

# Amazon Brazil Products 2023

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

```
In [2]: df = pd.read_csv(r"C:\Users\Admin\Downloads\amz_br_total_products_data_processed.csv", engine = 'python', encod
```

```
In [3]: df
```

```
Out[3]:
```

	asin	title	imgUrl	productURL	stars	reviews
0	B0C154VR2L	SAMSUNG Smart TV Crystal 50" 4K UHD CU7700 - A...	https://m.media-amazon.com/images/I/71vk2qFDSP...	https://www.amazon.com.br/dp/B0C154VR2L	4.6	269
1	B0C85J5HYD	Mini Projetor Portatil 5G Wifi 6 Bluetooth 5.0...	https://m.media-amazon.com/images/I/5133wKaOqS...	https://www.amazon.com.br/dp/B0C85J5HYD	4.5	816
2	B0C1538ZJ4	Samsung Smart TV Crystal 43" 4K UHD CU7700 - A...	https://m.media-amazon.com/images/I/71vk2qFDSP...	https://www.amazon.com.br/dp/B0C1538ZJ4	4.7	638
3	B0B4V379PC	Samsung 60BU8000 - Smart TV LED 60" 4K UHD, Wi...	https://m.media-amazon.com/images/I/8110fzpU-P...	https://www.amazon.com.br/dp/B0B4V379PC	4.6	675
4	B0899BRB4B	Smart TV LED 32" HD Samsung LH32BETBLGGXZD	https://m.media-amazon.com/images/I/61M025WzQd...	https://www.amazon.com.br/dp/B0899BRB4B	4.7	3945
...	...	...	...	...	...	...
1344508	B0CG81Q7YZ	Capa INFILAND Flex Cover Tab S8 Plus 12.4 pol ...	https://m.media-amazon.com/images/I/71L1XW1F6R...	https://www.amazon.com.br/dp/B0CG81Q7YZ	0.0	0
1344509	B0CJ418PM1	Teclado para Notebook compatÃ-vel com Thinkpad...	https://m.media-amazon.com/images/I/61c7FFn0Ji...	https://www.amazon.com.br/dp/B0CJ418PM1	0.0	0
1344510	B09YHX87RM	KKcare Tablet grÃ-fico LED A5 Light Pad Digita...	https://m.media-amazon.com/images/I/616v6AkvNV...	https://www.amazon.com.br/dp/B09YHX87RM	0.0	0
1344511	B0CKS1GBMM	Bateria para Notebook Ne57007b	https://m.media-amazon.com/images/I/41NnZx7wg+...	https://www.amazon.com.br/dp/B0CKS1GBMM	0.0	0
1344512	B00P10QYCC	Plantronics headset qd mono encorepro hw510 ca...	https://m.media-amazon.com/images/I/51EHaYprj1...	https://www.amazon.com.br/dp/B00P10QYCC	4.2	26

1344513 rows × 11 columns

```
In [4]: # Columns RENAME name
df.rename(columns={'productURL':'Product URL','title':'Title','imgUrl':'Image URL','stars':'Stars','reviews':'R
```

```
In [5]: # List of Columns Name
df.columns
```

```
Out[5]: Index(['asin', 'Title', 'Image URL', 'Product URL', 'Stars', 'Reviews',
        'Price', 'List Price', 'Category Name', 'Best Seller',
        'Bought In Last Month'],
        dtype='object')
```

```
In [6]: # Describe the table
df.describe
```

```

Out[6]: <bound method NDFrame.describe of          Title \
0      B0C154VR2L  SAMSUNG Smart TV Crystal 50" 4K UHD CU7700 - A...
1      B0C85J5HYD  Mini Projetor Portatil 5G Wifi 6 Bluetooth 5.0...
2      B0C1538ZJ4  Samsung Smart TV Crystal 43" 4K UHD CU7700 - A...
3      B0B4V379PC  Samsung 60BU8000 - Smart TV LED 60' 4K UHD, Wi...
4      B0899BRB4B  Smart TV LED 32'' HD Samsung LH32BETBLGGXZD
...
1344508 B0CG81Q7YZ  Capa INFILAND Flex Cover Tab S8 Plus 12.4 pol ...
1344509 B0CJ418PM1  Teclado para Notebook compatÃvel com Thinkpad...
1344510 B09YHX87RM  KKcare Tablet grÃfico LED A5 Light Pad Digita...
1344511 B0CKS1GBMM  Bateria para Notebook Ne57007b
1344512 B00P10QYCC  Plantronics headset qd mono encorepro hw510 ca...

          Image URL \
0      https://m.media-amazon.com/images/I/71vk2qFDSP...
1      https://m.media-amazon.com/images/I/5133wKa0qS...
2      https://m.media-amazon.com/images/I/71vk2qFDSP...
3      https://m.media-amazon.com/images/I/8110fzpU-P...
4      https://m.media-amazon.com/images/I/61M025WzQd...
...
1344508 https://m.media-amazon.com/images/I/71L1XW1F6R...
1344509 https://m.media-amazon.com/images/I/61c7FFn0Ji...
1344510 https://m.media-amazon.com/images/I/616v6AkvNV...
1344511 https://m.media-amazon.com/images/I/41NnZx7wg+...
1344512 https://m.media-amazon.com/images/I/51EHaYprjI...

          Product URL  Stars  Reviews  Price \
0      https://www.amazon.com.br/dp/B0C154VR2L  4.6    269    2279.05
1      https://www.amazon.com.br/dp/B0C85J5HYD  4.5    816    409.99
2      https://www.amazon.com.br/dp/B0C1538ZJ4  4.7    638    1969.00
3      https://www.amazon.com.br/dp/B0B4V379PC  4.6    675    3099.00
4      https://www.amazon.com.br/dp/B0899BRB4B  4.7    3945   1079.90
...
1344508 https://www.amazon.com.br/dp/B0CG81Q7YZ  0.0     0    229.00
1344509 https://www.amazon.com.br/dp/B0CJ418PM1  0.0     0    235.34
1344510 https://www.amazon.com.br/dp/B09YHX87RM  0.0     0    81.89
1344511 https://www.amazon.com.br/dp/B0CKS1GBMM  0.0     0    139.41
1344512 https://www.amazon.com.br/dp/B00P10QYCC  4.2    26    249.09

          List Price  Category Name  Best Seller \
0      0.0  TV, Ãudio e Cinema em Casa  False
1      0.0  TV, Ãudio e Cinema em Casa  True
2      2499.0  TV, Ãudio e Cinema em Casa  False
3      0.0  TV, Ãudio e Cinema em Casa  False
4      1369.0  TV, Ãudio e Cinema em Casa  True
...
1344508 254.0  Computadores e InformÃtica  False
1344509 0.0  Computadores e InformÃtica  False
1344510 0.0  Computadores e InformÃtica  False
1344511 0.0  Computadores e InformÃtica  False
1344512 0.0  Computadores e InformÃtica  False

          Bought In Last Month
0      0
1      5000
2      0
3      0
4      0
...
1344508 0
1344509 0
1344510 0
1344511 0
1344512 0

