

EDA_assign_@2

January 24, 2025

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

```
[2]: import os
os.path.exists(r"C:\Users\jarup\Downloads\data science papers\EDA\assignment_1\project\amazon.csv")
```

[2]: True

```
[5]: df=pd.read_csv(r"C:\Users\jarup\Downloads\data science papers\EDA\assignment_1\project\amazon.csv")
```

```
[7]: df
```

```
[7]:
```

	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	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 RGIQEG07R9HS2,R1SMWZQ86XIN8U,R2J3Y1WL29GWDE,RY...
 2 R3J3EQQ9TZI5ZJ,R3E7WBGK7ID0KV,RWU79XKQ6I1QF,R2...
 3 R3EEUZKKK9J36I,R3HJVYCLYOY554,REDECAZ7AMPQC,R1...
 4 R1BP4L2HH9TFUP,R16PVJEXKV6QZS,R2UPDB81N66T4P,R...
 ...
 1460 R3G3XFHPBFFOE8,R3COBZCD32EIGW,R2EBVBCN9QPD9R,R...
 1461 R3DDL2UPKQ2CK9,R2SYYU10ATVIU5,R1VM993161IYRW,R...
 1462 R1TLRJVV4STY5I,R2O455KRN493R1,R3Q5MVGEBRIAS2G,R...
 1463 R39Q2Y79MM9SWK,R3O79BG1NIH6MB,R29A31ZELTZNJM,R...
 1464 R2ORBRZOWEUIT9,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-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]

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

```

[9]: 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

```
[11]: df.shape
```

```
[11]: (1465, 16)
```

```
[13]: df.size
```

```
[13]: 23440
```

```
[15]: df.columns
```

```
[15]: 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')
```

```
[17]: df.describe()
```

```
[17]:
```

	product_id	product_name	\
count	1465	1465	
unique	1351	1337	
top	B07JW9H4J1	Fire-Boltt Ninja Call Pro Plus 1.83" Smart Wat...	
freq	3	5	

	category	discounted_price	\
count	1465	1465	
unique	211	550	
top	Computers&Accessories Accessories&Peripherals ...	199	
freq	233	53	

	actual_price	discount_percentage	rating	rating_count	\
count	1465	1465	1465	1463	
unique	449	92	28	1143	
top	999	50%	4.1	9,378	
freq	120	56	244	9	

	about_product	\
count	1465	
unique	1293	
top	[CHARGE & SYNC FUNCTION]- This cable comes wit...	
freq	6	

	user_id	\
count	1465	
unique	1194	
top	AHIKJUDTVJ4T6DV6IUGFYZ5LXMPA,AE55KTFVNXYPD5FPY...	

```

freq                                     10

                                     user_name \
count                                     1465
unique                                   1194
top    $@|\|T0$|-|,Sethu madhav,Akash Thakur,Burger P...
freq                                     10

                                     review_id \
count                                     1465
unique                                   1194
top    R3F4T5TRYPTMIG,R3DQIEC603E7AY,R104Z15FD40PV5,R...
freq                                     10

                                     review_title \
count                                     1465
unique                                   1194
top    Worked on iPhone 7 and didn't work on XR,Good ...
freq                                     10

                                     review_content \
count                                     1465
unique                                   1212
top    I am not big on camera usage, personally. I wa...
freq                                     8

                                     img_link \
count                                     1465
unique                                   1412
top    https://m.media-amazon.com/images/I/413sCRKobN...
freq                                     3

                                     product_link
count                                     1465
unique                                   1465
top    https://www.amazon.in/Wayona-Braided-WN3LG1-Sy...
freq                                     1

```

```
[19]: df.dtypes
```

```

[19]: 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

```

```
[21]: 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

```

```
[23]: df.fillna('0',inplace=True)
```

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

```

[25]: 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

```

```
[27]: df.drop_duplicates()
```

```

[27]:      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 AHITFY6AHALOFHOHOZEOC6XBP4FEA,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   RGIQEG07R9HS2,R1SMWZQ86XIN8U,R2J3Y1WL29GWDE,RY...

```

2 R3J3EQQ9TZI5ZJ,R3E7WBGK7IDOKV,RWU79XKQ6I1QF,R2...
 3 R3EEUZKKK9J36I,R3HJVYCLY0Y554,REDECAZ7AMPQC,R1...
 4 R1BP4L2HH9TFUP,R16PVJEXKV6QZS,R2UPDB81N66T4P,R...
 ...
 1460 R3G3XFHPBFF0E8,R3COBZCD32EIGW,R2EBVBCN9QPD9R,R...
 1461 R3DDL2UPKQ2CK9,R2SYYU10ATVIU5,R1VM993161IYRW,R...
 1462 R1TLRJVV4STY5I,R2O455KRN493R1,R3Q5MVGBRIAS2G,R...
 1463 R39Q2Y79MM9SWK,R3O79BG1NIH6MB,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-Konnect-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]
```

```
[29]: df.shape
```

```
[29]: (1465, 16)
```

```
[31]: df.columns
```

```
[31]: 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')
```

```
[33]: import warnings
      warnings.filterwarnings('ignore')
```

```
[35]: #What is the average rating for each product category?
```

```
[37]: df['rating'] = pd.to_numeric(df['rating'], errors='coerce') # Convert to_
      ↪numeric, coercing errors to NaN
      avg_rating_cat = df.groupby("category")['rating'].mean() # Calculate the mean
      print(avg_rating_cat)
```

```
category
Car&Motorbike|CarAccessories|InteriorAccessories|AirPurifiers&Ionizers
3.800000
Computers&Accessories|Accessories&Peripherals|Adapters|USBtoUSBAdapters
4.150000
Computers&Accessories|Accessories&Peripherals|Audio&VideoAccessories|PCHeadsets
3.500000
Computers&Accessories|Accessories&Peripherals|Audio&VideoAccessories|PCMicrophon
es
3.600000
```

Computers&Accessories|Accessories&Peripherals|Audio&VideoAccessories|PCSpeakers
4.050000

...

