

```

# IMPORTANT: RUN THIS CELL IN ORDER TO IMPORT YOUR KAGGLE DATA SOURCES
# TO THE CORRECT LOCATION (/kaggle/input) IN YOUR NOTEBOOK,
# THEN FEEL FREE TO DELETE THIS CELL.
# NOTE: THIS NOTEBOOK ENVIRONMENT DIFFERS FROM KAGGLE'S PYTHON
# ENVIRONMENT SO THERE MAY BE MISSING LIBRARIES USED BY YOUR
# NOTEBOOK.

import os
import sys
from tempfile import NamedTemporaryFile
from urllib.request import urlopen
from urllib.parse import unquote, urlparse
from urllib.error import HTTPError
from zipfile import ZipFile
import tarfile
import shutil

CHUNK_SIZE = 40960
DATA_SOURCE_MAPPING = 'amazon-product-reviews:https%3A%2F%2Fstorage.googleapis.com%2Fkaggle-data-sets%2F1181271%2F1976759%2Fbundle%2Far'

KAGGLE_INPUT_PATH='/kaggle/input'
KAGGLE_WORKING_PATH='/kaggle/working'
KAGGLE_SYMLINK='kaggle'

!umount /kaggle/input/ 2> /dev/null
shutil.rmtree('/kaggle/input', ignore_errors=True)
os.makedirs(KAGGLE_INPUT_PATH, 0o777, exist_ok=True)
os.makedirs(KAGGLE_WORKING_PATH, 0o777, exist_ok=True)

try:
    os.symlink(KAGGLE_INPUT_PATH, os.path.join(".", 'input'), target_is_directory=True)
except FileExistsError:
    pass
try:
    os.symlink(KAGGLE_WORKING_PATH, os.path.join(".", 'working'), target_is_directory=True)
except FileExistsError:
    pass

for data_source_mapping in DATA_SOURCE_MAPPING.split(','):
    directory, download_url_encoded = data_source_mapping.split(':')
    download_url = unquote(download_url_encoded)
    filename = urlparse(download_url).path
    destination_path = os.path.join(KAGGLE_INPUT_PATH, directory)
    try:
        with urlopen(download_url) as fileres, NamedTemporaryFile() as tfile:
            total_length = fileres.headers['content-length']
            print(f'Downloading {directory}, {total_length} bytes compressed')
            dl = 0
            data = fileres.read(CHUNK_SIZE)
            while len(data) > 0:
                dl += len(data)
                tfile.write(data)
                done = int(50 * dl / int(total_length))
                sys.stdout.write(f"\r[{ '=' * done }{' ' * (50-done)}] {dl} bytes downloaded")
                sys.stdout.flush()
                data = fileres.read(CHUNK_SIZE)
            if filename.endswith('.zip'):
                with ZipFile(tfile) as zfile:
                    zfile.extractall(destination_path)
            else:
                with tarfile.open(tfile.name) as tarfile:
                    tarfile.extractall(destination_path)
            print(f'\nDownloaded and uncompressed: {directory}')
    except HTTPError as e:
        print(f'Failed to load (likely expired) {download_url} to path {destination_path}')
        continue
    except OSError as e:
        print(f'Failed to load {download_url} to path {destination_path}')
        continue

print('Data source import complete.')

Downloading amazon-product-reviews, 3697835 bytes compressed
[=====] 3697835 bytes downloaded
Downloaded and uncompressed: amazon-product-reviews
Data source import complete.

import matplotlib.pyplot as plt
import pandas as pd

```

```
import numpy as np
import seaborn as sns
import math
import warnings
warnings.filterwarnings('ignore') # Hides warning
warnings.filterwarnings("ignore", category=DeprecationWarning)
warnings.filterwarnings("ignore",category=UserWarning)
sns.set_style("whitegrid") # Plotting style
np.random.seed(42) # seeding random number generator

df = pd.read_csv('../input/amazon-product-reviews/amazon.csv')
df.head()
```

	id	name	asins	brand	categories
0	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...
1	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...
2	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...
3	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...
4	AVqklhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...

