies
word_counts = Counter()
for tokens in df['tokens']:
    word_counts.update(tokens)

# Get most common words (adjust 'n' as needed)
common_words = word_counts.most_common(20)
print(common_words)

```

```

[('good', 942), ('product', 924), ('cable', 577), ('quality', 509), ('use',
439), ('remote', 352), ('charging', 343), ('easy', 322), ('water', 305), ('usb',
298), ('tv', 279), ('fast', 276), ('one', 276), ('also', 238), ('like', 236),
('price', 236), ('c', 225), ('1', 196), ('working', 196), ('well', 190)]

```

```

[141]: # Assuming df is your DataFrame with product reviews
# Calculate a weighted score combining rating and rating count
df['weighted_score'] = df['rating'] * df['rating_count']

# Sort reviews by weighted score in descending order to find the most popular
    ones
popular_reviews = df.sort_values(by='weighted_score', ascending=False)

# You can also consider other factors like engagement metrics, etc.
# For example, if you have 'likes' or 'comments' columns:
# popular_reviews = df.sort_values(by='likes', ascending=False)

```

```
# Select top N popular reviews
top_reviews = popular_reviews[['review_id', 'review_title', 'review_content', 'rating', 'rating_count']].head(10)

# Print or return the top reviews
top_reviews
```

```
[141]:
```

	review_id \
750	R1VCGAPSS4LWYQ,R3DS8EL4VV5LS6,R28MH1Y6O92EOP,R...
32	R2S0AYWUV349HP,R35OW9CYQNAYHY,R3B3DDF1D5NULK,R...
823	R2S0AYWUV349HP,R35OW9CYQNAYHY,R3B3DDF1D5NULK,R...
151	R2S0AYWUV349HP,R35OW9CYQNAYHY,R3B3DDF1D5NULK,R...
147	R3IUYQZ1BP7QPB,R3RCM1DK0EBGWB,R34I2C57PM50A3,R...
227	R26Z004978YU47,R13WAXAKPL2LIZ,RSOGJ8FAFL4E5,R3...
758	R1150W07XAD9VL,R3GGVCOWYVIRNV,R36CFZQPBAIJV8,R...
1262	R1785D08M4HFFD,R348X4GT06PQU9,R1VCNIW9SC311F,R...
1058	R1EU51LVE60B7C,R18PRSQIFU4R7M,R19E4QY5JWKCDD,R...
1272	R23G8LLBD9D4H3,R2SU25E3ZH4JEH,RCOH95A6KJB8J,RR...

	review_title \
750	Nice .,very good batteries received,Longteivity...
32	Great Cable, Charging Speeds Could Be Better,G...
823	Great Cable, Charging Speeds Could Be Better,G...
151	Great Cable, Charging Speeds Could Be Better,G...
147	Durable,Good Product,Okay ,So far so good,An ...
227	Good quality but not superb,Good,Good,Not work...
758	Decent quality product for the price,Nice prod...
1262	It's a good product in this price.,Nice produc...
1058	Good product, removes lint efficiently.,Good p...
1272	Ok,A good water heater,Easy to use inverter,Go...

	review_content	rating	rating_count
750	Nive,very good batteries received,Like,Good pr...	4.3	989.0
32	Not charging as fast as I'd have expected. May...	4.3	974.0
823	Not charging as fast as I'd have expected. May...	4.3	974.0
151	Not charging as fast as I'd have expected. May...	4.3	974.0
147	The cable does support fast PD charging, I hav...	4.2	919.0
227	Like ok,Product as specified,Good,,Yes it's ve...	4.2	910.0
758	I have been using it for the past week; no com...	4.2	903.0
1262	It's a good product and I'm using it since 1 w...	4.1	925.0
1058	I have been using this product from past few d...	4.2	900.0
1272	Ok,The product works very well. The support is...	3.9	959.0

```
[145]: #8. What is the correlation between discounted_price and rating?
# Convert 'discounted_price' to numeric type if necessary
df['discounted_price'] = pd.to_numeric(df['discounted_price'], errors='coerce')
```

```
# Convert 'rating' to numeric type if necessary
df['rating'] = pd.to_numeric(df['rating'], errors='coerce')

# Calculate the correlation between discounted_price and rating
correlation = df['discounted_price'].corr(df['rating'])

print(f"Correlation between discounted_price and rating: {correlation}")
```

Correlation between discounted\_price and rating: nan

```
[147]: #9. What are the Top 5 categories based on the highest ratings?

# Group by 'category' and calculate the average rating for each category
category_avg_rating = df.groupby('category')['rating'].mean()

# Sort the categories based on average rating in descending order
top_categories = category_avg_rating.sort_values(ascending=False).head(5)

# Print or return the top 5 categories with their average ratings
print("Top 5 categories based on highest ratings:")
top_categories
```

Top 5 categories based on highest ratings:

```
[147]: category
Computers&Accessories|Accessories&Peripherals|Keyboards,Mice&InputDevices|Mice
5.00
Electronics|Mobiles&Accessories|MobileAccessories|Photo&VideoAccessories|SelfieS
ticks      4.60
Electronics|HomeAudio|MediaStreamingDevices|StreamingClients
4.50
Home&Kitchen|Kitchen&HomeAppliances|WaterPurifiers&Accessories|WaterFilters&Puri
fiers      4.45
Electronics|Mobiles&Accessories|MobileAccessories|Décor|PhoneCharms
4.40
Name: rating, dtype: float64
```

```
[148]: #10. Identify any potential areas for improvement or optimization based on the
↳ data analysis.