```

```
[1344513 rows x 11 columns]>
```

```

In [7]: # Statisctical Information
df.describe().transpose()

```

```

Out[7]:

```

	count	mean	std	min	25%	50%	75%	max
Stars	1344513.0	1.471351	2.107560	0.0	0.0	0.0	4.20	5.0
Reviews	1344513.0	212.230983	4581.190312	0.0	0.0	0.0	3.00	1878552.0
Price	1344513.0	217.856979	856.208637	0.0	25.0	69.0	153.99	100000.0
List Price	1344513.0	65.303842	2874.167154	0.0	0.0	0.0	0.00	1999000.0
Bought In Last Month	1344513.0	4.350088	94.358861	0.0	0.0	0.0	0.00	40000.0

```

In [8]: # check Null Values
df.isnull().sum()

```

```
Out[8]: asin                0
        Title              1
        Image URL          0
        Product URL        0
        Stars              0
        Reviews            0
        Price              0
        List Price         0
        Category Name      0
        Best Seller        0
        Bought In Last Month 0
        dtype: int64
```

```
In [9]: # delete the null values
df1 = df.dropna()
```

```
In [10]: # check the null values after the delete in table
df1.isnull().sum()
```

```
Out[10]: asin                0
        Title              0
        Image URL          0
        Product URL        0
        Stars              0
        Reviews            0
        Price              0
        List Price         0
        Category Name      0
        Best Seller        0
        Bought In Last Month 0
        dtype: int64
```

```
In [11]: # check table size
df1.shape
```

```
Out[11]: (1344512, 11)
```

```
In [12]: # give the full information of table
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 1344512 entries, 0 to 1344512
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  -
0   asin                  1344512 non-null object
1   Title                 1344512 non-null object
2   Image URL             1344512 non-null object
3   Product URL           1344512 non-null object
4   Stars                 1344512 non-null float64
5   Reviews               1344512 non-null int64
6   Price                 1344512 non-null float64
7   List Price            1344512 non-null float64
8   Category Name         1344512 non-null object
9   Best Seller           1344512 non-null bool
10  Bought In Last Month  1344512 non-null int64
dtypes: bool(1), float64(3), int64(2), object(5)
memory usage: 114.1+ MB
```

```
In [13]: # top data from table
df1.head()
```

```
Out[13]:
```

	asin	Title	Image URL	Product URL	Stars	Reviews	Price	
0	B0C154VR2L	SAMSUNG Smart TV Crystal 50" 4K UHD CU7700 - A...	amazon.com/images/I/71vk2qFDSP...	https://m.media-amazon.com/images/I/71vk2qFDSP... https://www.amazon.com.br/dp/B0C154VR2L	4.6	269	2279.05	
1	B0C85J5HYD	Mini Projetor Portatil 5G Wifi 6 Bluetooth 5.0...	amazon.com/images/I/5133wKaQsS...	https://m.media-amazon.com/images/I/5133wKaQsS... https://www.amazon.com.br/dp/B0C85J5HYD	4.5	816	409.99	
2	B0C1538ZJ4	Samsung Smart TV Crystal 43" 4K UHD CU7700 - A...	amazon.com/images/I/71vk2qFDSP...	https://m.media-amazon.com/images/I/71vk2qFDSP... https://www.amazon.com.br/dp/B0C1538ZJ4	4.7	638	1969.00	24
3	B0B4V379PC	Samsung 60BU8000 - Smart TV LED 60" 4K UHD, Wi...	amazon.com/images/I/8110fzpU-P...	https://m.media-amazon.com/images/I/8110fzpU-P... https://www.amazon.com.br/dp/B0B4V379PC	4.6	675	3099.00	
4	B0899BRB4B	Smart TV LED 32" HD Samsung LH32BETBLGGXZD	amazon.com/images/I/61M025WzQd...	https://m.media-amazon.com/images/I/61M025WzQd... https://www.amazon.com.br/dp/B0899BRB4B	4.7	3945	1079.90	13

```
In [14]: # Bottom top data
df1.tail()
```