OfficeProducts|OfficePaperProducts|Paper|Stationery|Pens,Pencils&WritingSupplies
|Pens&Refills|GelInkRollerballPens 4.250000

OfficeProducts|OfficePaperProducts|Paper|Stationery|Pens,Pencils&WritingSupplies
|Pens&Refills|LiquidInkRollerballPens 4.150000

OfficeProducts|OfficePaperProducts|Paper|Stationery|Pens,Pencils&WritingSupplies
|Pens&Refills|RetractableBallpointPens 4.300000

OfficeProducts|OfficePaperProducts|Paper|Stationery|Pens,Pencils&WritingSupplies
|Pens&Refills|StickBallpointPens 4.133333

Toys&Games|Arts&Crafts|Drawing&PaintingSupplies|ColouringPens&Markers
4.300000

Name: rating, Length: 211, dtype: float64

```
[39]: print(df['rating_count'].dtype)
```

object

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

```
df = df.dropna(subset=['rating_count'])
```

```
df['rating_count'] = df['rating_count'].astype(int)
```

```
topRatedProd = df.groupby("category")['rating_count'].nlargest(5)
```

```
topRatedProd = topRatedProd.reset_index(level=0, drop=True)
```

```
print(topRatedProd)
```

900 817

32 974

151 974

823 974

147 919

...

1424 490

1449 159

1308 37

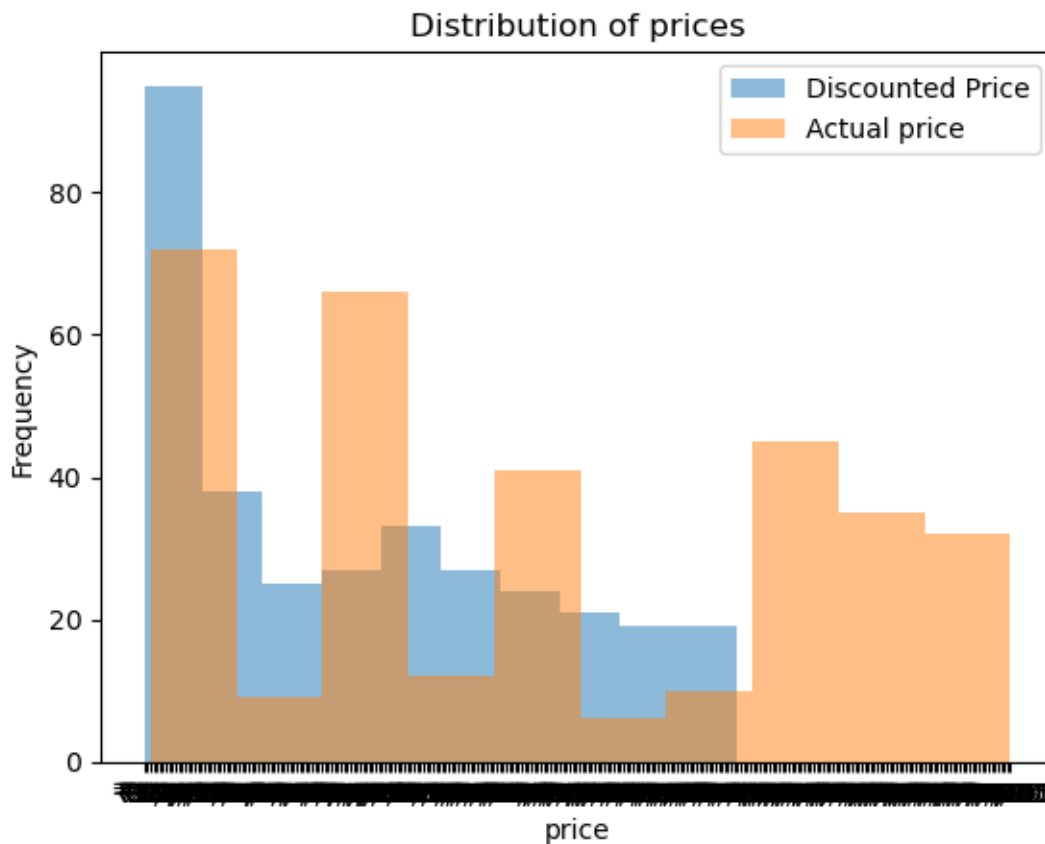
808 388

950 419

Name: rating_count, Length: 188, dtype: int32

```
[43]: #
plt.hist(df['discounted_price'],bins =10,alpha =0.5,label = 'Discounted Price')
plt.hist(df['actual_price'],bins=10,alpha=0.5,label = 'Actual price')
plt.xlabel('price')
plt.ylabel('Frequency')
plt.title("Distribution of prices")
```

```
plt.legend()
plt.show()
```



```
[44]: print(df['discount_percentage'].dtype)
```

object

```
[45]: df['discount_percentage'] = df['discount_percentage'].astype(str)
```

```
[49]: df['discount_percentage'] = df['discount_percentage'].str.rstrip('%').
      ↪astype(float)
```

```
[51]: avg_discount_per = df.groupby('category')['discount_percentage'].mean()
      print(avg_discount_per)
```

category

Computers&Accessories|Accessories&Peripherals|Cables&Accessories|Cables|SATACables
65.000000

Computers&Accessories|Accessories&Peripherals|Cables&Accessories|Cables|USBCables
59.081967