#Normalization of Data: Ensure consistent units and formats across numerical
↳ columns (e.g., price normalization) to facilitate accurate analysis and
↳ comparisons.

#Index Optimization: Review the efficiency of the current indexing strategy,
↳ especially if the dataset is large, to enhance query performance.
```

*#Data Integrity Checks: Implement checks to ensure data consistency and integrity, such as validating rating counts against actual reviews or ensuring product IDs are unique.*

*#Text Data Handling: Utilize text preprocessing techniques (like tokenization, stemming, or lemmatization) on columns such as 'about\_product' and 'review\_content' for better text analytics.*

*#Image Link Utilization: Consider leveraging the 'img\_link' for image processing or visualization tasks to enhance product understanding or user engagement.*

*#Review and Rating Analysis: Conduct sentiment analysis on 'review\_content' and explore correlations between 'rating', 'rating\_count', and other attributes for deeper insights.*

*#Data Storage Optimization: Evaluate the data storage format (e.g., CSV, JSON, database) based on access patterns and analytical needs for optimal performance.*

*#Feature Engineering: Derive additional features from existing data (e.g., calculate effective price after discount) to enrich analysis capabilities.*

```
[152]: #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.
```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[154]: d=pd.read_csv("C:\\Users\\Malleesh\\Downloads\\spotify.csv")
```

```
[155]: d
```

```
[155]:
```

	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  3F5Cg0j3wFlRv51JsHbxhe
..          ...          ...          ...
435           44        221863  3fBsE0nzwtlkpSOLxXAZhN
436           64        191900  7Ezwtgfw7khBrpvaNPtMoT
437           69        283693  7xVLFuuYdAvcTfcP3IG3dS
438           65        220408  4hGmQboiou09EwhcTWa0H6
439           68        267720  0nq6sfr8z1R5KJ4XUk396e

```

```
[440 rows x 5 columns]
```

```
[156]: d.head()
```

```

[156]:   Artist                Track Name  Popularity  Duration (ms)  \
0  Drake  Rich Baby Daddy (feat. Sexyy Red & SZA)         92        319191
1  Drake                        One Dance         91        173986
2  Drake                IDGAF (feat. Yeat)         90        260111
3  Drake  First Person Shooter (feat. J. Cole)         88        247444
4  Drake        Jimmy Cooks (feat. 21 Savage)         88        218364

      Track ID
0  1yeB8MUNeLo9Ek1UEpsyz6
1  1zi7xx7UVEFkmKfv06H8x0
2  2YSzYUF3jWqb9YP9VXmpjE
3  7aqfrAY2p9BUSiupwk3svU
4  3F5Cg0j3wFlRv51JsHbxhe

```

```
[157]: d.shape
```

```
[157]: (440, 5)
```

```
[158]: d.columns
```

```
[158]: Index(['Artist', 'Track Name', 'Popularity', 'Duration (ms)', 'Track ID'],
      dtype='object')
```

```
[159]: d.dtypes
```

```

[159]: Artist      object
      Track Name  object

```



```
Popularity      int64
Duration (ms)   int64
Track ID        object
dtype: object
```

```
[160]: d.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
```

```
[161]: d.isna()
```

```
[161]:
```

	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]
```

```
[162]: d.isna().sum()
```

```
[162]: Artist          0
Track Name          0
Popularity          0
Duration (ms)       0
Track ID            0
dtype: int64
```

```
[163]: d.dropna()
```

```
[163]:
```

	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	3F5CgOj3wF1Rv51JsHbxhe
..	...	...	...
435	44	221863	3fBsEOonzwtlkpSOLxXAZhN
436	64	191900	7Ezwtgfw7khBrpvaNPtMoT
437	69	283693	7xVLFuuYdAvcTfcP3IG3dS
438	65	220408	4hGmQboiou09EwhcTWa0H6
439	68	267720	0nq6sfr8z1R5KJ4XUk396e

[440 rows x 5 columns]

```
[166]: # Find duplicate rows based on all columns
duplicate_rows = d[d.duplicated()]
```

```
[167]: duplicate_rows
```

```
[167]:
```

	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	3F5Cg0j3wF1Rv51JsHbxhe
76	85	239359	1bDbXMyjaUIooNwFE9wnON
131	88	247444	7aqfrAY2p9BUSiupwk3svU
170	89	194786	5wG3HvLhF6Y5KTG1K0IW3J
182	74	222154	337kcYVjYXdLBItCw9ry3b
210	87	157560	ORiRZpuVRbi7oqRdSMwhQY
212	79	231746	39MK3d3fonIP8Mz9oHCTBB
222	80	187541	40un2ylbjFKMPTiaSbbCih
224	78	175099	4YQImHflXSiIMXntcwPxx8
242	87	198937	2tTmW7RDtMQtBk7m2rYeSw
270	74	223878	6T7FXSuXykeGktMLGp8WgE
280	96	264324	3w0w2T288decOmgeZZqoNN
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	4dASQi01Eoo3RJvt74FtXB
352	83	109750	741UUVE2kuITl0c6zuqqb0
422	79	231746	39MK3d3fonIP8Mz9oHCTBB
430	87	233901	3B54sVLJ402zGa6Xm4YGNe
435	44	221863	3fBsEOnzwtlkipSOLxXAZhN
439	68	267720	0nq6sfr8z1R5KJ4XUk396e

```
[168]: # Remove duplicate rows based on all columns
df_no_duplicates = d.drop_duplicates()
```

```
[169]: df_no_duplicates
```

```
[169]:
```

