

amazon-sales-dataset-analysis

July 16, 2025

```
[3]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import plotly.graph_objects as go
from plotly.subplots import make_subplots
import plotly.figure_factory as ff
```

```
[7]: df = pd.read_csv('amazon.csv', engine='python')
df.head()
```

```
[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...
```

```
                                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 AECPFYFQVRUWC3KGNLJIOREFP5LQ,AGYYVPDD7YG7FYNBX...
 2 AGU3BBQ2V2DDAMOAKGFAWDDQ6QHA,AESFLDV2PT363T2AQ...
 3 AEWAZDZZJLQUYVOVGBEUKSLXHQ5A,AG5HTSFRRE6NL3M5S...
 4 AE3Q6KSUK5P75D5HFYHCRAOLODSA,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,R3HJVYCLYOY554,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...

```

- 1 We had some eof string problem so we have to solve it in eda, thats why i used enigne = 'python'.
- 2 Copy original dataset in a new dataset for EDA & other processes

```
[8]: df1 = df.copy()
```

```

[9]: # Count columns in each row line
with open("amazon.csv", "r", encoding="utf-8") as f:
    lines = f.readlines()

expected_columns = 16
broken_rows = []

for i, line in enumerate(lines):
    if line.count(",") < expected_columns - 1:
        broken_rows.append((i, line.strip()))

print(f"Total broken rows: {len(broken_rows)}")
for row_num, content in broken_rows[:5]: # Show only first 5
    print(f"Row {row_num}: {content[:150]}...")

```

Total broken rows: 0

3 Now we are good to go

```
[10]: df1.shape
```

```
[10]: (932, 16)
```

```
[11]: df1.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 932 entries, 0 to 931
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -
0   product_id            932 non-null    object
1   product_name          932 non-null    object

```

```

2  category          932 non-null  object
3  discounted_price  932 non-null  object
4  actual_price     932 non-null  object
5  discount_percentage 932 non-null  object
6  rating           932 non-null  float64
7  rating_count     930 non-null  object
8  about_product    932 non-null  object
9  user_id          932 non-null  object
10 user_name        932 non-null  object
11 review_id        932 non-null  object
12 review_title     932 non-null  object
13 review_content   932 non-null  object
14 img_link         932 non-null  object
15 product_link     932 non-null  object
dtypes: float64(1), object(15)
memory usage: 116.6+ KB

```

```

[12]: # Clean discounted_price
df1['discounted_price'] = df1['discounted_price'].str.replace(' ', '').str.
      ↪replace(',', '').astype(float)

# Clean actual_price
df1['actual_price'] = df1['actual_price'].str.replace(' ', '').str.replace(',', '
      ↪').astype(float)

# Clean discount_percentage
df1['discount_percentage'] = df1['discount_percentage'].str.replace('%', '
      ↪').astype(float)

# Convert rating to float
df1['rating'] = pd.to_numeric(df1['rating'], errors='coerce')
df1['rating'].fillna(df1['rating'].mean(), inplace=True)

# Clean rating_count (remove commas and convert to int)
df1['rating_count'] = df1['rating_count'].str.replace(',', ' ')
df1['rating_count'] = pd.to_numeric(df1['rating_count'], errors='coerce').
      ↪fillna(0).astype(int)

# Extract main category (before first '/')
df1['main_category'] = df1['category'].str.split('/').str[0]

```

/tmp/ipython-input-12-3254353290.py:12: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an inplace method.

The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting values always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.

```
df1['rating'].fillna(df1['rating'].mean(), inplace=True)
```

```
[13]: df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 932 entries, 0 to 931
Data columns (total 17 columns):
#   Column                Non-Null Count  Dtype
---  -
0   product_id            932 non-null    object
1   product_name          932 non-null    object
2   category              932 non-null    object
3   discounted_price      932 non-null    float64
4   actual_price          932 non-null    float64
5   discount_percentage   932 non-null    float64
6   rating                932 non-null    float64
7   rating_count          932 non-null    int64
8   about_product         932 non-null    object
9   user_id              932 non-null    object
10  user_name             932 non-null    object
11  review_id            932 non-null    object
12  review_title         932 non-null    object
13  review_content       932 non-null    object
14  img_link             932 non-null    object
15  product_link         932 non-null    object
16  main_category        932 non-null    object
dtypes: float64(4), int64(1), object(12)
memory usage: 123.9+ KB
```