Computers&Accessories|Accessories&Peripherals|Keyboards,Mice&InputDevices|Graphi

```

cTablets                                68.333333
Computers&Accessories|Accessories&Peripherals|Keyboards,Mice&InputDevices|Keyboa
rd&MiceAccessories|MousePads           87.000000
Computers&Accessories|Accessories&Peripherals|Keyboards,Mice&InputDevices|Keyboa
rd&MouseSets                             36.000000
...
Home&Kitchen|Kitchen&HomeAppliances|Vacuum,Cleaning&Ironing|Vacuums&FloorCare|Va
cuums|HandheldVacuums                   38.000000
Home&Kitchen|Kitchen&HomeAppliances|WaterPurifiers&Accessories|WaterFilters&Puri
fiers                                   70.250000
Home&Kitchen|Kitchen&HomeAppliances|WaterPurifiers&Accessories|WaterPurifierAcce
ssories                                45.000000
OfficeProducts|OfficePaperProducts|Paper|Copy&PrintingPaper|ColouredPaper
0.000000
OfficeProducts|OfficePaperProducts|Paper|Stationery|Notebooks,WritingPads&Diarie
s|CompositionNotebooks                  0.000000
Name: discount_percentage, Length: 83, dtype: float64

```

```

[53]: #6
top_5_products = df['product_name'].value_counts().nlargest(5)
print(top_5_products)

```

```

product_name
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
3
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)
3
Portronics Konnect L 20W PD Quick Charge Type-C to 8-Pin USB Mobile Charging
Cable, 1.2M, Tangle Resistant, Fast Data Sync(Grey)
2
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
2
Name: count, dtype: int64

```

```

[150]: #7
top_5_product_keyword = df['product_id'].value_counts().nlargest(5)
print(top_5_product_keyword)

```

```

product_id
B09C6HXFC1    3
B09W5XR9RT    3
B09CMP1SC8    3
B09Q8HMKZX    2

```

```
BOBMXLSMM      2
Name: count, dtype: int64
```

```
[56]: top_5_review = df['review_title'].value_counts().nlargest(5)
      print(top_5_review)
```

```
review_title
Good cable for car,Good substitute for orginal,Better Value for money
Product,Way better than the original,Absolutely amazing.,Namm hi kafi hai,Very
good,As the names say Durable cell it a durable cable ;-)
3
Great Cable, Charging Speeds Could Be Better,Good,A good cable.,One of the best
type c cable,Works as intended.,A good buy. The extra length helps a
lot.,Good,Ok
3
Get it with bundled discounts.,Heating & Touch screen,Buy for normal usage. NOT
FOR CAMERA,The phone is a good device and I am happy with the purchase.,Average
quality.,Exlent mobile,Valueable buy,A GOOD AND AFFORDABLE PRODUCT      3
Very good product.,Using as a spare cable in car,Sturdy, Durable, Fast
Charging!,Good brand,It's like original apple cable,One of the best wire
..,Super well build. Quality product worth the money,Good product
3
Good product,Its good, but micro usb doesn't fit my phone.,Good and useful
item,It is very best cable,good,2 in 1 Charging Cable.,Sturdy cable overall,Nice
& Best Charger Cabel                                                    3
Name: count, dtype: int64
```

```
[64]: df['discounted_price'] = df['discounted_price'].replace({' ': '', ',': ''},
      ↪ regex=True).astype(float)
      correlation = df['discounted_price'].corr(df['rating'])
      print("Correlation between discounted price and rating:", correlation)
```

```
Correlation between discounted price and rating: 0.07996602303100599
```

```
[66]: top_5_category_rating = df.groupby('category')['rating'].mean().nlargest(5)
      print(top_5_category_rating)
```

```
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
```

```
[ ]: #10
# analyzing how dynamic pricing affects sales can help identify which products
  ↳are most responsive to discounts, allowing for better-targeted promotions.
  ↳By studying reviews over time, we can predict future sales patterns and
  ↳adjust strategies based on shifts in customer demand.
# including both short and long reviews, can uncover important issues that
  ↳affect their satisfaction
#using these insights can improve pricing strategies, optimize product
  ↳offerings, and drive higher sales, all while ensuring that decisions are
  ↳aligned with customer expectations and needs
```

```
[68]: import os
os.path.exists(r"C:\Users\jarup\Downloads\data science papers\EDA\assignment_
  ↳project\spotify.csv")
```

[68]: True

```
[70]: df1=pd.read_csv(r"C:\Users\jarup\Downloads\data science papers\EDA\assignment_
  ↳project\spotify.csv")
```

[72]: df1

```
[72]:
```

	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	3F5Cg0j3wF1Rv51JsHbxhe
..
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]

```
[74]: #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.
      df1.shape
```

[74]: (440, 5)

```
[112]: import warnings
      warnings.filterwarnings('ignore')
```

```
[76]: df1.size
```

[76]: 2200

```
[78]: df1.dtypes
```

```
[78]: Artist          object
      Track Name      object
      Popularity      int64
      Duration (ms)   int64
      Track ID        object
      dtype: object
```

```
[80]: df1.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
```

```
[82]: df1.isnull().sum()
```

```
[82]: Artist          0
      Track Name      0
      Popularity      0
      Duration (ms)   0
      Track ID        0
```

dtype: int64

```
[84]: df1.describe()
```

```
[84]:
```

	Popularity	Duration (ms)
count	440.000000	440.000000
mean	75.736364	206810.040909
std	9.886534	53576.930289
min	29.000000	81666.000000
25%	70.000000	172778.500000
50%	77.000000	201866.000000
75%	83.000000	235119.750000
max	97.000000	501648.000000

```
[86]: df1.drop_duplicates(inplace=True)
```

```
[88]: df1
```

```
[88]:
```