	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)

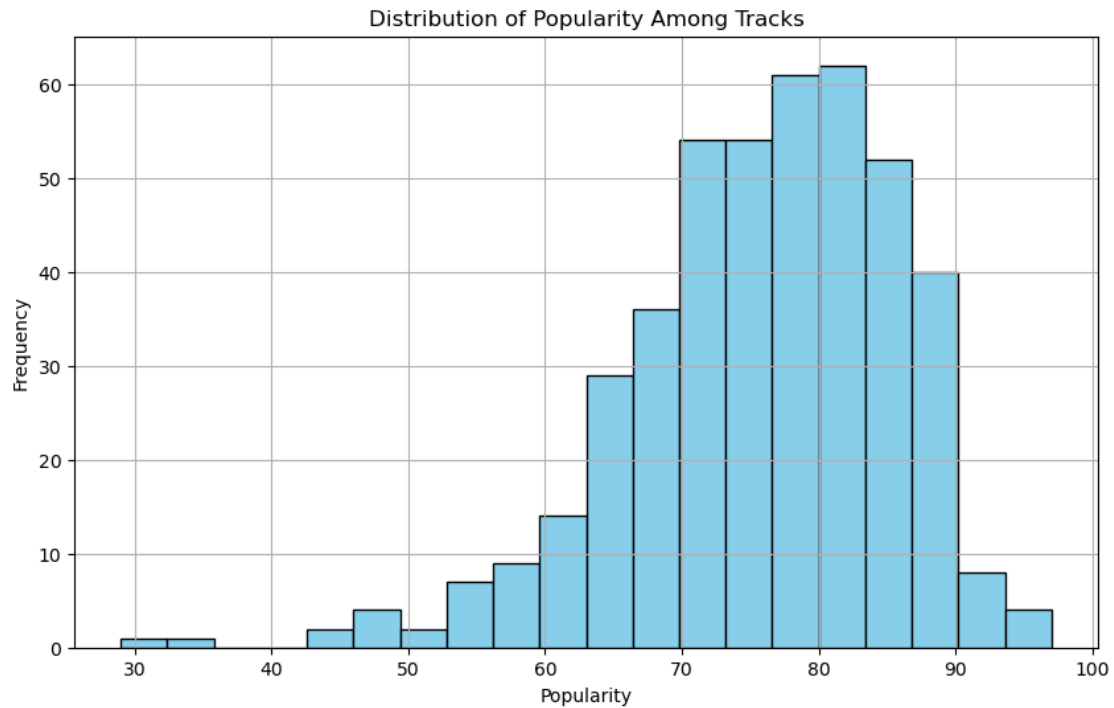
  

	Popularity	Duration (ms)	Track ID
0	92	319191	1yeB8MUNeLo9Ek1UEpsyz6
1	91	173986	1zi7xx7UVEFkmKfv06H8x0
2	90	260111	2YSzYUF3jWqb9YP9VXmpjE
3	88	247444	7aqfrAY2p9BUSiupwk3svU
4	88	218364	3F5Cg0j3wF1Rv51JsHbxhe
..	...	...	...
433	54	163971	01CHrTerCzyRpMI1MzQ4fz
434	65	187521	0TY3jVGwGDwDabLyQLVRQQ
436	64	191900	7Ezwtgfw7khBrpvaNPtMoT
437	69	283693	7xVLFuuYdAvcTfcP3IG3dS
438	65	220408	4hGmQboiou09EwhcTWa0H6

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

```
[171]: #2.What is the distribution of popularity among the tracks in the dataset?
↳Visualize it using a histogram.
```

```
# Plotting the histogram
plt.figure(figsize=(10, 6))
plt.hist(d['Popularity'], bins=20, color='skyblue', edgecolor='black')
plt.xlabel('Popularity')
plt.ylabel('Frequency')
plt.title('Distribution of Popularity Among Tracks')
plt.grid(True)
plt.show()
```