Out[14]:

	asin	Title	Image URL	Product URL	Stars	Reviews	Price	l Pr
1344508	B0CG81Q7YZ	Capa INFILAND Flex Cover Tab S8 Plus 12.4 pol ...	amazon.com/images/I/71L1XW1F6R... https://m.media-	https://www.amazon.com.br/dp/B0CG81Q7YZ	0.0	0	229.00	25
1344509	B0CJ418PM1	Teclado para Notebook compatÃ-vel com Thinkpad...	amazon.com/images/I/61c7FFn0Ji... https://m.media-	https://www.amazon.com.br/dp/B0CJ418PM1	0.0	0	235.34	
1344510	B09YHX87RM	KKcare Tablet grÃ¡fico LED A5 Light Pad Digita...	amazon.com/images/I/616v6AkVNV... https://m.media-	https://www.amazon.com.br/dp/B09YHX87RM	0.0	0	81.89	
1344511	B0CKS1GBMM	Bateria para Notebook Ne57007b	amazon.com/images/I/41NnZx7wg+... https://m.media-	https://www.amazon.com.br/dp/B0CKS1GBMM	0.0	0	139.41	
1344512	B00P10QYCC	Plantronics headset qd mono encorepro hw510 ca...	amazon.com/images/I/51EHaYprjl... https://m.media-	https://www.amazon.com.br/dp/B00P10QYCC	4.2	26	249.09	

```
In [15]: # show unique values of columns
df1['asin'].unique()
```

Out[15]: array(['B0C154VR2L', 'B0C85J5HYD', 'B0C1538ZJ4', ..., 'B09YHX87RM', 'B0CKS1GBMM', 'B00P10QYCC'], dtype=object)

```
In [16]: # count the total values
df1['asin'].value_counts()
```

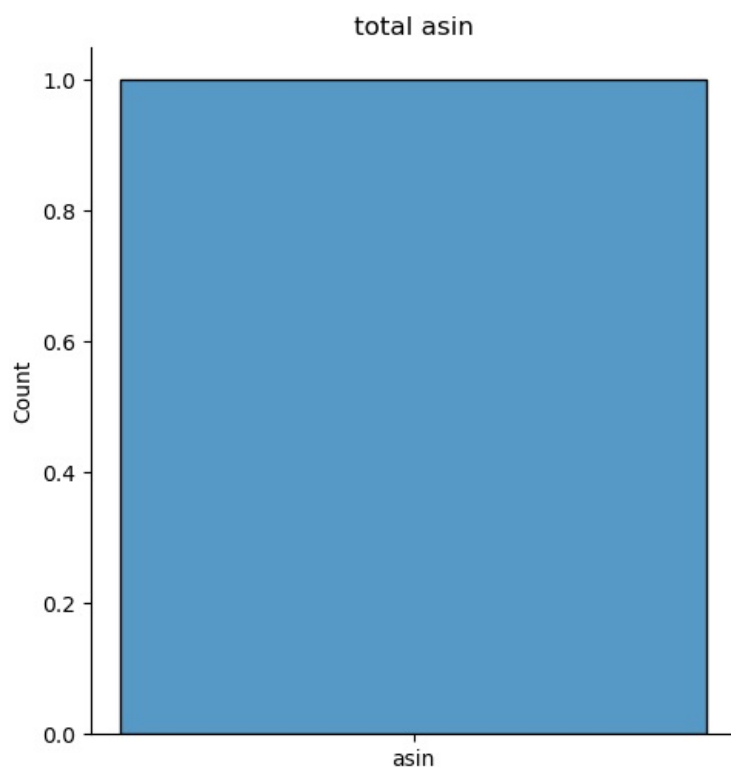
Out[16]:

asin	
B0C154VR2L	1
B08ZM1G6DP	1
B08L9MYQ4T	1
B08TR76NFT	1
B079H2H7P8	1
..	
B0C3WSHPPX	1
B0C9SQV2TL	1
B0C9VKVKS1	1
B0CGKN2GKX	1
B00P10QYCC	1

Name: count, Length: 1344512, dtype: int64

```
In [17]: # check the indivual references bar chart of columns
plt.figure(figsize=(10,5))
sns.displot('asin')
plt.title('total asin')
plt.show()
```