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

```
[14]: 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
main_category      0
dtype: int64

```

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

```

[15]:      discounted_price  actual_price  discount_percentage  rating \
count      932.000000      932.000000      932.000000  932.000000
mean      3599.548948      6206.442361        51.162017    4.117275
std       8182.995121     12474.283180        22.010800    0.269327
min        39.000000       50.000000         0.000000    2.800000
25%       279.000000       699.000000        35.000000    4.000000
50%       577.000000      1490.000000        55.000000    4.200000
75%      1799.000000      4992.250000        67.000000    4.300000
max      77990.000000     139900.000000       94.000000    5.000000

      rating_count
count      932.000000
mean     24677.965665
std     51031.325280
min         0.000000
25%     1779.750000
50%     8732.500000
75%     24269.000000
max    426973.000000

```

```
[16]: df1.duplicated().sum()
```

```
[16]: np.int64(0)
```

```
[17]: df1.head()
```

```

[17]:      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|...      399.0
1  Computers&Accessories|Accessories&Peripherals|...      199.0

```

| | | |
|---|---|-------|
| 2 | Computers&Accessories Accessories&Peripherals ... | 199.0 |
| 3 | Computers&Accessories Accessories&Peripherals ... | 329.0 |
| 4 | Computers&Accessories Accessories&Peripherals ... | 154.0 |

| | actual_price | discount_percentage | rating | rating_count | \ |
|---|--------------|---------------------|--------|--------------|---|
| 0 | 1099.0 | 64.0 | 4.2 | 24269 | |
| 1 | 349.0 | 43.0 | 4.0 | 43994 | |
| 2 | 1899.0 | 90.0 | 3.9 | 7928 | |
| 3 | 699.0 | 53.0 | 4.2 | 94363 | |
| 4 | 399.0 | 61.0 | 4.2 | 16905 | |

| | 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 | AE3Q6KSUK5P75D5HFYHCRAOLODSA,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,R3HJVYCLYOY554,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          main_category
0 https://www.amazon.in/Wayona-Braided-WN3LG1-Sy... Computers&Accessories
1 https://www.amazon.in/Ambrane-Unbreakable-Char... Computers&Accessories
2 https://www.amazon.in/Sounce-iPhone-Charging-C... Computers&Accessories
3 https://www.amazon.in/Deuce-300-Resistant-Tang... Computers&Accessories
4 https://www.amazon.in/Portronics-Konnect-POR-1... Computers&Accessories

```

4 We already added main_technology column and now we should do more feature engineering.

```
[18]: df1['main_category'].value_counts()
```

```

[18]: main_category
Electronics          507
Computers&Accessories 393
OfficeProducts        22
Home&Kitchen          6
MusicalInstruments    2
HomeImprovement        1
Toys&Games            1
Name: count, dtype: int64

```

```
[19]: df1['discount_amount'] = df1['actual_price'] - df1['discounted_price']
```

```
[20]: df1['discount_ratio'] = df1['discount_amount'] / df1['actual_price'] * 100
```

```
[21]: df1[df1['discounted_price'] <= 0]
```

```

[21]: Empty DataFrame
Columns: [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, main_category,
discount_amount, discount_ratio]

```


Index: []

```
[22]: df1['is_highly_rated'] = df1['rating'].apply(lambda x: 1 if x >= 4.0 else 0)
      df1['is_highly_rated'].value_counts()
```

```
[22]: is_highly_rated
      1    735
      0    197
      Name: count, dtype: int64
```

```
[23]: df1['is_popular'] = df1['rating_count'].apply(lambda x: 1 if x >= 1000 else 0)
      df1['is_popular'].value_counts()
```

```
[23]: is_popular
      1    766
      0    166
      Name: count, dtype: int64
```

5 Means Everything Is Discounted

6 Means Most Of The Things Are Highly Rated

7 Means They Are Highly Rated By Many Customers

```
[24]: def price_bucket(price):
      if price < 500:
          return 'Low'
      elif price < 1500:
          return 'Mid'
      elif price < 3000:
          return 'High'
      else:
          return 'Premium'

      df1['price_bucket'] = df1['discounted_price'].apply(price_bucket)
      df1.head()
```