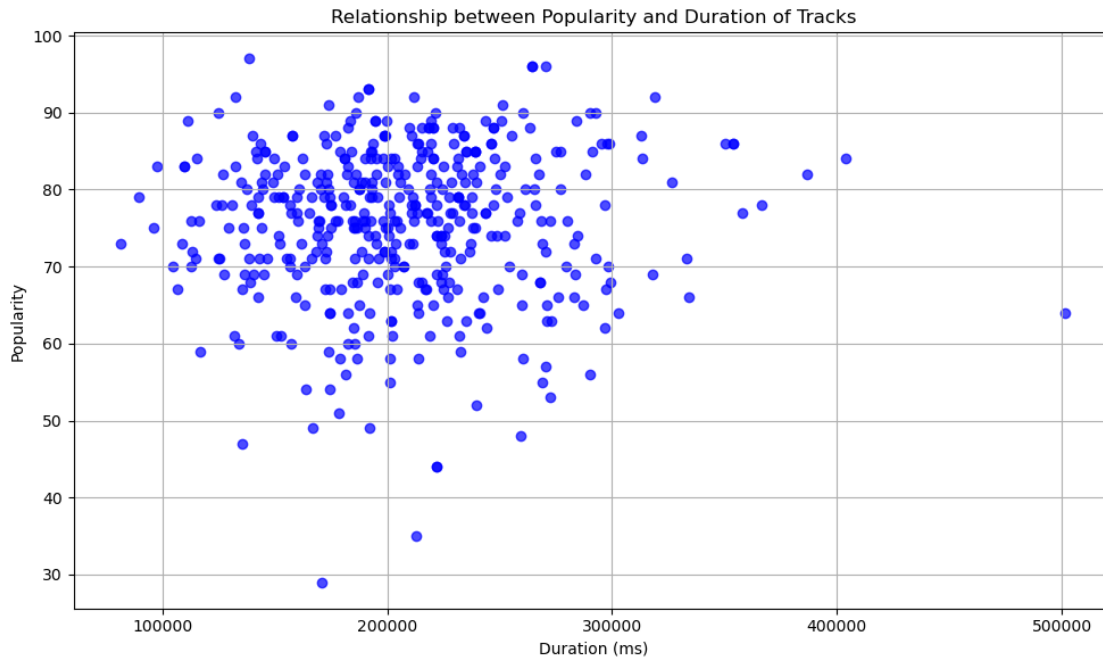


[175]: #3. Is there any relationship between the popularity and the duration of tracks? ┐  
→ Explore this using a scatter plot.

```
# Scatter plot
plt.figure(figsize=(10, 6))
plt.scatter(d['Duration (ms)'], d['Popularity'], marker='o', color='b', alpha=0.
    →7)

# Add titles and labels
plt.title('Relationship between Popularity and Duration of Tracks')
plt.xlabel('Duration (ms)')
plt.ylabel('Popularity')

# Show plot
plt.grid(True)
plt.tight_layout()
plt.show()
```



[178]: #4. Which artist has the highest number of tracks in the dataset? Display the count of tracks for each artist using a countplot.

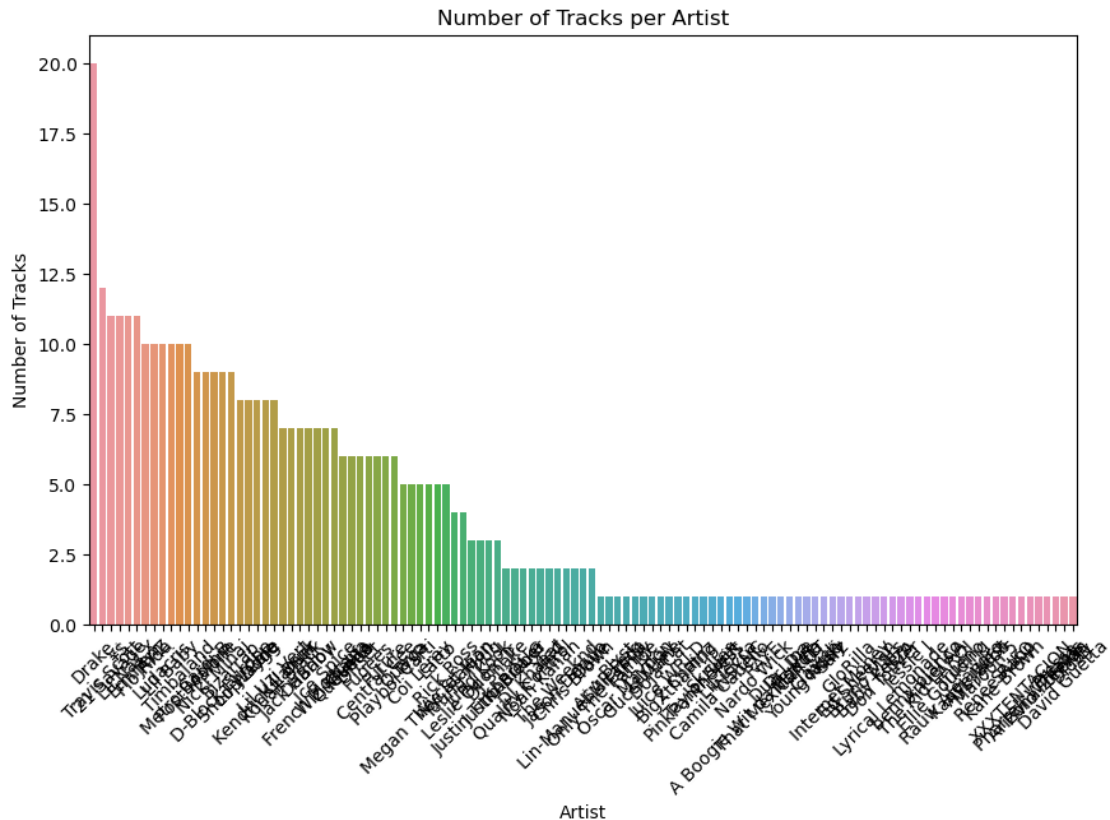
```
# Counting tracks per artist
track_counts = d['Artist'].value_counts()

# Finding the artist with the highest number of tracks
artist_with_most_tracks = track_counts.idxmax()
max_tracks = track_counts.max()

print(f"The artist with the highest number of tracks is {artist_with_most_tracks} with {max_tracks} tracks.")

# Plotting the countplot
plt.figure(figsize=(10, 6))
sns.countplot(data=d, x='Artist', order=d['Artist'].value_counts().index)
plt.xticks(rotation=45)
plt.title('Number of Tracks per Artist')
plt.xlabel('Artist')
plt.ylabel('Number of Tracks')
plt.show()
```