C:\Users\Admin\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning: The figure layout has changed to tight  
self.\_figure.tight\_layout(\*args, \*\*kwargs)  
<Figure size 1000x500 with 0 Axes>



```
In [18]: df['Title'].value_counts()
```

```
Out[18]: Title
Mochila escolar durável para faculdade, mochila casual resistente à água para mulheres e homens, mochila para laptop campus para viagens, negócios, caminhadas, Multicor, Large 184
Óculos de Titânio Armação Feminina Lentes Fotocromáticas Rosa Roxo cor Oculos Miopia -3.0 Oculos Femininos 90
Lona transparente com ilhãs, lona de cobertura à prova d'água, capa de chuva multifuncional externa de 0,35 mm, cobertura de dossel de planta de pátio de toldo externo (2,8 x 4 m/9,2 x 13 pés) () 63
Tênis masculino 62
Sandália Sandália My shoes, My shoes, feminino 61
...
ADOCARN Removedor De Unhas 3 Peças Kit De Pedicure Doméstico Ferramentas Domésticas Corretor Ferramenta De Remoção De Unhas Encravadas Conjunto De Cutícula De Manicure Ferramenta 1
Beavorty Protetores De Pés De Gel De Silicone Para Pés Protetores De Calcanhar Meias Fascite Plantar Manga De Silicone Sapatos Saltos Meias Cuidados Com Os Pés Meias De Calcanhar De Gel 1
Beavorty Massageador Facial Massageador Elétrico Para Os Olhos Massageador Elétrico Para Os Olhos Mini Massageador Para Os Olhos Caneta De Massagem Bateria Seca Choque Rosa Rugas 1
FRCOLOR 6 Folhas De Adesivos De Unhas Decorações De Unhas Para Arte De Unhas Decalques De Arte De Unhas Adesivos Para Unhas Adesivos Luminosos De Unhas 3D Decalque De Arte De Unhas De 1
Plantronics headset qd mono encorepro hw510 cancelamento de ruído 1
Name: count, Length: 1308168, dtype: int64
```

```
In [19]: df['Title'].unique()
```

```
Out[19]: array(['SAMSUNG Smart TV Crystal 50" 4K UHD CU7700 - Alexa built in, Samsung Gaming Hub',
      'Mini Projetor Portatil 5G Wifi 6 Bluetooth 5.0 Android 11, Projetor 4K 1080P Full HD Suporte 8000 Lumen
s, Projetor Led Auto de Correção de Trapezóide Horizontal, 180° Girável Projector para Telefónico',
      'Samsung Smart TV Crystal 43" 4K UHD CU7700 - Alexa built in, Samsung Gaming Hub',
      ...,
      'KKcare Tablet gráfico LED A5 Light Pad Digital Tablet Copyboard com brilho ajustável de 3 níveis
para rastreamento de desenho cãpia visualização arte DIY artesanato joia de diamante suprimentos de pintura'
      ,
      'Bateria para Notebook Ne57007b',
      'Plantronics headset qd mono encorepro hw510 cancelamento de ruído'],
      dtype=object)
```

```
In [20]: df['Stars'].unique()
```

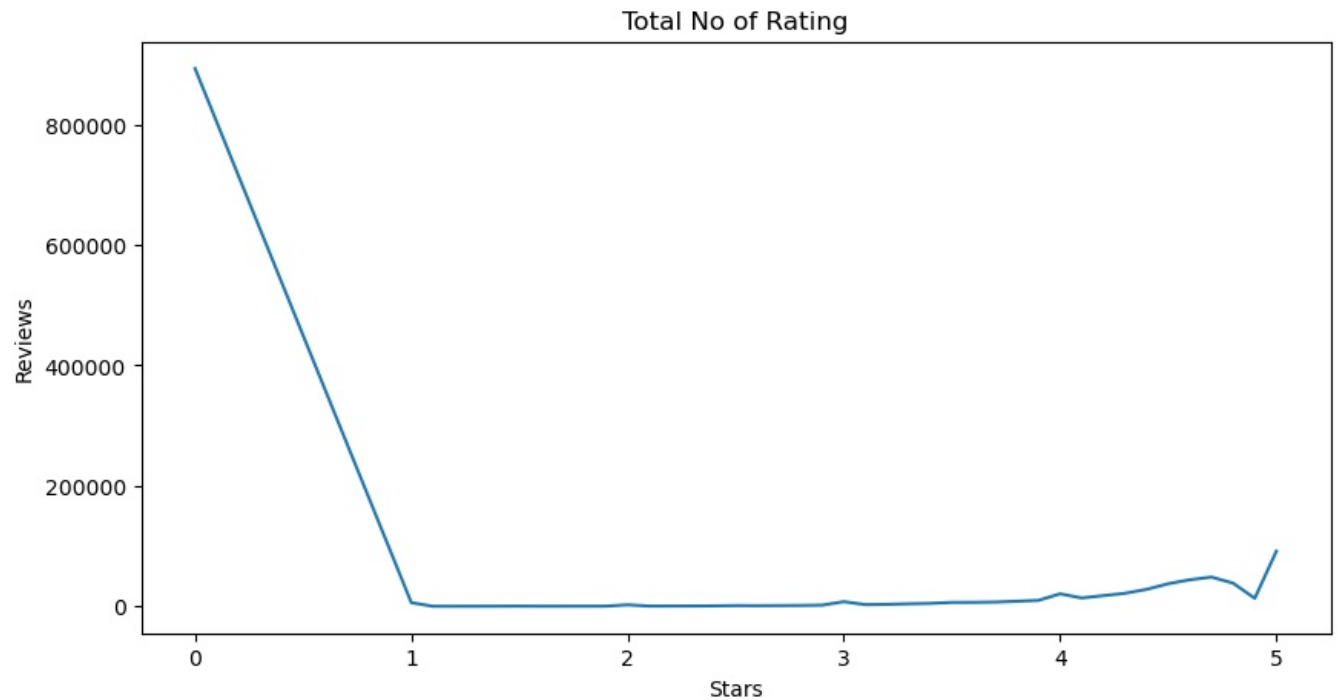
```
Out[20]: array([4.6, 4.5, 4.7, 4.4, 4.3, 4.9, 4.8, 3.9, 4.1, 5. , 4.2, 4. , 3.6,
      3.3, 3.7, 3.8, 3.5, 0. , 3. , 3.4, 2.7, 2.5, 2.9, 3.1, 2. , 1.7,
      3.2, 1. , 2.6, 1.9, 2.1, 2.8, 2.3, 1.8, 2.4, 2.2, 1.6, 1.4, 1.5,
      1.2, 1.3, 1.1])
```

```
In [21]: df['Stars'].value_counts()
```

```
Out[21]: Stars
0.0    892728
5.0     91408
4.7     48410
4.6     43997
4.8     38335
4.5     37354
4.4     28053
4.3     21499
4.0     20619
4.2     17602
4.1     13827
4.9     13229
3.9      9804
3.8      8292
3.0      7586
3.7      6939
3.6      6455
3.5      6373
1.0      5948
3.4      4787
3.3      4082
3.2      3159
3.1      2725
2.0      2373
2.9      1848
2.8      1351
2.7      1134
2.5      1106
2.6       949
2.4       631
2.3       420
2.2       338
2.1       305
1.5       224
1.9       170
1.8       131
1.4       102
1.7        92
1.6        77
1.3        42
1.2         8
1.1         1
Name: count, dtype: int64
```

```
In [37]: #Total No of Rating
model_year_counts = df.groupby('Stars')['Reviews'].count()
plt.figure(figsize=(10,5))
plt.plot(model_year_counts.index,model_year_counts.values)
plt.xlabel('Stars')
plt.ylabel('Reviews')
plt.title('Total No of Rating')
```

```
Out[37]: Text(0.5, 1.0, 'Total No of Rating')
```



```
In [23]: df['Price'].value_counts()
```

```
Out[23]: Price
0.00      264114
152.17     7929
39.90     4701
209.00     4435
49.90     3999
...
829.08         1
435.51         1
715.45         1
1044.76        1
1986.90        1
Name: count, Length: 85794, dtype: int64
```

```
In [24]: df['Price'].unique()
```

```
Out[24]: array([2279.05, 409.99, 1969. , ..., 7199.9 , 3169.22, 1986.9 ])
```

```
In [25]: df['List Price'].value_counts()
```

```
Out[25]: List Price
0.00      1186414
39.90     1419
59.90     1401
49.90     1321
29.90     1185
...
2136.39         1
140.12         1
150.83         1
216.72         1
1128.00         1
Name: count, Length: 30668, dtype: int64
```

```
In [26]: df['List Price'].unique()
```