5 rows × 21 columns

```
data = df.copy()
data.describe()
```

	reviews.id	reviews.numHelpful	reviews.rating	reviews.userCity	reviews.user
count	1.0	34131.000000	34627.000000	0.0	
mean	111372787.0	0.630248	4.584573	NaN	
std	NaN	13.215775	0.735653	NaN	
min	111372787.0	0.000000	1.000000	NaN	
25%	111372787.0	0.000000	4.000000	NaN	
50%	111372787.0	0.000000	5.000000	NaN	
75%	111372787.0	0.000000	5.000000	NaN	
max	111372787.0	814.000000	5.000000	NaN	

```
data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 34660 entries, 0 to 34659
Data columns (total 21 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    34660 non-null  object
1   name                  27900 non-null  object
```

```

2  asins                34658 non-null object
3  brand                34660 non-null object
4  categories           34660 non-null object
5  keys                 34660 non-null object
6  manufacturer         34660 non-null object
7  reviews.date         34621 non-null object
8  reviews.dateAdded    24039 non-null object
9  reviews.dateSeen     34660 non-null object
10 reviews.didPurchase  1 non-null object
11 reviews.doRecommend  34066 non-null object
12 reviews.id           1 non-null float64
13 reviews.numHelpful   34131 non-null float64
14 reviews.rating       34627 non-null float64
15 reviews.sourceURLs   34660 non-null object
16 reviews.text         34659 non-null object
17 reviews.title        34655 non-null object
18 reviews.userCity     0 non-null float64
19 reviews.userProvince 0 non-null float64
20 reviews.username     34658 non-null object
dtypes: float64(5), object(16)
memory usage: 5.6+ MB

```

```
data["asins"].unique()
```

```

array(['B01AHB9CN2', 'B00VINDBJK', 'B005PB2T0S', 'B002Y27P3M',
      'B01AHB9CYG', 'B01AHB9C1E', 'B01J2G4VBG', 'B00ZV9PXP2',
      'B0083Q04TA', 'B018Y2290U', 'B00REQKWGA', 'B00IOYAM4I',
      'B018T075DC', nan, 'B00DU15MU4', 'B018Y225IA', 'B005PB2T2Q',
      'B018Y23MNM', 'B00QVZDJM', 'B00IOY8XWQ', 'B00LO29KXQ',
      'B00QJDU3KY', 'B018Y22C2Y', 'B01BFIBRIE', 'B01J40RNHU',
      'B018SZT3BK', 'B00UH4D8G2', 'B018Y22BI4', 'B00TSUGXKE',
      'B00L9EPT80', 'B01E6A069U', 'B018Y23P7K', 'B00X4WHP5E', 'B00QFQRELG',
      'B00LW9X0JM', 'B00QL1ZN3G', 'B0189XYY0Q', 'B01BH83OOM',
      'B00BFJAHF8', 'B00U3FPN4U', 'B002Y27P6Y', 'B006GW05NE',
      'B006GW05WK'], dtype=object)

```

```

asins_unique = len(data["asins"].unique())
print("Number of Unique ASINs: " + str(asins_unique))

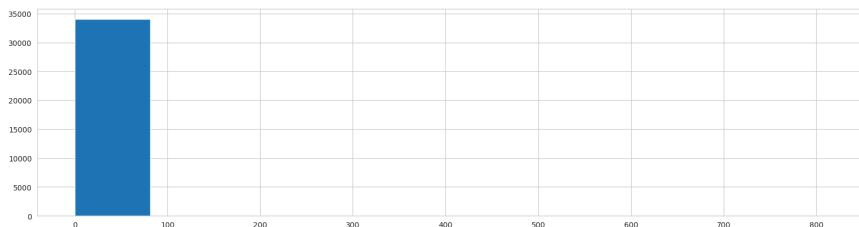
```

```
Number of Unique ASINs: 42
```

```

data["reviews.numHelpful"].hist(figsize=(20,5))
plt.show()

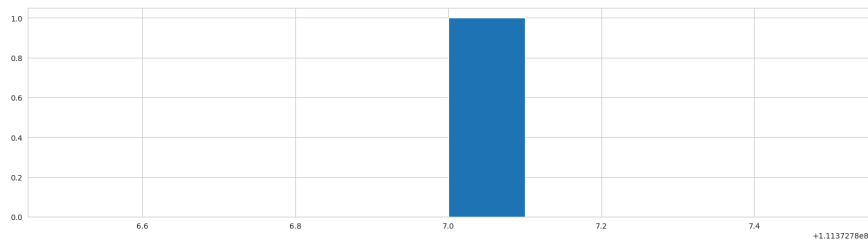
```