```
[24]: 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  B08HJDJ86NZ  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|...      399.0
```

| | | |
|---|---|-------|
| 1 | Computers&Accessories Accessories&Peripherals ... | 199.0 |
| 2 | Computers&Accessories Accessories&Peripherals ... | 199.0 |
| 3 | Computers&Accessories Accessories&Peripherals ... | 329.0 |
| 4 | Computers&Accessories Accessories&Peripherals ... | 154.0 |

| | actual_price | discount_percentage | rating | rating_count | \ |
|---|--------------|---------------------|--------|--------------|---|
| 0 | 1099.0 | 64.0 | 4.2 | 24269 | |
| 1 | 349.0 | 43.0 | 4.0 | 43994 | |
| 2 | 1899.0 | 90.0 | 3.9 | 7928 | |
| 3 | 699.0 | 53.0 | 4.2 | 94363 | |
| 4 | 399.0 | 61.0 | 4.2 | 16905 | |

| | 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 | AECPFYFQVRUWC3KGNLJIOREFP5LQ,AGYYVPDD7YG7FYNBX... | ... | |
| 2 | AGU3BBQ2V2DDAMOAKGFAWDDQ6QHA,AESFLDV2PT363T2AQ... | ... | |
| 3 | AEWAZDZZJLQUYVOVGBEUKSLXHQ5A,AG5HTSFRRE6NL3M5S... | ... | |
| 4 | AE3Q6KSUK5P75D5HFYHCRAOLODSA,AFUGIFH5ZAFXRDSZH... | ... | |

| | 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 | main_category \ |
|---|--|-----------------------|
| 0 | https://www.amazon.in/Wayona-Braided-WN3LG1-Sy... | Computers&Accessories |
| 1 | https://www.amazon.in/Ambrane-Unbreakable-Char... | Computers&Accessories |
| 2 | https://www.amazon.in/Sounce-iPhone-Charging-C... | Computers&Accessories |
| 3 | https://www.amazon.in/Deuce-300-Resistant-Tang... | Computers&Accessories |
| 4 | https://www.amazon.in/Portronics-Konnnect-POR-1... | Computers&Accessories |

| | discount_amount | discount_ratio | is_highly_rated | is_popular | price_bucket |
|---|-----------------|----------------|-----------------|------------|--------------|
| 0 | 700.0 | 63.694268 | 1 | 1 | Low |
| 1 | 150.0 | 42.979943 | 1 | 1 | Low |
| 2 | 1700.0 | 89.520800 | 0 | 1 | Low |
| 3 | 370.0 | 52.932761 | 1 | 1 | Low |
| 4 | 245.0 | 61.403509 | 1 | 1 | Low |

[5 rows x 22 columns]

```
[25]: df1['price_bucket'].value_counts()
```

```
[25]: price_bucket
Low      444
Mid       227
Premium  162
High      99
Name: count, dtype: int64
```

```
[26]: df1['sub_category'] = df1['category'].str.split('|').str[-1]
df1['sub_category'].value_counts()
```

```
[26]: sub_category
USBCables      219
SmartWatches   75
Smartphones    68
SmartTelevisions 62
In-Ear         50
...
GamingKeyboards 1
SoundbarSpeakers 1
Earbuds          1
InkjetPrinters   1
ColouringPens&Markers 1
Name: count, Length: 122, dtype: int64
```

```
[27]: df1['category_depth'] = df1['category'].str.count('|') + 1
df1['category_depth'].value_counts()
```

```
[27]: category_depth
83      221
```

| | |
|-----|----|
| 45 | 91 |
| 70 | 68 |
| 61 | 65 |
| 62 | 65 |
| 63 | 64 |
| 64 | 30 |
| 73 | 26 |
| 80 | 23 |
| 71 | 22 |
| 77 | 20 |
| 94 | 18 |
| 74 | 13 |
| 90 | 12 |
| 58 | 10 |
| 75 | 9 |
| 57 | 9 |
| 102 | 9 |
| 85 | 8 |
| 49 | 8 |
| 95 | 7 |
| 50 | 7 |
| 110 | 7 |
| 79 | 7 |
| 92 | 7 |
| 78 | 6 |
| 81 | 6 |
| 66 | 6 |
| 88 | 5 |
| 55 | 5 |
| 53 | 5 |
| 69 | 5 |
| 67 | 5 |
| 82 | 4 |
| 72 | 4 |
| 87 | 4 |
| 104 | 4 |
| 76 | 4 |
| 60 | 3 |
| 54 | 3 |
| 84 | 3 |
| 59 | 3 |
| 46 | 2 |
| 56 | 2 |
| 114 | 2 |
| 106 | 2 |
| 111 | 2 |
| 116 | 2 |