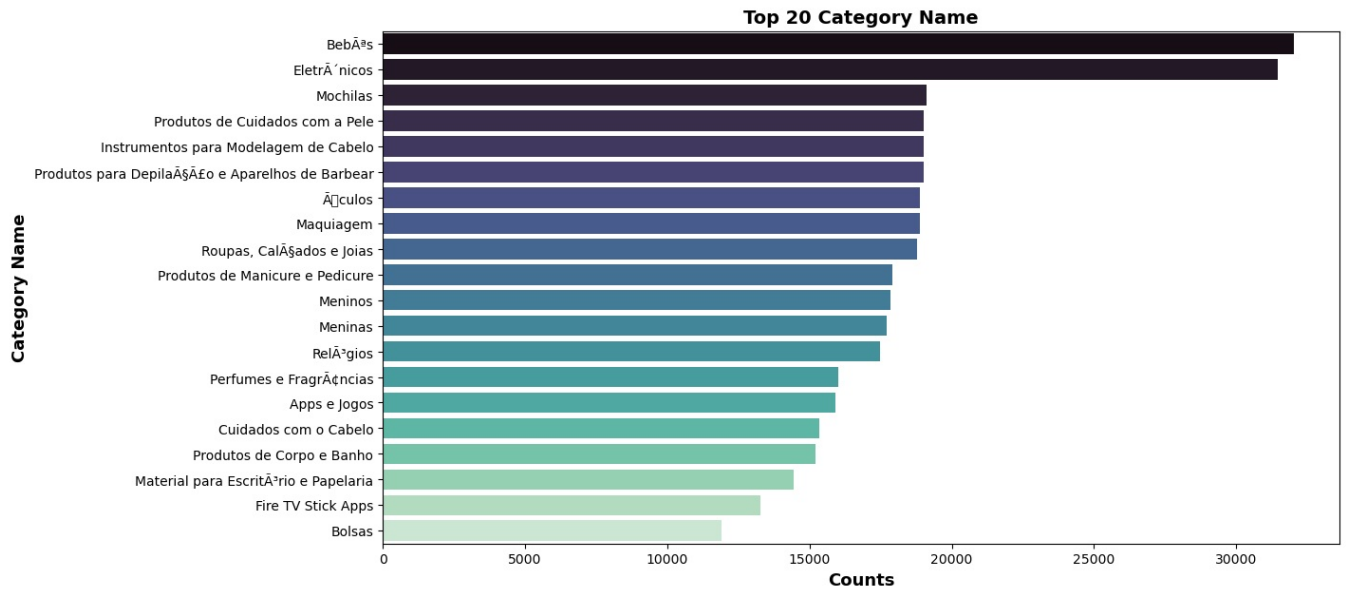
```
Out[26]: array([ 0. , 2499. , 1369. , ..., 182.73, 117.87, 227.03])
```

```
In [27]: aws_counts = df1['Category Name'].value_counts().nlargest(20)
plt.figure(figsize=(13, 7))
sns.barplot(x=aws_counts.values, y=aws_counts.index, palette="mako")
plt.title('Top 20 Category Name', fontweight='bold', fontsize=14)
```

```
plt.xlabel('Counts', fontweight='bold', fontsize=13)
plt.ylabel('Category Name', fontweight='bold', fontsize=13)
plt.show()
```