	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	3F5CgOj3wF1Rv51JsHbxhe
..
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]

```
[90]: df1.shape
```

```
[90]: (413, 5)
```

```
[92]: df1.size
```

```
[92]: 2065
```

```
[94]: df1['Popularity'].value_counts()
```

```
[94]: Popularity
```

76	23
79	21
84	19
82	18
78	18
71	18
75	16
77	16
81	16
85	15
73	14
80	14
74	13
67	13
86	13
69	12
83	12
72	11
68	10
70	10
87	10
64	9
88	9
89	8
66	7
61	7
65	7
90	6
63	6
58	5
60	4
92	4
62	3
59	3
96	2
55	2
56	2
54	2

```

49      2
91      2
29      1
51      1
97      1
47      1
35      1
53      1
52      1
48      1
57      1
44      1
93      1
Name: count, dtype: int64

```

```
[96]: df1['Track Name'].value_counts()
```

```

[96]: Track Name
Dial Drunk (with Post Malone)      2
Rich Baby Daddy (feat. Sexyy Red & SZA)  1
New Tank                          1
Low (feat. T-Pain)                 1
Life Goes On                       1
..
Eagle                             1
Overseas                         1
Prada                            1
She Knows (feat. Amber Coffman & Cults)  1
Another One Of Me (feat. 21 Savage)  1
Name: count, Length: 412, dtype: int64

```

```
[98]: df1['Duration (ms)'].value_counts()
```

```

[98]: Duration (ms)
219333      2
252262      2
213817      2
162053      1
232946      1
..
174545      1
143200      1
297290      1
222154      1
220408      1
Name: count, Length: 410, dtype: int64

```

```
[100]: df1['Track ID'].value_counts()
```

```
[100]: Track ID
1yeB8MUNeLo9Ek1UEpsyz6      1
4txKMpsSfZRV6durPuHVq0      1
OCAfXk7DXMnon4gLudAp7J      1
2bSk87AVkCIIC3Bcligq1z      1
37Nqx7iavZpotJSDXZWbJ3      1
..
337kcYVjYXdLBIItCw9ry3b      1
59NraMJsLaMCVtwXTSia8i      1
OFs9cdPDhptWEDJmiCbKEW      1
2e3Ea0o24lReQFR4FA7yXH      1
4hGmQboiou09EwhcTWa0H6      1
Name: count, Length: 413, dtype: int64
```

```
[102]: df1['Artist'].value_counts()
```

```
[102]: Artist
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: count, Length: 115, dtype: int64
```

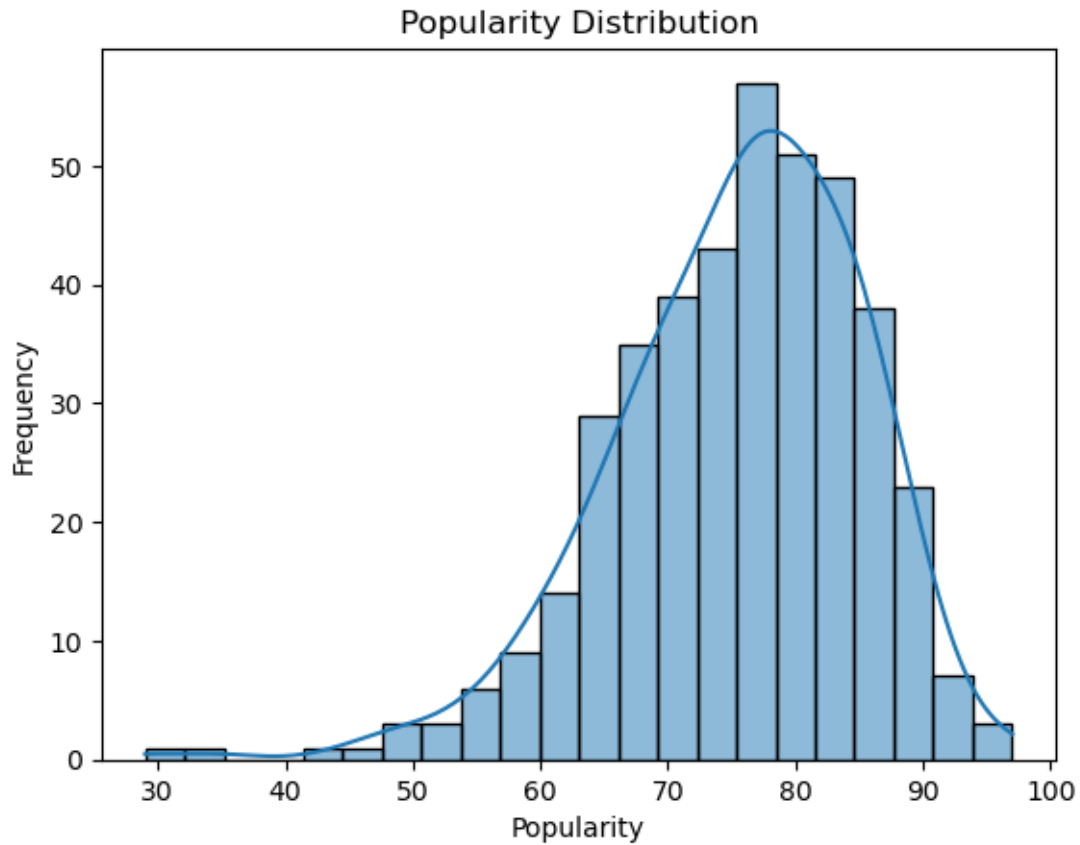
```
[104]: # there is no null values present in the data and removed the duplicate data
```

```
[106]: df1.columns
```