```

42      2
41      2
52      2
32      2
100     2
48      2
120     2
44      1
47      1
108     1
51      1
65      1
Name: count, dtype: int64

```

```
[28]: df1['review_length'] = df1['review_content'].str.len()
```

```
[29]: from textblob import TextBlob

# Function to get sentiment polarity
def get_sentiment(text):
    if not isinstance(text, str) or text.strip() == '':
        return 'Neutral'
    polarity = TextBlob(text).sentiment.polarity
    if polarity > 0.1:
        return 'Positive'
    elif polarity < -0.1:
        return 'Negative'
    else:
        return 'Neutral'

df1['title_sentiment'] = df1['review_title'].apply(get_sentiment)
df1['review_content_sentiment'] = df1['review_content'].apply(get_sentiment)
df1.head()
```

```
[29]:
```

| | product_id | product_name \ | category | discounted_price \ |
|---|-------------|---|---|--------------------|
| 0 | B07JW9H4J1 | Wayona Nylon Braided USB to Lightning Fast Cha... | Computers&Accessories Accessories&Peripherals ... | 399.0 |
| 1 | B098NS6PVG | Ambrane Unbreakable 60W / 3A Fast Charging 1.5... | Computers&Accessories Accessories&Peripherals ... | 199.0 |
| 2 | B096MSW6CT | Sounce Fast Phone Charging Cable & Data Sync U... | Computers&Accessories Accessories&Peripherals ... | 199.0 |
| 3 | B08HJDJ86NZ | boAt Deuce USB 300 2 in 1 Type-C & Micro USB S... | Computers&Accessories Accessories&Peripherals ... | 329.0 |
| 4 | B08CF3B7N1 | Portronics Konnect L 1.2M Fast Charging 3A 8 P... | Computers&Accessories Accessories&Peripherals ... | 154.0 |

| | actual_price | discount_percentage | rating | rating_count | \ |
|---|--------------|---------------------|--------|--------------|---|
| 0 | 1099.0 | 64.0 | 4.2 | 24269 | |
| 1 | 349.0 | 43.0 | 4.0 | 43994 | |
| 2 | 1899.0 | 90.0 | 3.9 | 7928 | |
| 3 | 699.0 | 53.0 | 4.2 | 94363 | |
| 4 | 399.0 | 61.0 | 4.2 | 16905 | |

| | 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 | ... | discount_amount | \ |
|---|---|-----|-----------------|---|
| 0 | AG3D604STAQKAY2UVGEUV46KN35Q,AHMY5CWJMMK5BJRBB... | ... | 700.0 | |
| 1 | AECPFYFQVRUWC3KGNLJIOREFP5LQ,AGYYVPDD7YG7FYNBX... | ... | 150.0 | |
| 2 | AGU3BBQ2V2DDAMOAKGFAWDDQ6QHA,AESFLDV2PT363T2AQ... | ... | 1700.0 | |
| 3 | AEWAZDZZJLQUYVOVGBEUKSLXHQ5A,AG5HTSFRRE6NL3M5S... | ... | 370.0 | |
| 4 | AE3Q6KSUK5P75D5HFYHCRAOLODSA,AFUGIFH5ZAFXRDSZH... | ... | 245.0 | |

| | discount_ratio | is_highly_rated | is_popular | price_bucket | sub_category | \ |
|---|----------------|-----------------|------------|--------------|--------------|---|
| 0 | 63.694268 | 1 | 1 | Low | USBCables | |
| 1 | 42.979943 | 1 | 1 | Low | USBCables | |
| 2 | 89.520800 | 0 | 1 | Low | USBCables | |
| 3 | 52.932761 | 1 | 1 | Low | USBCables | |
| 4 | 61.403509 | 1 | 1 | Low | USBCables | |

| | category_depth | review_length | title_sentiment | review_content_sentiment |
|---|----------------|---------------|-----------------|--------------------------|
| 0 | 83 | 483 | Positive | Positive |
| 1 | 83 | 1186 | Positive | Positive |
| 2 | 83 | 271 | Positive | Positive |
| 3 | 83 | 443 | Positive | Positive |
| 4 | 83 | 2451 | Positive | Positive |

[5 rows x 27 columns]

```
[30]: print(df1['title_sentiment'].value_counts())
df1['review_content_sentiment'].value_counts()
```

```
title_sentiment
Positive    752
Neutral    149
Negative     31
Name: count, dtype: int64
```

```
[30]: review_content_sentiment
      Positive      881
      Neutral       44
      Negative        7
      Name: count, dtype: int64
```

#Highly Number Of Positives. For title_sentiment: Positive: 81.43%, Neutral: 14.75%, Negative: 3.40%

8 For review__content__sentiment:

Positive: 92.43%, Neutral: 6.83%, Negative: 0.75%

9 Now i am removing some unnecessary columns, bcz i think they are not valueable for the upcoming analysis and insights.