C:\Users\Admin\anaconda3\Lib\site-packages\IPython\core\pylabtools.py:152: UserWarning: Glyph 147 (\x93) missing from current font.

```
fig.canvas.print_figure(bytes_io, **kw)
```



```
In [28]: df['Stars'].count()
```

```
Out[28]: 1344513
```

```
In [29]: df['Best Seller'].value_counts()
```

```
Out[29]: Best Seller
False    1342124
True      2389
Name: count, dtype: int64
```

```
In [30]: df['Best Seller'].unique()
```

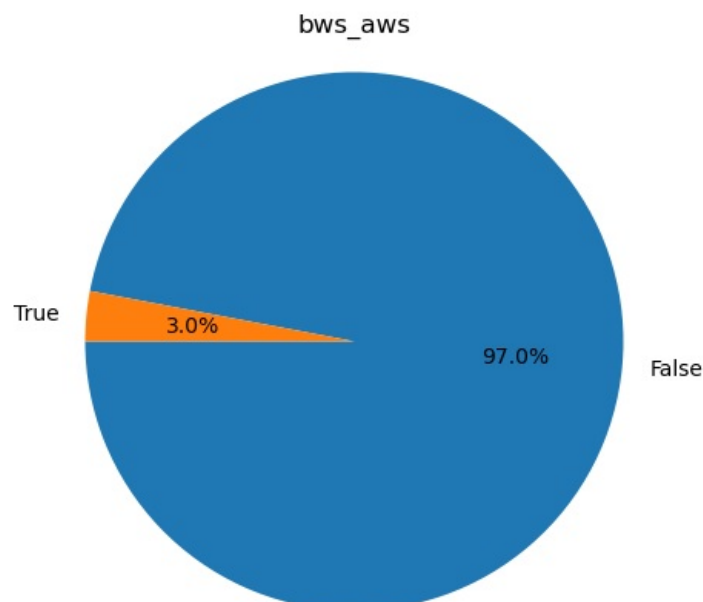
```
Out[30]: array([False,  True])
```

```
In [31]: bws_aws = ['False', 'True']
percentages = [97, 3] # These should add up to 100%

# Create a pie chart
plt.figure(figsize=(10, 5))
plt.pie(percentages, labels=bws_aws, autopct='%1.1f%%', startangle=180)

# Add a title
plt.title("bws_aws")

# Display the pie chart
plt.axis('equal') # Equal aspect ratio ensures that the pie is drawn as a circle.
plt.show()
```





```
In [32]: df['Bought In Last Month'].unique()
```

```
Out[32]: array([ 0, 5000, 2000, 1000, 700, 3000, 300, 200, 800,
        900, 500, 600, 400, 100, 50, 4000, 10000, 9000,
        7000, 6000, 8000, 40000, 30000, 20000], dtype=int64)
```

```
In [33]: top_model = [0, 5000, 2000, 1000, 700, 3000, 300, 200, 800,
                    900, 500, 600, 400, 100, 50, 4000, 10000, 9000,
                    7000, 6000, 8000, 40000, 30000, 20000]

model_evtype = df1[df1["Bought In Last Month"].isin(top_model)]

top_model_evtype = model_evtype.groupby(["Price", "Bought In Last Month"]).size().reset_index(name="Count")

top_model_evtype
```

Out[33]:

	Price	Bought In Last Month	Count
0	0.00	0	263730
1	0.00	50	175
2	0.00	100	118
3	0.00	200	31
4	0.00	300	19
...	...	...	...
100153	67020.00	0	1
100154	67858.78	0	1
100155	73434.41	0	1
100156	85916.98	0	1
100157	100000.00	0	1