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

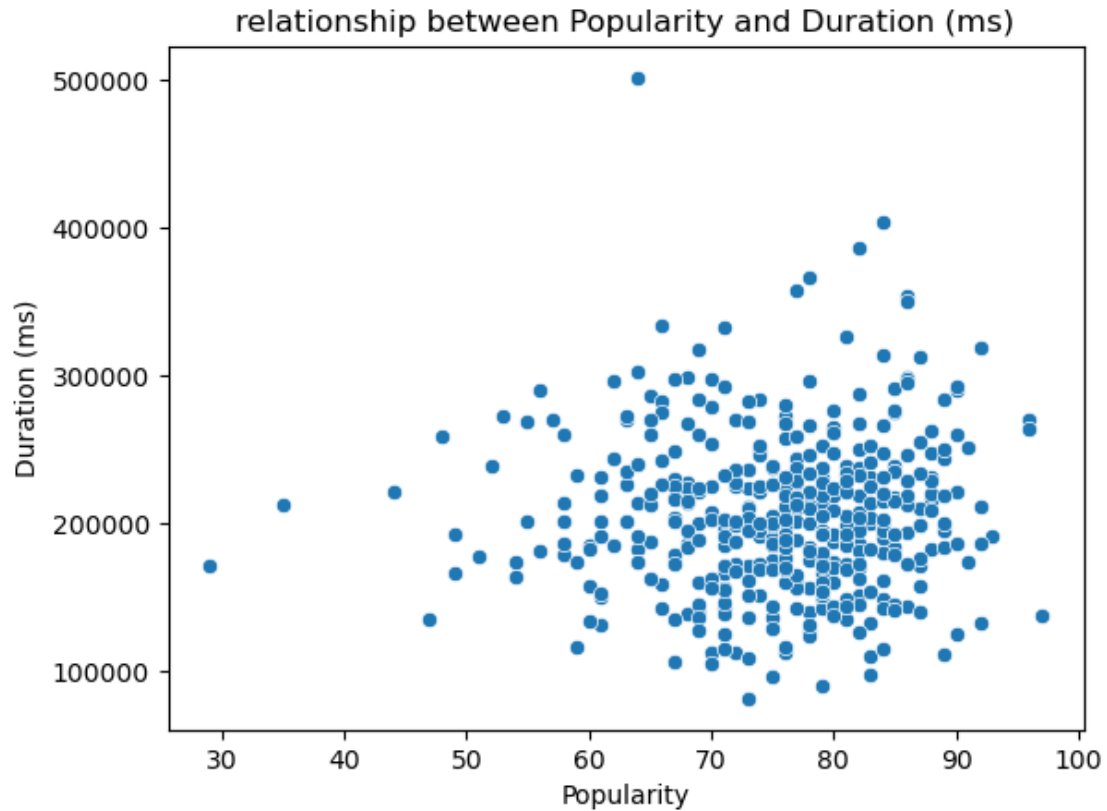
```
[91]: #2.What is the distribution of popularity among the tracks in the dataset?
      ↪Visualize it using a histogram.
```

```
[108]: sns.histplot(df1.Popularity,kde=True)
plt.xlabel('Popularity')
plt.ylabel('Frequency')
plt.title('Popularity Distribution')
plt.show()
```



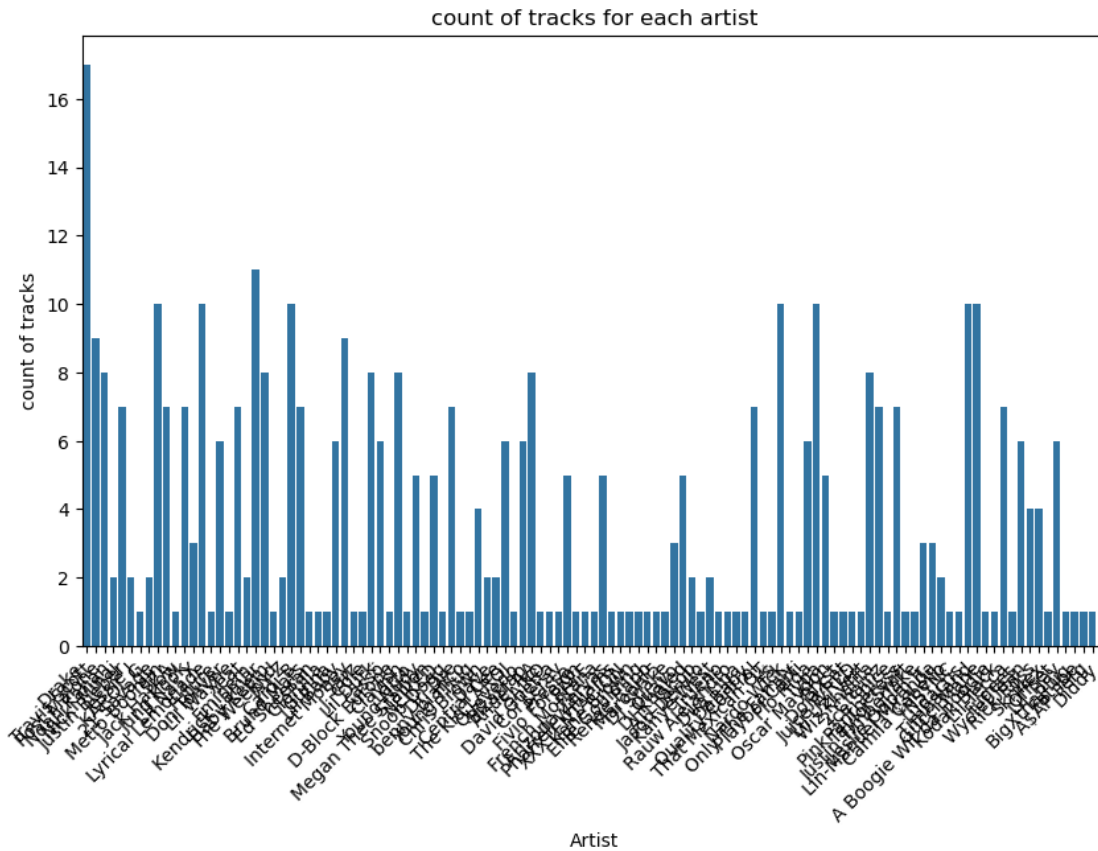
[95]: *#3. Is there any relationship between the popularity and the duration of tracks? ↴*
↪ Explore this using a scatter plot.

```
[97]: sns.scatterplot(x='Popularity',y='Duration (ms)',data=df1)
plt.xlabel('Popularity')
plt.ylabel('Duration (ms)')
plt.title("relationship between Popularity and Duration (ms)")
plt.show()
```