```
[31]: df1.drop(columns=['user_name', 'review_id', 'product_link'], inplace=True)
```

10 Now i have to download the clean for upcoming analysis and for power bi also.

```
[32]: df1.to_csv('amazon_sails_clean_data.csv', index=False)
```

11 Now we are ready to some analysis.

```
[33]: df1.head()
```

```
[33]:
```

| | product_id | product_name \ | category | discounted_price \ | actual_price | discount_percentage | rating | rating_count \ |
|---|-------------|---|---|--------------------|--------------|---------------------|--------|----------------|
| 0 | B07JW9H4J1 | Wayona Nylon Braided USB to Lightning Fast Cha... | Computers&Accessories Accessories&Peripherals ... | 399.0 | 1099.0 | 64.0 | 4.2 | 24269 |
| 1 | B098NS6PVG | Ambrane Unbreakable 60W / 3A Fast Charging 1.5... | Computers&Accessories Accessories&Peripherals ... | 199.0 | | | | |
| 2 | B096MSW6CT | Sounce Fast Phone Charging Cable & Data Sync U... | Computers&Accessories Accessories&Peripherals ... | 199.0 | | | | |
| 3 | B08HJDJ86NZ | boAt Deuce USB 300 2 in 1 Type-C & Micro USB S... | Computers&Accessories Accessories&Peripherals ... | 329.0 | | | | |
| 4 | B08CF3B7N1 | Portronics Konnect L 1.2M Fast Charging 3A 8 P... | Computers&Accessories Accessories&Peripherals ... | 154.0 | | | | |

| | | | | |
|---|--------|------|-----|-------|
| 1 | 349.0 | 43.0 | 4.0 | 43994 |
| 2 | 1899.0 | 90.0 | 3.9 | 7928 |
| 3 | 699.0 | 53.0 | 4.2 | 94363 |
| 4 | 399.0 | 61.0 | 4.2 | 16905 |

```

                                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 | discount_amount |
|---|---|-----------------|
| 0 | AG3D604STAQKAY2UVGEUV46KN35Q,AHMY5CWJMMK5BJRBB... | 700.0 |
| 1 | AECPFYFQVRUWC3KGNLJIOREFP5LQ,AGYYVPDD7YG7FYNBX... | 150.0 |
| 2 | AGU3BBQ2V2DDAMOAKGFAWDDQ6QHA,AESFLDV2PT363T2AQ... | 1700.0 |
| 3 | AEWAZDZZJLQUYVOVGBEUKSLXHQ5A,AG5HTSFRRE6NL3M5S... | 370.0 |
| 4 | AE3Q6KSUK5P75D5HFYHCRAOLODSA,AFUGIFH5ZAFXRDSZH... | 245.0 |

| | discount_ratio | is_highly_rated | is_popular | price_bucket | sub_category |
|---|----------------|-----------------|------------|--------------|--------------|
| 0 | 63.694268 | 1 | 1 | Low | USBCables |
| 1 | 42.979943 | 1 | 1 | Low | USBCables |
| 2 | 89.520800 | 0 | 1 | Low | USBCables |
| 3 | 52.932761 | 1 | 1 | Low | USBCables |
| 4 | 61.403509 | 1 | 1 | Low | USBCables |

| | category_depth | review_length | title_sentiment | review_content_sentiment |
|---|----------------|---------------|-----------------|--------------------------|
| 0 | 83 | 483 | Positive | Positive |
| 1 | 83 | 1186 | Positive | Positive |
| 2 | 83 | 271 | Positive | Positive |
| 3 | 83 | 443 | Positive | Positive |
| 4 | 83 | 2451 | Positive | Positive |

[5 rows x 24 columns]

```

[34]: # Top 10 rated products (with review count > 1000)
      # We have already a column named is_popular for that, so

top_rated_popular = (
    df1[df1['is_popular'] == 1]
    .sort_values(by='rating', ascending=False)
    .drop_duplicates(subset='product_id')
    .head(10)[[
        'product_name',
        'rating',
        'rating_count',
        'discounted_price',