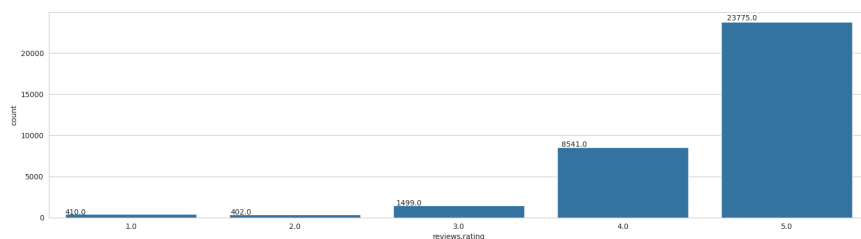
```

data["reviews.id"].hist( figsize=(20,5))
plt.show()

```



```
plt.figure(figsize=(20,5))
ax1=sns.countplot(x="reviews.rating", data=data)
for p in ax1.patches:
    ax1.annotate(str(p.get_height()), (p.get_x() * 1.01 , p.get_height() * 1.01))
plt.show()
```



```
from sklearn.model_selection import StratifiedShuffleSplit
print("Before : {}".format(len(data)))
dataAfter = data.dropna(subset=["reviews.rating"])
# Removes all NAN in reviews.rating
print("After : {}".format(len(dataAfter)))
dataAfter["reviews.rating"] = dataAfter["reviews.rating"].astype(int)

split = StratifiedShuffleSplit(n_splits=10, test_size=0.2)
for train_index, test_index in split.split(dataAfter,
                                           dataAfter["reviews.rating"]):
    strat_train = dataAfter.reindex(train_index)
    strat_test = dataAfter.reindex(test_index)

    Before : 34660
    After : 34627

print(len(strat_train))
print(len(strat_test))
round((strat_test["reviews.rating"].value_counts()* 100/len(strat_test)),2)

27701
6926
5.0    68.81
4.0    24.56
3.0     4.09
2.0     1.29
1.0     1.17
Name: reviews.rating, dtype: float64

reviews = strat_train.copy()
reviews.head()
```

		id	name	asins	brand	
27895	AVpfl8cLLJeJML43AE3S		New Amazon Kindle Fire Hd 9w Powerfast Adapter...	B00L9EPT8O,B01E6AO69U	Amazon	5 C Ech
9252	AVphgVaX1cnluZ0-DR74		Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes...	B018Y229OU	Amazon	Tablets, Tat {
1955	AVqklhwDv8e3D1O-lebb		All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Ele Tablets
33173	AV1YE_muvKc47QAVgpwE		NaN	B00U3FPN4U	Amazon Fire Tv	Back To (Electroni
14385	AVphgVaX1cnluZ0-DR74		Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes...	B018Y229OU	Amazon	Tablets, Tat {