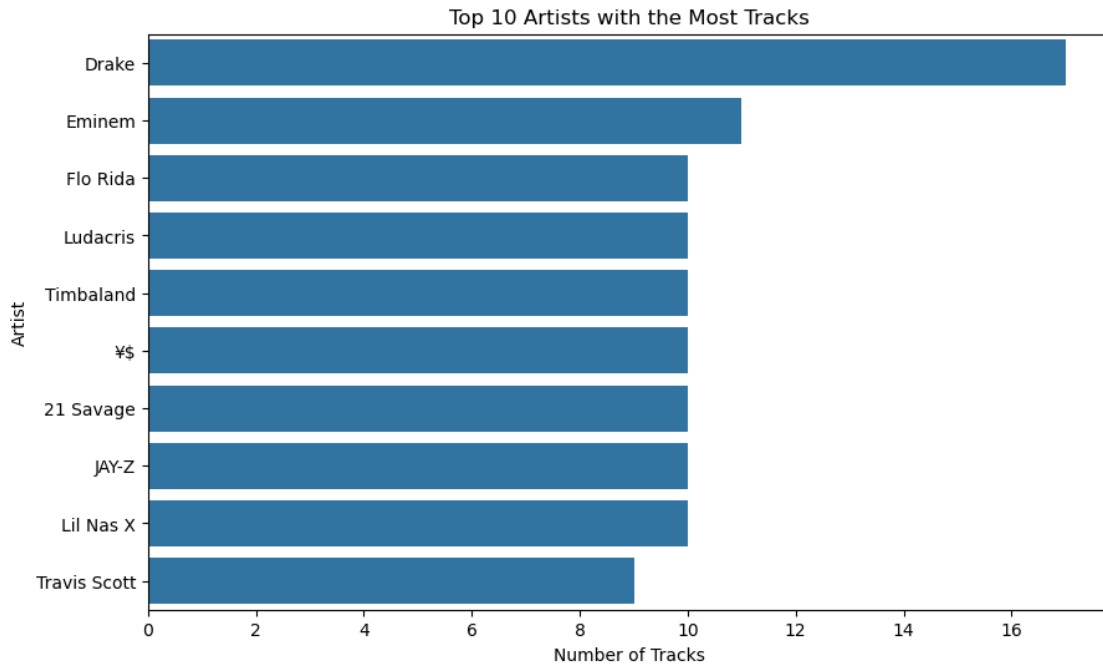


[]: #4. Which artist has the highest number of tracks in the dataset? Display the
 ↳ count of tracks for each artist using a countplot.

```
[116]: plt.figure(figsize=(10,6))
sns.countplot(x="Artist",data=df1)
plt.title("count of tracks for each artist")
plt.xlabel("Artist")
plt.ylabel("count of tracks")
plt.xticks(rotation=45, ha="right")
plt.show()
```

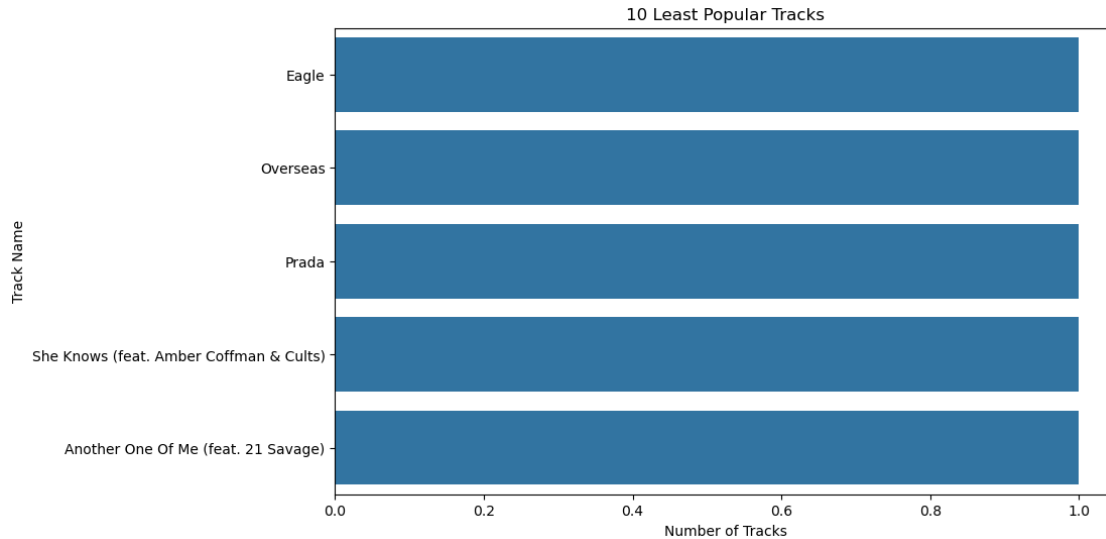


```
[144]: top_10_artists = artist_counts.head(10)
plt.figure(figsize=(10, 6))
sns.countplot(data=df1, y='Artist', order=top_10_artists.index)
plt.title('Top 10 Artists with the Most Tracks')
plt.xlabel('Number of Tracks')
plt.ylabel('Artist')
plt.show()
```