```



```

        'is_highly_rated',
        'is_popular'
    ]]
)
topRatedPopular.reset_index(drop=True, inplace=True)
topRatedPopular

```

```

[34]:
           product_name  rating  rating_count  \
0  Spigen EZ Fit Tempered Glass Screen Protector ...    4.7          7779
1  Sony Bravia 164 cm (65 inches) 4K Ultra HD Sma...    4.7          5935
2  Logitech M331 Silent Plus Wireless Mouse, 2.4G...    4.6         12375
3  Spigen EZ Fit Tempered Glass Screen Protector ...    4.6          6129
4  Redgear MP35 Speed-Type Gaming Mousepad (Black...    4.6         33434
5  Logitech Pebble M350 Wireless Mouse with Bluet...    4.6         10652
6  10k 8k 4k HDMI Cable, Certified 48Gbps 1ms Ult...    4.6          3664
7  Spigen EZ Fit Tempered Glass Screen Protector ...    4.6         26603
8  Seagate Expansion 1TB External HDD - USB 3.0 f...    4.5         50810
9  Dell MS116 1000Dpi USB Wired Optical Mouse, Le...    4.5         33176

   discounted_price  is_highly_rated  is_popular
0             999.0                1           1
1          77990.0                1           1
2           1295.0                1           1
3             999.0                1           1
4             299.0                1           1
5           1490.0                1           1
6             999.0                1           1
7             999.0                1           1
8           4098.0                1           1
9             299.0                1           1

```

```

[35]: # Average discount per sub-category

avg_discount_per_sub_category = df1.groupby('sub_category')['discount_amount'].
    ↪mean().reset_index().sort_values(by='discount_amount', ascending=False).
    ↪head(10)
avg_discount_per_sub_category

```

```

[35]:
   sub_category  discount_amount
96  SmartTelevisions    15361.887097
81      Projectors      8303.333333
60      Monitors      8231.000000
101  StandardTelevisions    8148.833333
99  SoundbarSpeakers    7500.000000
97      SmartWatches    6271.253333
98      Smartphones    4838.955882
107  Tabletop&TravelTripods    3191.000000

```

| | | |
|-----|------------|-------------|
| 105 | StylusPens | 2589.142857 |
| 71 | Over-Ear | 2525.000000 |

```
[36]: # Sentiment distribution from titles
sentiment_distribution = df1['title_sentiment'].value_counts().reset_index()
sentiment_distribution.columns = ['Sentiment', 'Count']
sentiment_distribution
```

```
[36]: Sentiment  Count
0  Positive    752
1  Neutral    149
2  Negative     31
```