The artist with the highest number of tracks is 'Drake' with 20 tracks.



```
[179]: #5.What are the top 5 least popular tracks in the dataset? Provide the artist,
      ↪ name and track name for each.
```

```
# Sort the dataset by 'Popularity' in ascending order
df_sorted = d.sort_values(by='Popularity')

# Select the top 5 least popular tracks
top_5_least_popular = df_sorted.head(5)

# Print the Artist and Track Name for each of the top 5 least popular tracks
for index, row in top_5_least_popular.iterrows():
    print(f"Artist: {row['Artist']}, Track Name: {row['Track Name']}")
```

Artist: Pressa, Track Name: Attachments (feat. Coi Leray)  
Artist: Justin Bieber, Track Name: Intentions  
Artist: French Montana, Track Name: Splash Brothers  
Artist: French Montana, Track Name: Splash Brothers  
Artist: Lil Baby, Track Name: On Me - Remix

[180] :

```

#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.
# Group by Artist and calculate average popularity
average_popularity = d.groupby('Artist')['Popularity'].mean().reset_index()

# Sort by average popularity in descending order
average_popularity_sorted = average_popularity.sort_values(by='Popularity',
↳ascending=False)

# Get the top 5 most popular artists
top_5_artists = average_popularity_sorted.head(5)

print(top_5_artists)

```

	Artist	Popularity
113	cassö	92.000000
104	Trueno	89.000000
24	David Guetta	87.000000
103	Travis Scott	87.000000
114	¥\$	86.090909

[182]: #7.For the top 5 most popular artists, what are their most popular tracks? List  
↳the track name for each artist.

```

# Sort by 'Popularity' in descending order and select top 5
top_artists = d.sort_values(by='Popularity', ascending=False).head(5)

# Iterate through top 5 artists and print their most popular track
for index, row in top_artists.iterrows():
    artist_name = row['Artist']
    most_popular_track = d[d['Artist'] == artist_name].
↳sort_values(by='Popularity', ascending=False).iloc[0]['Track Name']
    print(f"Artist: {artist_name} | Most Popular Track: {most_popular_track}")

```

```

Artist: Jack Harlow | Most Popular Track: Lovin On Me
Artist: ¥$ | Most Popular Track: CARNIVAL
Artist: 21 Savage | Most Popular Track: redrum
Artist: ¥$ | Most Popular Track: CARNIVAL
Artist: Travis Scott | Most Popular Track: FE!N (feat. Playboi Carti)

```

[183]: #8.Visualize relationships between multiple numerical variables simultaneously  
↳using a pair plot.

```

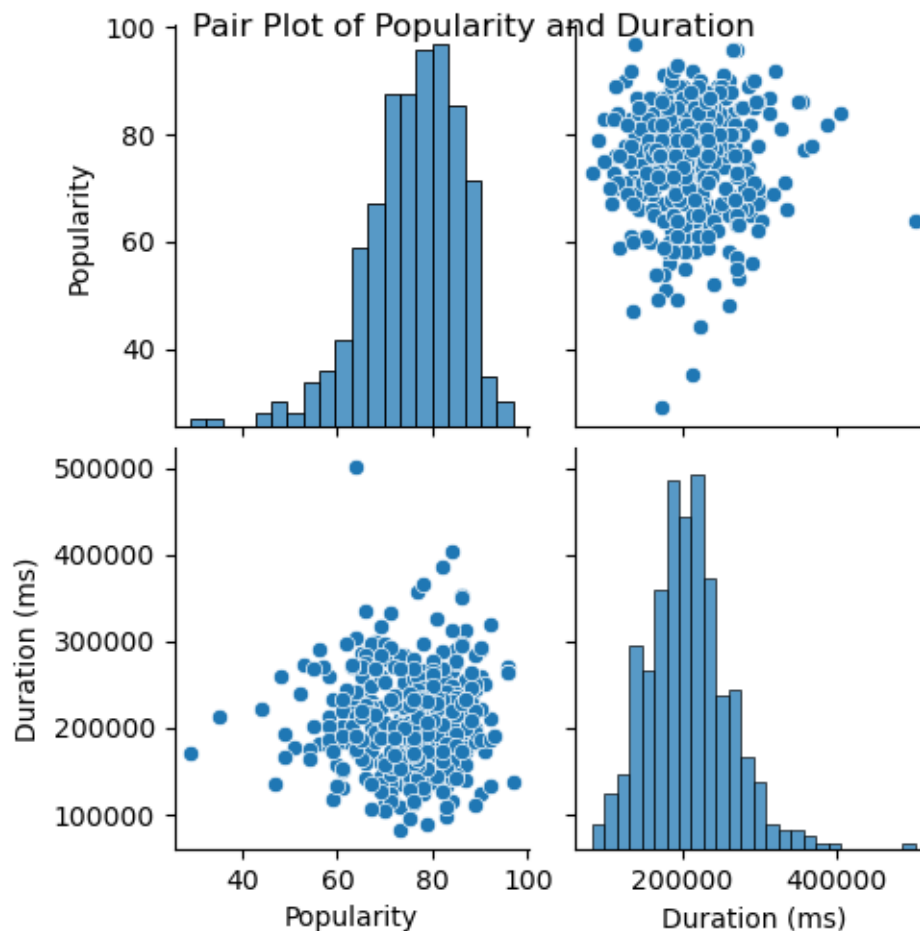
# Specify the numerical columns to include in the pair plot
numerical_vars = ['Popularity', 'Duration (ms)']