100158 rows × 3 columns

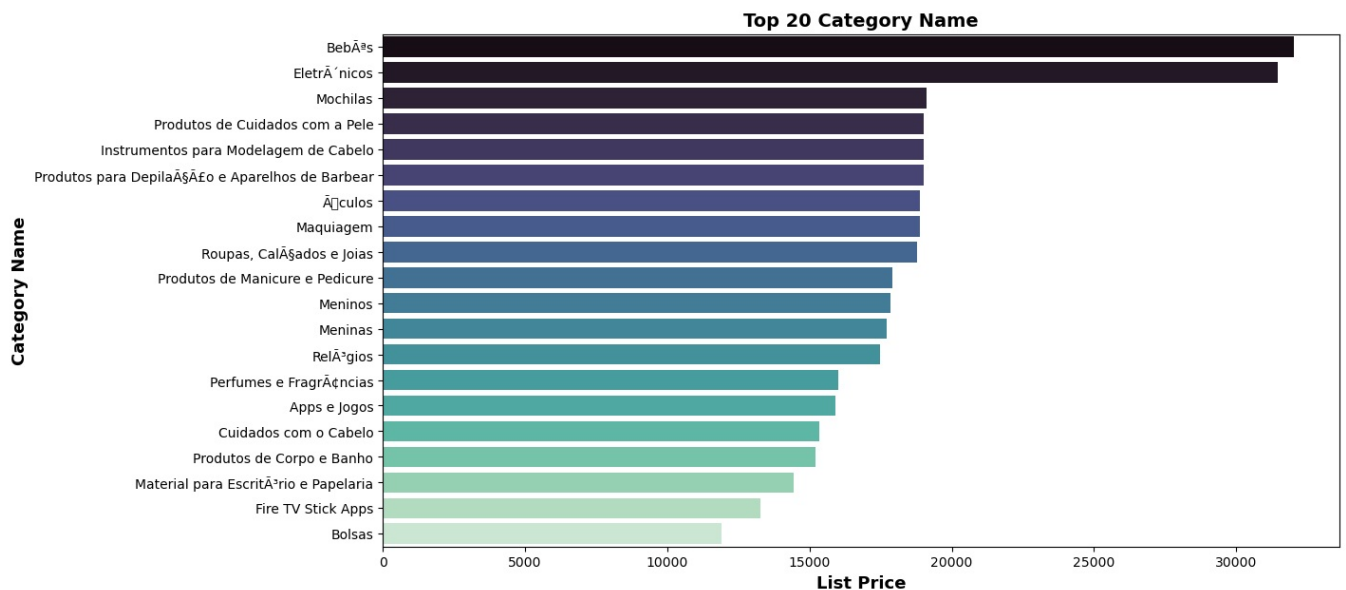
```
In [35]: df.sample(1).T
```

Out[35]:

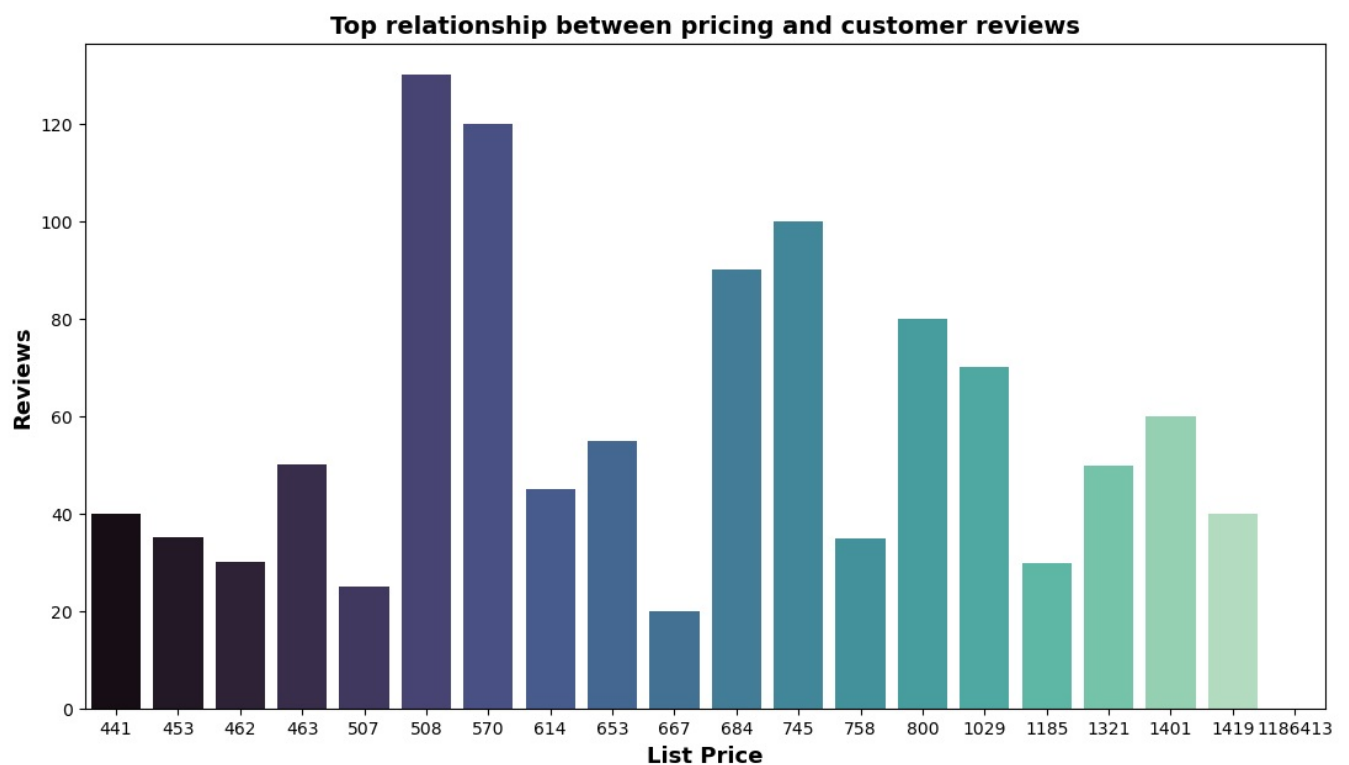
	778781
asin	B0CJBLG5S3
Title	KKcare Carregador USB para motocicleta Super F...
Image URL	https://m.media-amazon.com/images/I/617D3SMKut...
Product URL	https://www.amazon.com.br/dp/B0CJBLG5S3
Stars	0.0
Reviews	0
Price	88.49
List Price	0.0
Category Name	Carregadores de Celular
Best Seller	False
Bought In Last Month	0

```
In [36]: # List Price vs trend of Category Name
aws_counts = df1['Category Name'].value_counts().nlargest(20)
plt.figure(figsize=(13, 7))
sns.barplot(x=aws_counts.values, y=aws_counts.index, palette="mako")
plt.title('Top 20 Category Name', fontweight='bold', fontsize=14)
plt.xlabel('List Price', fontweight='bold', fontsize=13)
plt.ylabel('Category Name', fontweight='bold', fontsize=13)
plt.show()
```