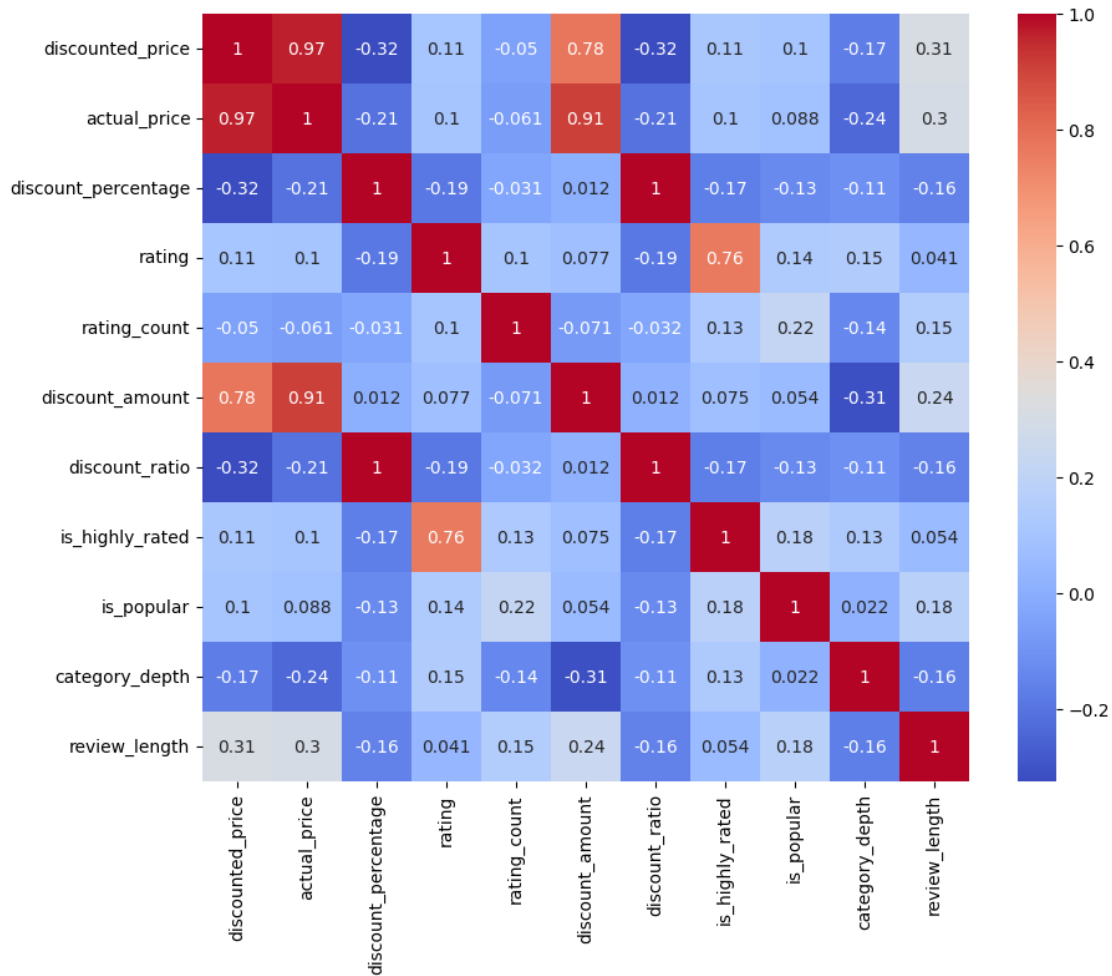
```
[37]: # Average review length by category
avg_review_length_by_category = (
    df1.groupby('main_category')['review_length']
        .mean()
        .reset_index()
        .sort_values(by='review_length', ascending=False)
)
avg_review_length_by_category
```

```
[37]:      main_category  review_length
4  MusicalInstruments    5118.500000
1           Electronics    1808.966469
0  Computers&Accessories    1183.447837
6           Toys&Games    1001.000000
5      OfficeProducts     639.272727
3      HomeImprovement     480.000000
2      Home&Kitchen     409.500000
```

```
[38]: # Correlation: Discount vs. Rating
correlation = df1['discount_percentage'].corr(df1['rating'])
print("Correlation between Discount and rating", round(correlation, 4))
```

Correlation between Discount and rating -0.1854

```
[41]: # Correlation By Heatmap
numeric_df = df1.select_dtypes(include=['int64', 'float64'])
correlation_matrix = numeric_df.corr()
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', square=True)
plt.show()
```



```
[43]: # Price bucket vs. avg rating
avg_rating_by_bucket = (
    df1.groupby('price_bucket')['rating']
    .mean()
    .reset_index()
    .sort_values(by='rating', ascending=False)
)
avg_rating_by_bucket
```

```
[43]: price_bucket  rating
3      Premium  4.145679
2         Mid   4.135242
0        High   4.124242
1         Low   4.096171
```

```
[49]: # Popularity Score (custom feature)
df1['popularity_score'] = df1['rating'] * df1['rating_count']

top_popular_products = (
    df1.sort_values(by='popularity_score', ascending=False)
    .drop_duplicates(subset='product_id')
    .head(10)[[
        'product_name',
        'rating',
        'rating_count',
        'popularity_score',
        'discounted_price',
        'price_bucket'
    ]])
top_popular_products.reset_index(drop=True, inplace=True)
top_popular_products
```

```
[49]:
```