5 rows × 21 columns

```
len(reviews["name"].unique())
```

47

```
len(reviews["asins"].unique())
```

36

```
reviews.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 27701 entries, 27895 to 27018
Data columns (total 21 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    27674 non-null  object
1   name                  22303 non-null  object
2   asins                 27672 non-null  object
3   brand                 27674 non-null  object
4   categories            27674 non-null  object
5   keys                  27674 non-null  object
6   manufacturer          27674 non-null  object
7   reviews.date          27651 non-null  object
8   reviews.dateAdded     19237 non-null  object
9   reviews.dateSeen      27674 non-null  object
10  reviews.didPurchase   1 non-null     object
11  reviews.doRecommend  27239 non-null  object
12  reviews.id            1 non-null     float64
13  reviews.numHelpful    27291 non-null  float64
14  reviews.rating        27674 non-null  float64
15  reviews.sourceURLs    27674 non-null  object
16  reviews.text          27673 non-null  object
17  reviews.title         27671 non-null  object
18  reviews.userCity      0 non-null     float64
19  reviews.userProvince  0 non-null     float64
20  reviews.username      27673 non-null  object
dtypes: float64(5), object(16)
memory usage: 4.6+ MB
```

```
reviews.groupby("asins")["name"].unique()
```

```
asins
B005PB2T0S      [Kindle Keyboard,,,\r\nKindle Keyboard,,, Ama...
B005PB2T2Q      [Fire Kids Edition Tablet, 7 Display, Wi-Fi, 1...
B00DU15MU4      [Brand New Amazon Kindle Fire 16gb 7 Ips Displ...
B00IOY8XWQ      [Kindle Paperwhite E-reader - White, 6 High-Re...
B00IOYAM4I      [Brand New Amazon Kindle Fire 16gb 7 Ips Displ...
```

```

B00L9EPT80,B01E6A069U [New Amazon Kindle Fire Hd 9w Powerfast Adapte...
B00L029KXQ [Fire Tablet, 7 Display, Wi-Fi, 8 GB - Include...
B00LW9X0JM [nan]
B00OQVZDJM [Amazon Kindle Paperwhite - eBook reader - 4 G...
B00QFQRELG [nan]
B00QJDU3KY [Fire Tablet, 7 Display, Wi-Fi, 8 GB - Include...
B00QL1ZN3G [nan]
B00REQKWGA [Brand New Amazon Kindle Fire 16gb 7 Ips Displ...
B00TSUGXKE [Echo (White),,,\r\nEcho (White),,,]
B00U3FPN4U [nan]
B00UH4D8G2 [Echo (White),,,\r\nEcho (White),,,]
B00VINDBJK [Kindle Oasis E-reader with Leather Charging C...
B00X4WHP5E [nan]
B00ZV9PXP2 [Fire Tablet, 7 Display, Wi-Fi, 8 GB - Include...
B0189XY0Q [nan]
B018SZT3BK [Fire Tablet, 7 Display, Wi-Fi, 8 GB - Include...
B018T075DC [Brand New Amazon Kindle Fire 16gb 7 Ips Displ...
B018Y225IA [Brand New Amazon Kindle Fire 16gb 7 Ips Displ...
B018Y229OU [Fire Tablet, 7 Display, Wi-Fi, 8 GB - Include...
B018Y22BI4 [Echo (White),,,\r\nEcho (White),,,]
B018Y22C2Y [Fire Tablet, 7 Display, Wi-Fi, 8 GB - Include...
B018Y23MNM [Amazon Kindle Paperwhite - eBook reader - 4 G...
B018Y23P7K [nan]
B01AHB9C1E [Amazon 5W USB Official OEM Charger and Power ...
B01AHB9CN2 [All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi...
B01AHB9CYG [All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi...
B01BFIBRIE [Fire Tablet, 7 Display, Wi-Fi, 8 GB - Include...
B01BH830OM [nan]
B01J2G4VVG [Amazon 5W USB Official OEM Charger and Power ...
B01J40RNHU [Fire Tablet, 7 Display, Wi-Fi, 8 GB - Include...
Name: name, dtype: object

```

```

different_names = reviews[reviews["asins"] ==
                           "B00L9EPT80,B01E6A069U"]["name"].unique()
for name in different_names:
    print(name)