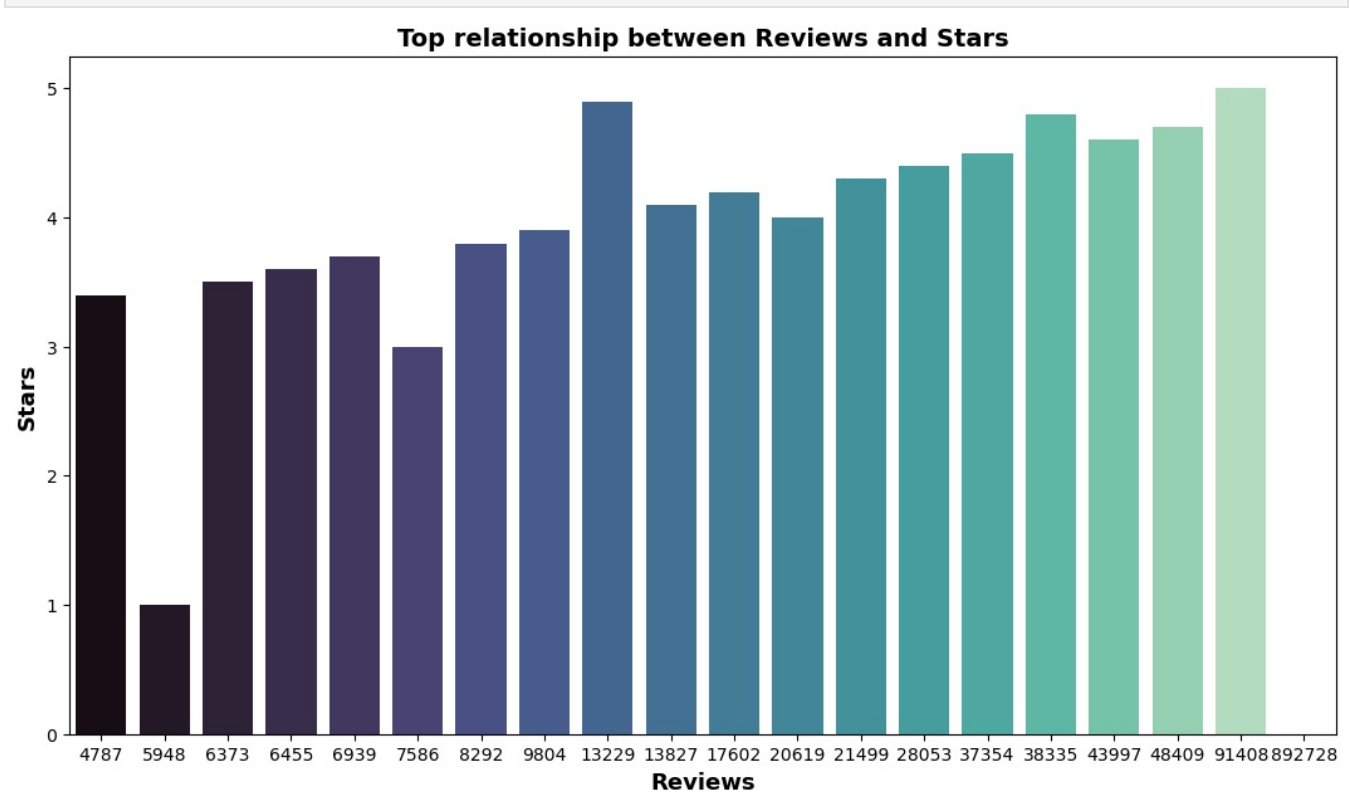
C:\Users\Admin\anaconda3\Lib\site-packages\IPython\core\pylabtools.py:152: UserWarning: Glyph 147 (\x93) missing from current font.  
fig.canvas.print\_figure(bytes\_io, \*\*kw)



```
In [38]: # Study the relationship between pricing and customer reviews
aws_counts = df1['List Price'].value_counts().nlargest(20)
plt.figure(figsize=(13, 7))
sns.barplot(x=aws_counts.values, y=aws_counts.index, palette="mako")
plt.title('Top relationship between pricing and customer reviews', fontweight='bold', fontsize=14)
plt.xlabel('List Price', fontweight='bold', fontsize=13)
plt.ylabel('Reviews', fontweight='bold', fontsize=13)
plt.show()
```



```
In [42]: # Top relationship between Reviews and Stars
aws_counts = df1['Stars'].value_counts().nlargest(20)
plt.figure(figsize=(13, 7))
sns.barplot(x=aws_counts.values, y=aws_counts.index, palette="mako")
plt.title('Top relationship between Reviews and Stars', fontweight='bold', fontsize=14)
plt.xlabel('Reviews', fontweight='bold', fontsize=13)
plt.ylabel('Stars', fontweight='bold', fontsize=13)
plt.show()
```



In [ ]:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js