| | product_name | rating | rating_count \ |
|---|---|--------|----------------|
| 0 | AmazonBasics Flexible Premium HDMI Cable (Blac... | 4.4 | 426973 |
| 1 | Amazon Basics High-Speed HDMI Cable, 6 Feet - ... | 4.4 | 426973 |
| 2 | Amazon Basics High-Speed HDMI Cable, 6 Feet (2... | 4.4 | 426973 |
| 3 | boAt Bassheads 100 in Ear Wired Earphones with... | 4.1 | 363713 |
| 4 | boAt Bassheads 100 in Ear Wired Earphones with... | 4.1 | 363713 |
| 5 | boAt BassHeads 100 in-Ear Wired Headphones wit... | 4.1 | 363711 |
| 6 | Redmi 9 Activ (Carbon Black, 4GB RAM, 64GB Sto... | 4.1 | 313836 |
| 7 | Redmi 9A Sport (Coral Green, 2GB RAM, 32GB Sto... | 4.1 | 313836 |
| 8 | Redmi 9A Sport (Coral Green, 3GB RAM, 32GB Sto... | 4.1 | 313832 |
| 9 | Redmi 9A Sport (Carbon Black, 2GB RAM, 32GB St... | 4.1 | 313832 |

| | popularity_score | discounted_price | price_bucket |
|---|------------------|------------------|--------------|
| 0 | 1878681.2 | 219.0 | Low |
| 1 | 1878681.2 | 309.0 | Low |
| 2 | 1878681.2 | 309.0 | Low |
| 3 | 1491223.3 | 349.0 | Low |
| 4 | 1491223.3 | 379.0 | Low |
| 5 | 1491215.1 | 365.0 | Low |
| 6 | 1286727.6 | 8499.0 | Premium |
| 7 | 1286727.6 | 6499.0 | Premium |
| 8 | 1286711.2 | 7499.0 | Premium |
| 9 | 1286711.2 | 6499.0 | Premium |

```
[50]: # Same Thing With Some Sklearn Preprocessing
from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler()
df1['popularity_score_scaled'] = scaler.fit_transform(df1[['popularity_score']])
df1['popularity_score_scaled'] = (df1['popularity_score_scaled'] * 100).round(2)
```

```

top_scaled_popular = (
    df1.sort_values(by='popularity_score_scaled', ascending=False)
        .drop_duplicates(subset='product_id')
        .head(10)[[
            'product_name',
            'rating',
            'rating_count',
            'popularity_score',
            'popularity_score_scaled',
            'discounted_price',
            'price_bucket'
        ]])
top_scaled_popular.reset_index(drop=True, inplace=True)
top_scaled_popular

```

```

[50]:

```

| | product_name | rating | rating_count | \ |
|---|---|--------|--------------|---|
| 0 | AmazonBasics Flexible Premium HDMI Cable (Blac... | 4.4 | 426973 | |
| 1 | Amazon Basics High-Speed HDMI Cable, 6 Feet - ... | 4.4 | 426973 | |
| 2 | Amazon Basics High-Speed HDMI Cable, 6 Feet (2... | 4.4 | 426973 | |
| 3 | boAt BassHeads 100 in-Ear Wired Headphones wit... | 4.1 | 363711 | |
| 4 | boAt Bassheads 100 in Ear Wired Earphones with... | 4.1 | 363713 | |
| 5 | boAt Bassheads 100 in Ear Wired Earphones with... | 4.1 | 363713 | |
| 6 | Redmi 9A Sport (Coral Green, 3GB RAM, 32GB Sto... | 4.1 | 313832 | |
| 7 | Redmi 9A Sport (Coral Green, 2GB RAM, 32GB Sto... | 4.1 | 313836 | |
| 8 | Redmi 9 Activ (Carbon Black, 4GB RAM, 64GB Sto... | 4.1 | 313836 | |
| 9 | Redmi 9A Sport (Carbon Black, 2GB RAM, 32GB St... | 4.1 | 313832 | |

| | popularity_score | popularity_score_scaled | discounted_price | price_bucket |
|---|------------------|-------------------------|------------------|--------------|
| 0 | 1878681.2 | 100.00 | 219.0 | Low |
| 1 | 1878681.2 | 100.00 | 309.0 | Low |
| 2 | 1878681.2 | 100.00 | 309.0 | Low |
| 3 | 1491215.1 | 79.38 | 365.0 | Low |
| 4 | 1491223.3 | 79.38 | 379.0 | Low |
| 5 | 1491223.3 | 79.38 | 349.0 | Low |
| 6 | 1286711.2 | 68.49 | 7499.0 | Premium |
| 7 | 1286727.6 | 68.49 | 6499.0 | Premium |
| 8 | 1286727.6 | 68.49 | 8499.0 | Premium |
| 9 | 1286711.2 | 68.49 | 6499.0 | Premium |

- 12 *So many low price product, that's quite surprising and interesting.*
- 13 *So that's it for now, we will meet in power bi with an awesome dashboard.*