```

```

New Amazon Kindle Fire Hd 9w Powerfast Adapter Charger + Micro Usb Angle Cable,,,
New Amazon Kindle Fire Hd 9w Powerfast Adapter Charger + Micro Usb Angle Cable,,,
Amazon - Amazon Tap Portable Bluetooth and Wi-Fi Speaker - Black,,,
Amazon - Amazon Tap Portable Bluetooth and Wi-Fi Speaker - Black,,,
Amazon Fire Tv,,,
Amazon Fire Tv,,,
Echo (White),,,
Echo (White),,,
nan
Kindle Dx Leather Cover, Black (fits 9.7 Display, Latest and 2nd Generation Kindle Dxs),,
Amazon 9W PowerFast Official OEM USB Charger and Power Adapter for Fire Tablets and Kindle eReaders,,,
Amazon 9W PowerFast Official OEM USB Charger and Power Adapter for Fire Tablets and Kindle eReaders,,,
Amazon Fire Hd 10 Tablet, Wi-Fi, 16 Gb, Special Offers - Silver Aluminum,,,
Amazon Fire Hd 10 Tablet, Wi-Fi, 16 Gb, Special Offers - Silver Aluminum,,,
Amazon Kindle Fire 5ft USB to Micro-USB Cable (works with most Micro-USB Tablets),,,
Amazon Kindle Fire 5ft USB to Micro-USB Cable (works with most Micro-USB Tablets),,,
Echo (White),,,
Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes Special Offers, Tangerine"
Amazon Fire Hd 6 Standing Protective Case(4th Generation - 2014 Release), Cayenne Red,,,
Amazon Fire Hd 6 Standing Protective Case(4th Generation - 2014 Release), Cayenne Red,,,
Echo (Black),,,
Amazon 9W PowerFast Official OEM USB Charger and Power Adapter for Fire Tablets and Kindle eReaders,,,
Amazon 5W USB Official OEM Charger and Power Adapter for Fire Tablets and Kindle eReaders,,,
Amazon 5W USB Official OEM Charger and Power Adapter for Fire Tablets and Kindle eReaders,,,
Amazon Fire Hd 6 Standing Protective Case(4th Generation - 2014 Release), Cayenne Red,,,
Amazon 5W USB Official OEM Charger and Power Adapter for Fire Tablets and Kindle eReaders,,,
Amazon Fire Tv,,,
Kindle Dx Leather Cover, Black (fits 9.7 Display, Latest and 2nd Generation Kindle Dxs)",,
Echo (Black),,,
Echo (Black),,,

```

```
reviews[reviews["asins"] == "B00L9EPT80,B01E6A069U"]["name"].value_counts()
```

```

Echo (White),,,\r\nEcho (White),,,
2299
Amazon Fire Tv,,, \r\nAmazon Fire Tv,,,
2021
Amazon - Amazon Tap Portable Bluetooth and Wi-Fi Speaker - Black,,, \r\nAmazon - Amazon Tap Portable Bluetooth and Wi-Fi Speaker -
Black,,,
262
Amazon Fire Hd 10 Tablet, Wi-Fi, 16 Gb, Special Offers - Silver Aluminum,,, \r\nAmazon Fire Hd 10 Tablet, Wi-Fi, 16 Gb, Special
Offers - Silver Aluminum,,,
102
Amazon 9W PowerFast Official OEM USB Charger and Power Adapter for Fire Tablets and Kindle eReaders,,, \r\nAmazon 9W PowerFast
Official OEM USB Charger and Power Adapter for Fire Tablets and Kindle eReaders,,,
29
Kindle Dx Leather Cover, Black (fits 9.7 Display, Latest and 2nd Generation Kindle Dxs),,
8
Amazon Fire Hd 6 Standing Protective Case(4th Generation - 2014 Release), Cayenne Red,,, \r\nAmazon Fire Hd 6 Standing Protective
Case(4th Generation - 2014 Release), Cayenne Red,,,
6
Amazon 5W USB Official OEM Charger and Power Adapter for Fire Tablets and Kindle eReaders,,, \r\nAmazon 5W USB Official OEM Charger
and Power Adapter for Fire Tablets and Kindle eReaders,,,
6
New Amazon Kindle Fire Hd 9w Powerfast Adapter Charger + Micro Usb Angle Cable,,, \r\nNew Amazon Kindle Fire Hd 9w Powerfast Adapter