```



```
# Create pair plot
sns.pairplot(d[numerical_vars])
plt.suptitle('Pair Plot of Popularity and Duration')
plt.show()
```

C:\Users\Mallesha\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118:  
UserWarning: The figure layout has changed to tight  
self.\_figure.tight\_layout(\*args, \*\*kwargs)



[185]: #9.Does the duration of tracks vary significantly across different artists?\_  
→Explore this visually using a box plot or violin plot.

```
# Calculate average duration per artist
avg_duration_per_artist = d.groupby('Artist')['Duration (ms)'].mean().
    →reset_index()

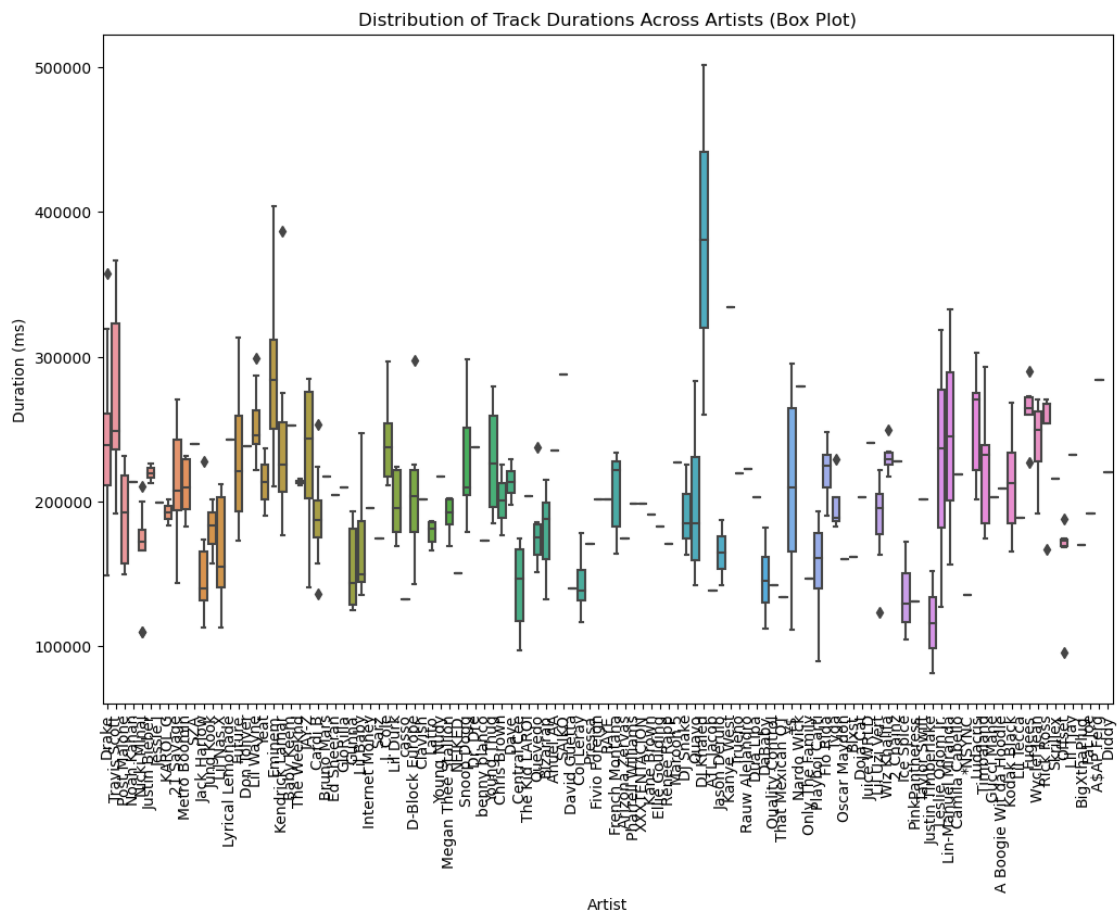
plt.figure(figsize=(12, 8))
```