[]: *#What are the top 5 least popular tracks in the dataset? Provide the artist_ name and track name for each*

```
[100]: track_name_counts = df1['Track Name'].value_counts()
least_popular_tracks = track_name_counts.tail(5)
plt.figure(figsize=(10, 6))
sns.barplot(x=least_popular_tracks.values, y=least_popular_tracks.index)
plt.title('10 Least Popular Tracks')
plt.xlabel('Number of Tracks')
plt.ylabel('Track Name')
plt.show()
```



```
[120]: top_5_least_pop = df1.sort_values(by="Popularity",ascending=True).head(5)
print(top_5_least_pop[["Artist","Popularity"]])
```

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

```
[126]: #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.
top_5_most_pop=df1.groupby("Artist")["Popularity"].mean().nlargest(5)
print(top_5_most_pop)
```

Artist	
cassö	92.000000
Trueno	89.000000
David Guetta	87.000000
Travis Scott	86.555556
¥\$	85.100000

Name: Popularity, dtype: float64

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

```
[132]: top_5_products = df1['Track Name'].value_counts().nlargest(5)
print(top_5_products)
```

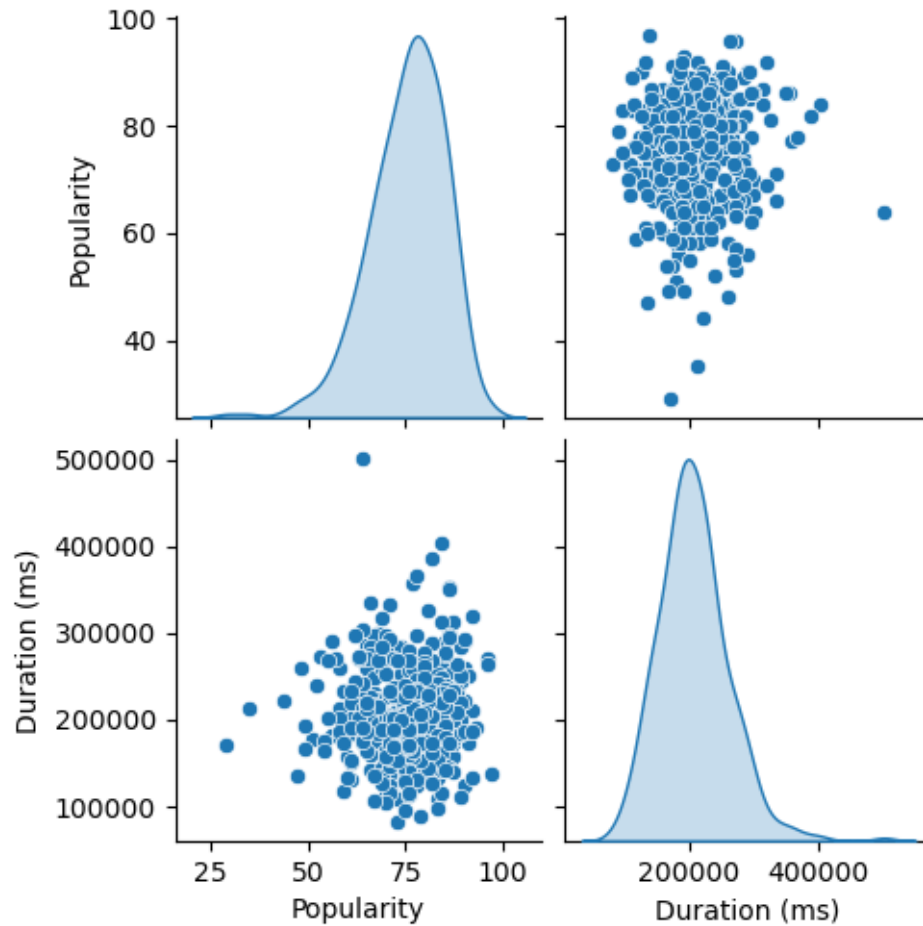
Track Name	
Dial Drunk (with Post Malone)	2
Rich Baby Daddy (feat. Sexyy Red & SZA)	1
New Tank	1
Low (feat. T-Pain)	1
Life Goes On	1
Name: count, dtype: int64	

```
[130]: print(df.columns)
```

```
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')
```