```

```

Charger + Micro Usb Angle Cable,,,
Amazon Kindle Fire 5ft USB to Micro-USB Cable (works with most Micro-USB Tablets),,,\r\nAmazon Kindle Fire 5ft USB to Micro-USB
Cable (works with most Micro-USB Tablets),,,
Echo (White),,,\r\nFire Tablet, 7 Display, Wi-Fi, 8 GB - Includes Special Offers, Tangerine"
1
Echo (Black),,,\r\nAmazon 9W PowerFast Official OEM USB Charger and Power Adapter for Fire Tablets and Kindle eReaders,,,
1
Amazon Fire Hd 6 Standing Protective Case(4th Generation - 2014 Release), Cayenne Red,,, \r\nAmazon 5W USB Official OEM Charger and
Power Adapter for Fire Tablets and Kindle eReaders,,,
1
Amazon Fire Tv,,, \r\nKindle Dx Leather Cover, Black (fits 9.7 Display, Latest and 2nd Generation Kindle Dxs)",,
1
Echo (Black),,, \r\nEcho (Black),,,
1
Name: name, dtype: int64

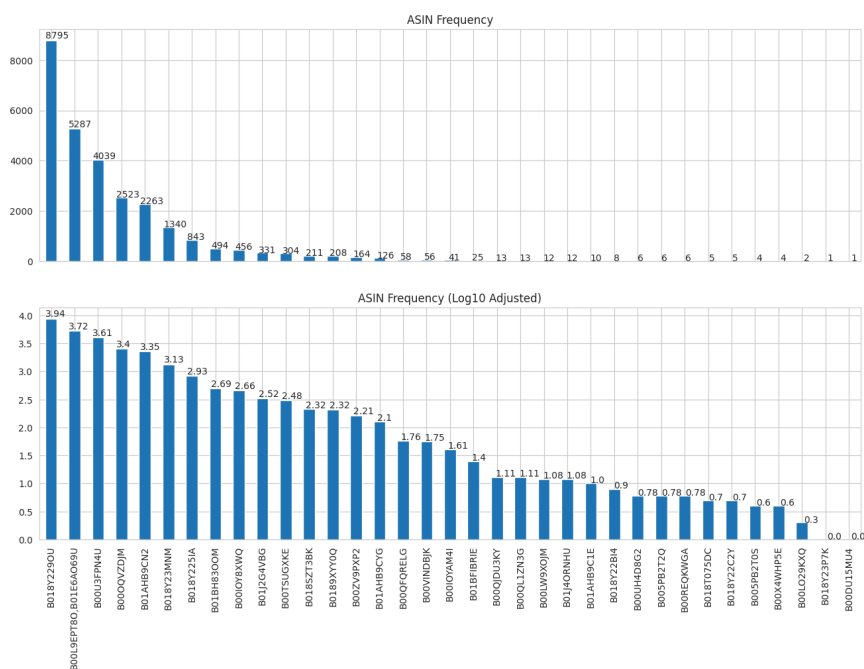
```

```

fig = plt.figure(figsize=(16,10))
ax1 = plt.subplot(211)
ax2 = plt.subplot(212, sharex = ax1)
reviews["asins"].value_counts().plot(kind="bar", ax=ax1, title="ASIN Frequency")
np.log10(reviews["asins"].value_counts()).plot(kind="bar", ax=ax2,
                                                title="ASIN Frequency (Log10 Adjusted)")

for p in ax1.patches:
    ax1.annotate(str(p.get_height()), (p.get_x() * 1.01, p.get_height() * 1.01))
for p in ax2.patches:
    ax2.annotate(str(round((p.get_height()),2)), (p.get_x() * 1.01, p.get_height() * 1.01))
plt.show()

```



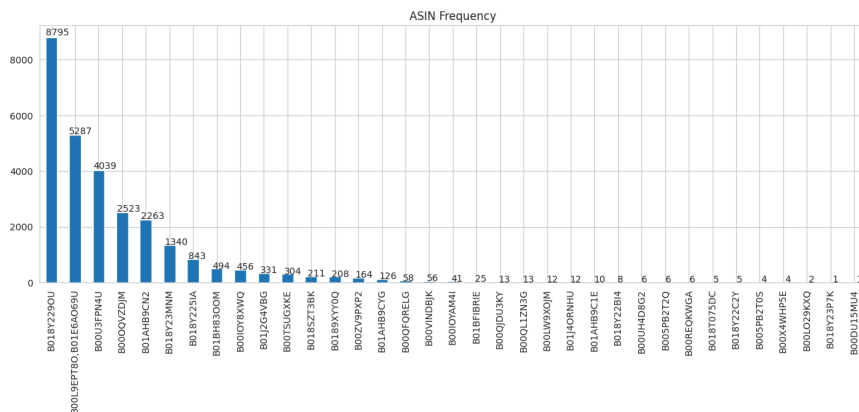
```
reviews["reviews.rating"].mean()
```

```
4.585278600852786
```

```

asins_count_ix = reviews["asins"].value_counts().index
fig = plt.figure(figsize=(16,5))
ax=reviews["asins"].value_counts().plot(kind="bar", title="ASIN Frequency")
for p in ax.patches:
    ax.annotate(str(p.get_height()), (p.get_x() * 1.01 , p.get_height() * 1.01))
plt.xticks(rotation=90)
plt.show()

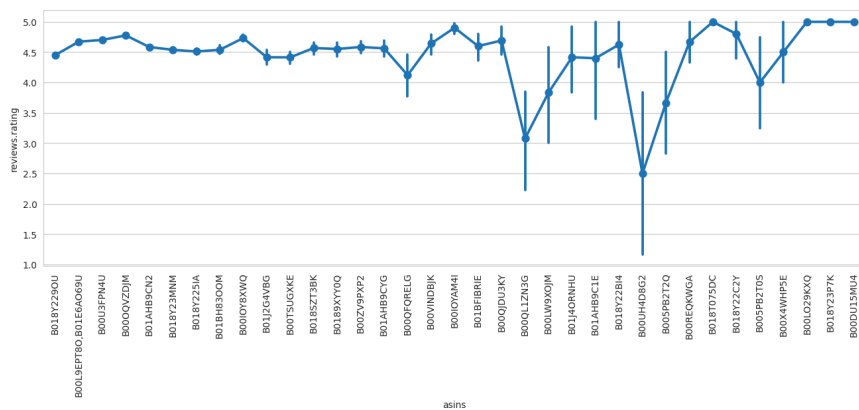
```



```

fig = plt.figure(figsize=(16,5))
sns.pointplot(x="asins", y="reviews.rating", order=asins_count_ix, data=reviews)
plt.xticks(rotation=90)
plt.show()