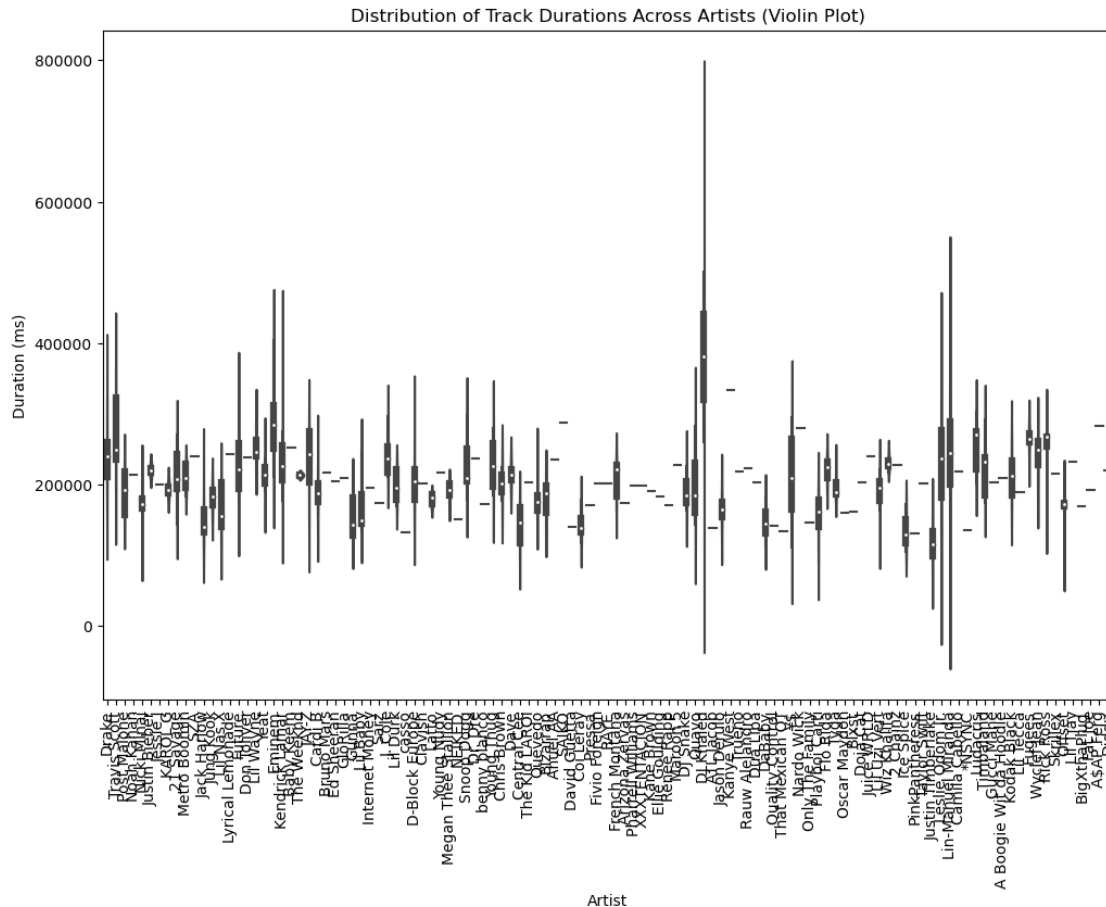
```

sns.boxplot(x='Artist', y='Duration (ms)', data=d)
plt.xticks(rotation=90)
plt.title('Distribution of Track Durations Across Artists (Box Plot)')
plt.show()

plt.figure(figsize=(12, 8))
sns.violinplot(x='Artist', y='Duration (ms)', data=d)
plt.xticks(rotation=90)
plt.title('Distribution of Track Durations Across Artists (Violin Plot)')
plt.show()

```





[186]: #10. How does the distribution of track popularity vary for different artists? ↪ Visualize this using a swarm plot or a violin plot.