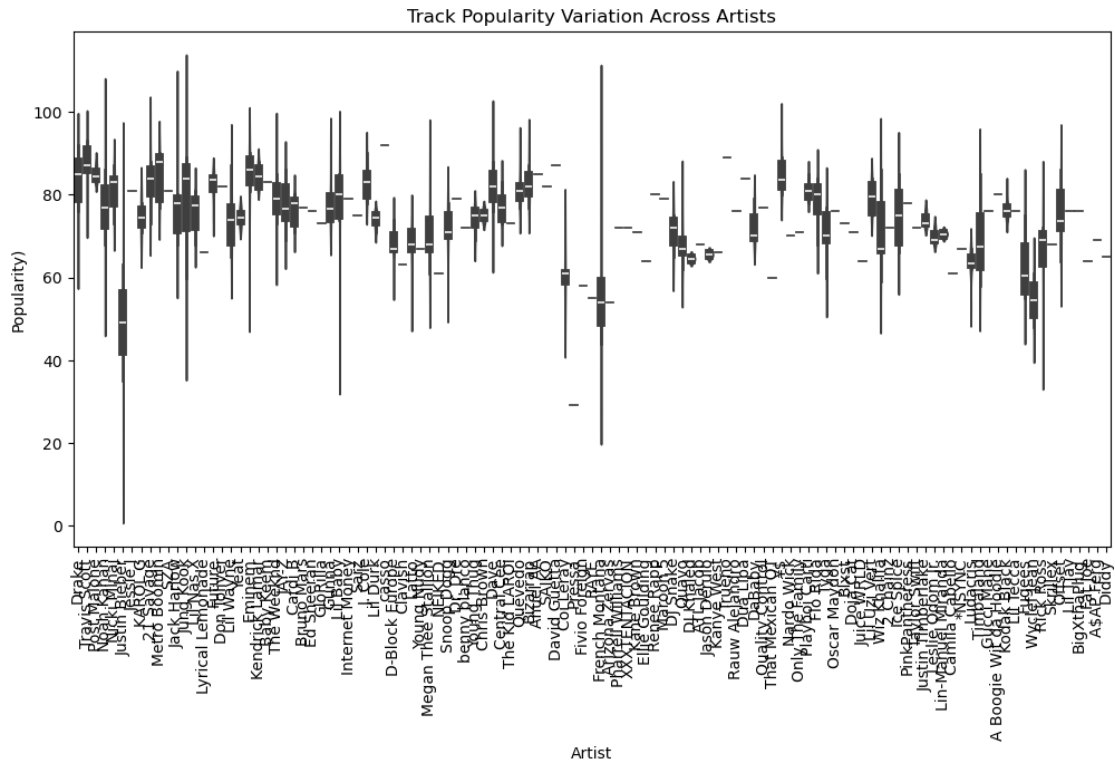
```
[103]: #.Visualize relationships between multiple numerical variables simultaneously,
       →using a pair plot
numeric_columns = df1.select_dtypes(include=['number'])
sns.pairplot(numeric_columns,diag_kind='kde')
```

```
[103]: <seaborn.axisgrid.PairGrid at 0x1bdb042f9e0>
```

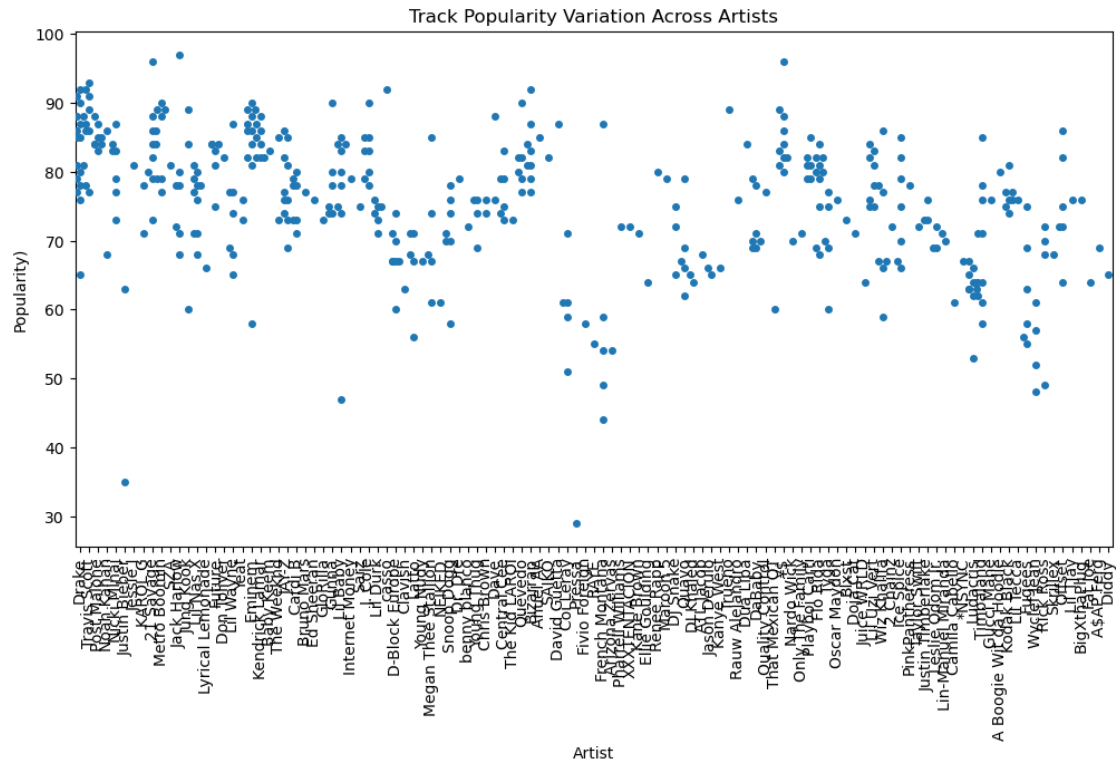


[105]: *#Does the duration of tracks vary significantly across different artists?
#Explore this visually using a box plot or violin plot.*

```
plt.figure(figsize=(12, 6))
sns.boxplot(data=df1, x="Artist", y="Duration (ms)")
plt.title('Track Duration Variation Across Artists')
plt.xlabel('Artist')
plt.ylabel('Duration (ms)')
plt.xticks(rotation=90)
plt.show()
```

```
[121]: plt.figure(figsize=(12, 6))
sns.swarmplot(data=df1, x="Artist", y="Popularity")
plt.title('Track Popularity Variation Across Artists')
plt.xlabel('Artist')
plt.ylabel('Popularity')
plt.xticks(rotation=90)
plt.show()
```



```
[117]: import warnings
warnings.filterwarnings('ignore')
```

```
[ ]:
```