```



✓ Sentiment Analysis

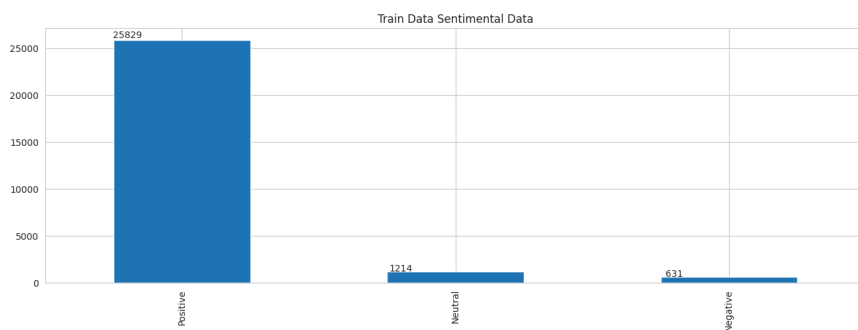
```
def sentiments(rating):
    if (rating == 5) or (rating == 4):
        return "Positive"
    elif rating == 3:
        return "Neutral"
    elif (rating == 2) or (rating == 1):
        return "Negative"
# Add sentiments to the data
strat_train["Sentiment"] = strat_train["reviews.rating"].apply(sentiments)
strat_test["Sentiment"] = strat_test["reviews.rating"].apply(sentiments)
print(strat_train["Sentiment"][:15])
```

```
27895    Positive
9252     Positive
1955     Positive
33173    Positive
14385    Positive
11337    Positive
25289    Positive
11280    Positive
9329     Positive
11394    Positive
27278    Positive
14660    Positive
23561    Positive
21529     Neutral
1087     Positive
Name: Sentiment, dtype: object
```

```
round((strat_train["Sentiment"].value_counts()*100/len(strat_train)),2)
```

```
Positive    93.24
Neutral      4.38
Negative     2.28
Name: Sentiment, dtype: float64
```

```
fig = plt.figure(figsize=(16,5))
ax=strat_train["Sentiment"].value_counts().plot(kind="bar", title="Train Data Sentimental Data")
for p in ax.patches:
    ax.annotate(str(p.get_height()), (p.get_x() * 1.01 , p.get_height() * 1.01))
plt.show()
```



```
round((strat_test["Sentiment"].value_counts()*100/len(strat_test)),2)
```

```
Positive    93.37
Neutral      4.09
Negative     2.45
Name: Sentiment, dtype: float64
```

```
fig = plt.figure(figsize=(16,5))
ax=strat_test["Sentiment"].value_counts().plot(kind="bar", title="Test Data Sentimental Data")
for p in ax.patches:
```

```
for p in ax.patches:  
    ax.annotate(str(p.get_height()), (p.get_x() * 1.01 , p.get_height() * 1.01))  
plt.show()
```