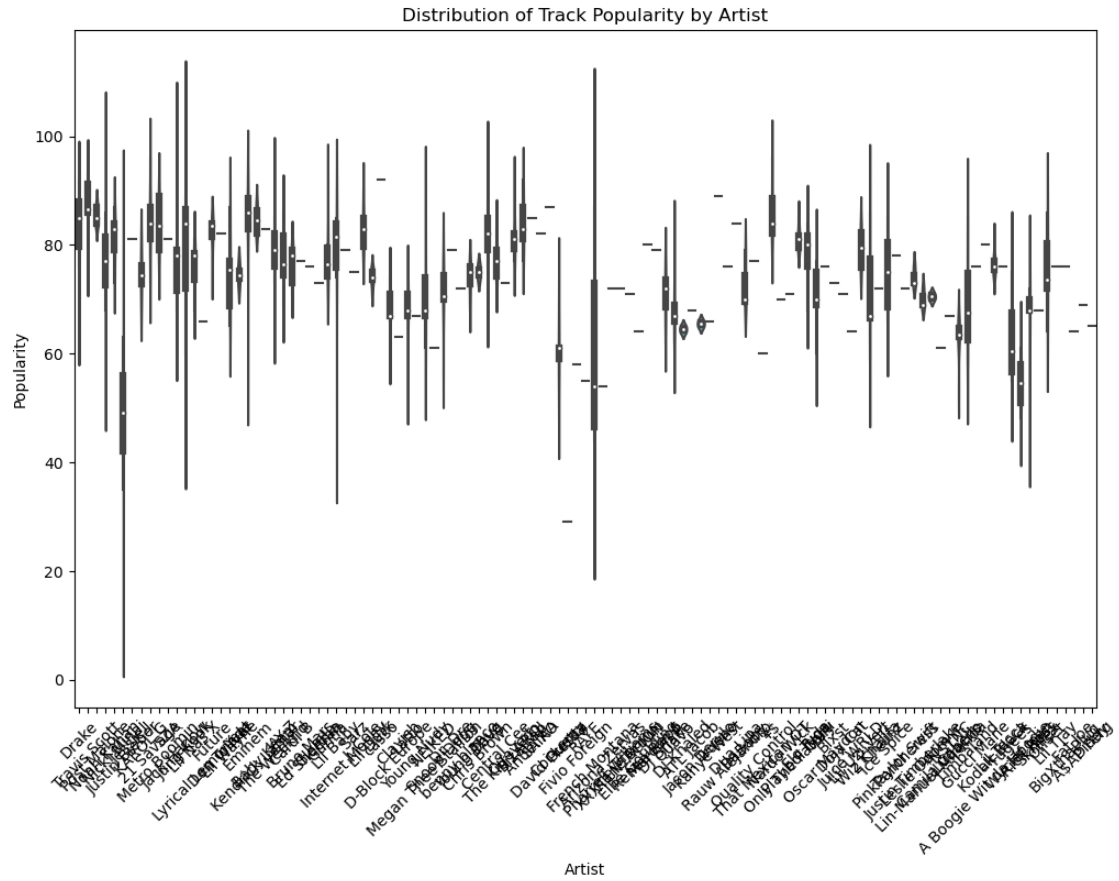
```
plt.figure(figsize=(12, 8))
sns.swarmplot(x='Artist', y='Popularity', data=d)
plt.title('Distribution of Track Popularity by Artist')
plt.xticks(rotation=45)
plt.show()

plt.figure(figsize=(12, 8))
sns.violinplot(x='Artist', y='Popularity', data=d)
plt.title('Distribution of Track Popularity by Artist')
plt.xticks(rotation=45)
plt.show()
```

C:\Users\Mallesha\anaconda3\Lib\site-packages\seaborn\categorical.py:3544:  
 UserWarning: 30.0% of the points cannot be placed; you may want to decrease the  
 size of the markers or use stripplot.  
 warnings.warn(msg, UserWarning)

```
C:\Users\Mallesha\anaconda3\Lib\site-packages\seaborn\categorical.py:3544:
UserWarning: 33.3% of the points cannot be placed; you may want to decrease the
size of the markers or use stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\Mallesha\anaconda3\Lib\site-packages\seaborn\categorical.py:3544:
UserWarning: 44.4% of the points cannot be placed; you may want to decrease the
size of the markers or use stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\Mallesha\anaconda3\Lib\site-packages\seaborn\categorical.py:3544:
UserWarning: 27.3% of the points cannot be placed; you may want to decrease the
size of the markers or use stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\Mallesha\anaconda3\Lib\site-packages\seaborn\categorical.py:3544:
UserWarning: 50.0% of the points cannot be placed; you may want to decrease the
size of the markers or use stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\Mallesha\anaconda3\Lib\site-packages\seaborn\categorical.py:3544:
UserWarning: 14.3% of the points cannot be placed; you may want to decrease the
size of the markers or use stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\Mallesha\anaconda3\Lib\site-packages\seaborn\categorical.py:3544:
UserWarning: 25.0% of the points cannot be placed; you may want to decrease the
size of the markers or use stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\Mallesha\anaconda3\Lib\site-packages\seaborn\categorical.py:3544:
UserWarning: 18.2% of the points cannot be placed; you may want to decrease the
size of the markers or use stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\Mallesha\anaconda3\Lib\site-packages\seaborn\categorical.py:3544:
UserWarning: 10.0% of the points cannot be placed; you may want to decrease the
size of the markers or use stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\Mallesha\anaconda3\Lib\site-packages\seaborn\categorical.py:3544:
UserWarning: 16.7% of the points cannot be placed; you may want to decrease the
size of the markers or use stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\Mallesha\anaconda3\Lib\site-packages\seaborn\categorical.py:3544:
UserWarning: 20.0% of the points cannot be placed; you may want to decrease the
size of the markers or use stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\Mallesha\anaconda3\Lib\site-packages\seaborn\categorical.py:3544:
UserWarning: 28.6% of the points cannot be placed; you may want to decrease the
size of the markers or use stripplot.
    warnings.warn(msg, UserWarning)
C:\Users\Mallesha\anaconda3\Lib\site-packages\seaborn\categorical.py:3544:
UserWarning: 22.2% of the points cannot be placed; you may want to decrease the
size of the markers or use stripplot.
    warnings.warn(msg, UserWarning)
```





[ ]:

[ ]: