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

In [3]:
```
df = pd.read_csv(r"C:\Users\AISWARYA\Downloads\amazon.csv")
```

```
C:\Users\AISWARYA\AppData\Local\Temp\ipykernel_14556\3193914938.py:1: DtypeWa
rning: Columns (1,10) have mixed types. Specify dtype option on import or set
low_memory=False.
  df = pd.read_csv(r"C:\Users\AISWARYA\Downloads\amazon.csv")
```

In [5]:
```
pip install --user -U nltk
```

```
Requirement already satisfied: nltk in c:\users\aiswarya\anaconda3\lib\site-p
ackages (3.9.1)
Requirement already satisfied: click in c:\users\aiswarya\anaconda3\lib\site-
packages (from nltk) (8.1.7)
Requirement already satisfied: joblib in c:\users\aiswarya\anaconda3\lib\site
-packages (from nltk) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in c:\users\aiswarya\anaconda3
\lib\site-packages (from nltk) (2024.9.11)
Requirement already satisfied: tqdm in c:\users\aiswarya\anaconda3\lib\site-p
ackages (from nltk) (4.66.5)
Requirement already satisfied: colorama in c:\users\aiswarya\anaconda3\lib\si
te-packages (from click->nltk) (0.4.6)
Note: you may need to restart the kernel to use updated packages.
```

In [7]:
```
df.head()
```

Out[7]:

	id	name	asins	brand	categories	
0	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazc
1	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazc
2	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazc
3	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazc
4	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazc

5 rows × 21 columns

In [9]:
```
df.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         34654 non-null  object 
 18  reviews.userCity      0 non-null      float64
 19  reviews.userProvince  0 non-null      float64
 20  reviews.username      34653 non-null  object 
dtypes: float64(5), object(16)
memory usage: 5.6+ MB
```

In [13]:
```
df.isnull().sum()
```

Out[13]:
```
id                       0
name                  6760
asins                    2
brand                    0
categories               0
keys                     0
manufacturer             0
reviews.date            39
reviews.dateAdded    10621
reviews.dateSeen         0
reviews.didPurchase  34659
reviews.doRecommend    594
reviews.id           34659
reviews.numHelpful     529
reviews.rating          33
reviews.sourceURLs       0
reviews.text             1
reviews.title            6
reviews.userCity     34660
reviews.userProvince 34660
reviews.username         7
dtype: int64
```

In [15]:
```
df.describe()
```

Out[15]:

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

In [17]:
```
data = df[["reviews.text", "reviews.rating"]]
data.head()
```

Out[17]:

	reviews.text	reviews.rating
0	This product so far has not disappointed. My c...	5.0
1	great for beginner or experienced person. Boug...	5.0
2	Inexpensive tablet for him to use and learn on...	5.0
3	I've had my Fire HD 8 two weeks now and I love...	4.0
4	I bought this for my grand daughter when she c...	5.0

In [19]:
```
df.isnull().sum()
```

Out[19]:
```
id                        0
name                   6760
asins                     2
brand                     0
categories                0
keys                      0
manufacturer              0
reviews.date             39
reviews.dateAdded      10621
reviews.dateSeen           0
reviews.didPurchase    34659
reviews.doRecommend      594
reviews.id             34659
reviews.numHelpful       529
reviews.rating            33
reviews.sourceURLs         0
reviews.text               1
reviews.title              6
reviews.userCity       34660
reviews.userProvince   34660
reviews.username           7
dtype: int64
```

In [21]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 34660 entries, 0 to 34659
Data columns (total 2 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   reviews.text    34659 non-null  object 
 1   reviews.rating  34627 non-null  float64
dtypes: float64(1), object(1)
memory usage: 541.7+ KB
```

In [23]:
```
#drop null values
data.dropna(inplace=True)
data.isnull().sum()
```

```
C:\Users\AISWARYA\AppData\Local\Temp\ipykernel_14556\4210551023.py:2: Setting
WithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy
  data.dropna(inplace=True)
```

Out[23]:
```
reviews.text      0
reviews.rating    0
dtype: int64
```

In [25]:
```
import random
n_samples = 5

for _ in range(n_samples):
    i = random.choice(range(data.shape[0]))
    print(f"REVIEW TEXT:\n{data['reviews.text'][i]} \n\nRATE:\n{data['review
    print('\n', 90*"-", '\n')
```

```
REVIEW TEXT:
This tablet is a great price. Perfect for those who use Amazon and want to us
e Amazon apps.Don't let the low price and screen size fool you. Works flawles
sly. It has long battery life and a bright display. This is a must-have item
for $50.

RATE:
4.0

 --------------------------------------------------------------------------
--------------

REVIEW TEXT:
Very helpful enjoy asking Alexa questions on history each day

RATE:
4.0

 --------------------------------------------------------------------------
--------------

REVIEW TEXT:
This is my first tablet. It is simple to use and very efficient.The only thin
g that I would have liked would be larger keys on the key pad. Other than tha
t I am very happy with it.

RATE:
4.0

 --------------------------------------------------------------------------
--------------

REVIEW TEXT:
This is a great tablet for the price and is easy for my kids to use they love
it

RATE:
4.0

 --------------------------------------------------------------------------
--------------

REVIEW TEXT:
I have a few top end Bluetooth speakers and this Amazon tap rates right up wi
th them in sound and built quality plus it has alexa built in

RATE:
5.0

 --------------------------------------------------------------------------
--------------
```

In [27]:
```
# descriptive statistics
data.describe()
```

Out[27]:

	reviews.rating
count	34626.000000
mean	4.584561
std	0.735660
min	1.000000
25%	4.000000
50%	5.000000
75%	5.000000
max	5.000000

In [29]:
```
# distribution of rating
data['reviews.rating'].value_counts().sort_index(ascending=False)
```

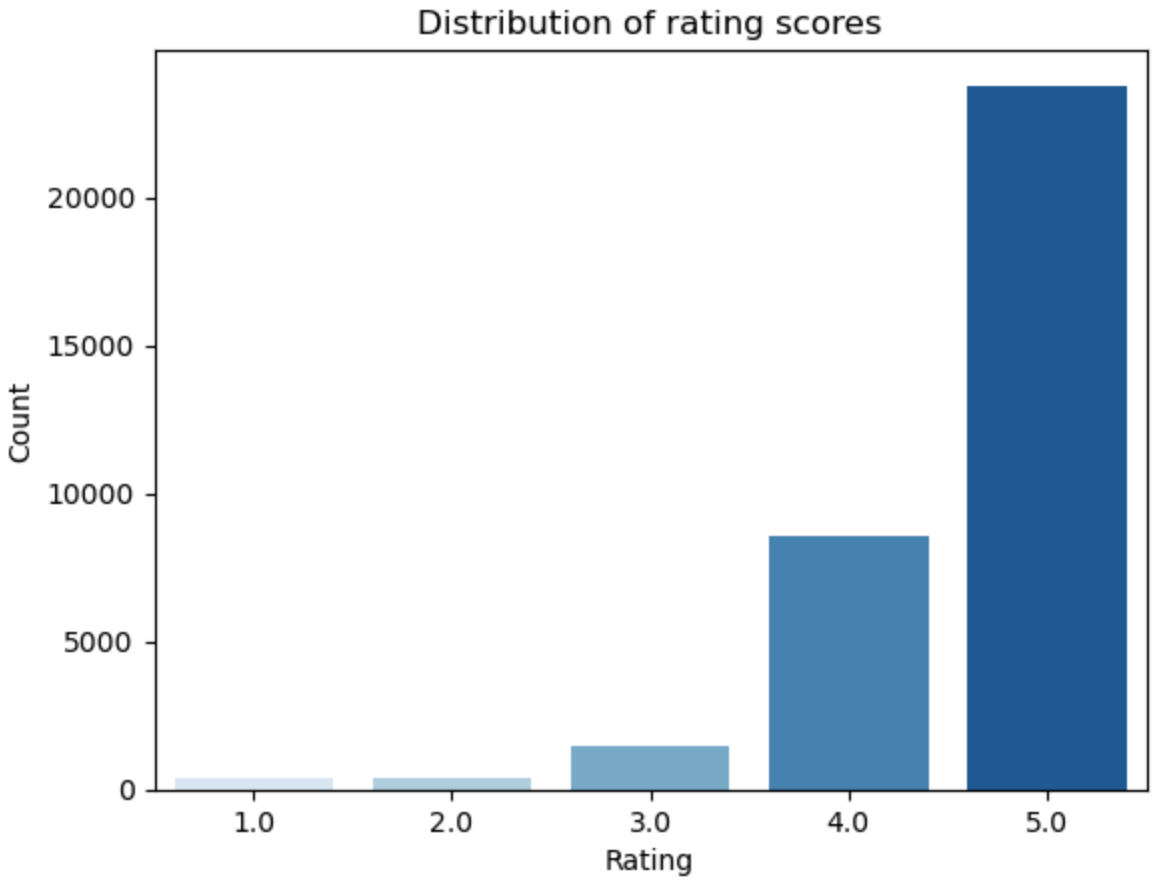
Out[29]:
```
reviews.rating
5.0    23774
4.0     8541
3.0     1499
2.0      402
1.0      410
Name: count, dtype: int64
```

In [31]:
```
rating = data["reviews.rating"].value_counts
rating
```

Out[31]:
```
<bound method IndexOpsMixin.value_counts of 0        5.0
1        5.0
2        5.0
3        4.0
4        5.0
        ...
34655    3.0
34656    1.0
34657    1.0
34658    1.0
34659    1.0
Name: reviews.rating, Length: 34626, dtype: float64>
```

In [33]:
```
# distribution of rating
sns.countplot(x = data['reviews.rating'], palette='Blues')

plt.title('Distribution of rating scores')
plt.xlabel('Rating')
plt.ylabel('Count')
plt.show()
```

```
C:\Users\AISWARYA\AppData\Local\Temp\ipykernel_14556\2890151192.py:2: FutureW
arning:

Passing `palette` without assigning `hue` is deprecated and will be removed i
n v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the sa
me effect.

  sns.countplot(x = data['reviews.rating'], palette='Blues')
```



In [51]:
```
# loading new data
data_2 = pd.read_csv(r"C:\Users\AISWARYA\Downloads\amazoncustomerproductsrev
data_2 = data_2[["reviews.text", "reviews.rating"]]

# using only lower ratings. less than or equals 3
data_2 = data_2[data_2["reviews.rating"] <= 3].reset_index(drop=True)

#now for 3rd Dataset as well

data_3 = pd.read_csv(r"C:\Users\AISWARYA\Downloads\amazoncustomer.csv")
data_3 = data_3[["reviews.text", "reviews.rating"]]
data_3 = data_3[data_3["reviews.rating"] <=3 ].reset_index(drop=True)
```

In [53]:
```
data_2["reviews.rating"].value_counts().sort_index(ascending=False)
```

Out[53]:
```
reviews.rating
3    1206
2     616
1     965
Name: count, dtype: int64
```

In [55]:
```
data_3["reviews.rating"].value_counts().sort_index(ascending=False)
```

Out[55]:
```
reviews.rating
3    197
2     54
1     63
Name: count, dtype: int64
```

In [57]:
```
#concatination
data = pd.concat([data, data_2, data_3])

data.head()
```

Out[57]:

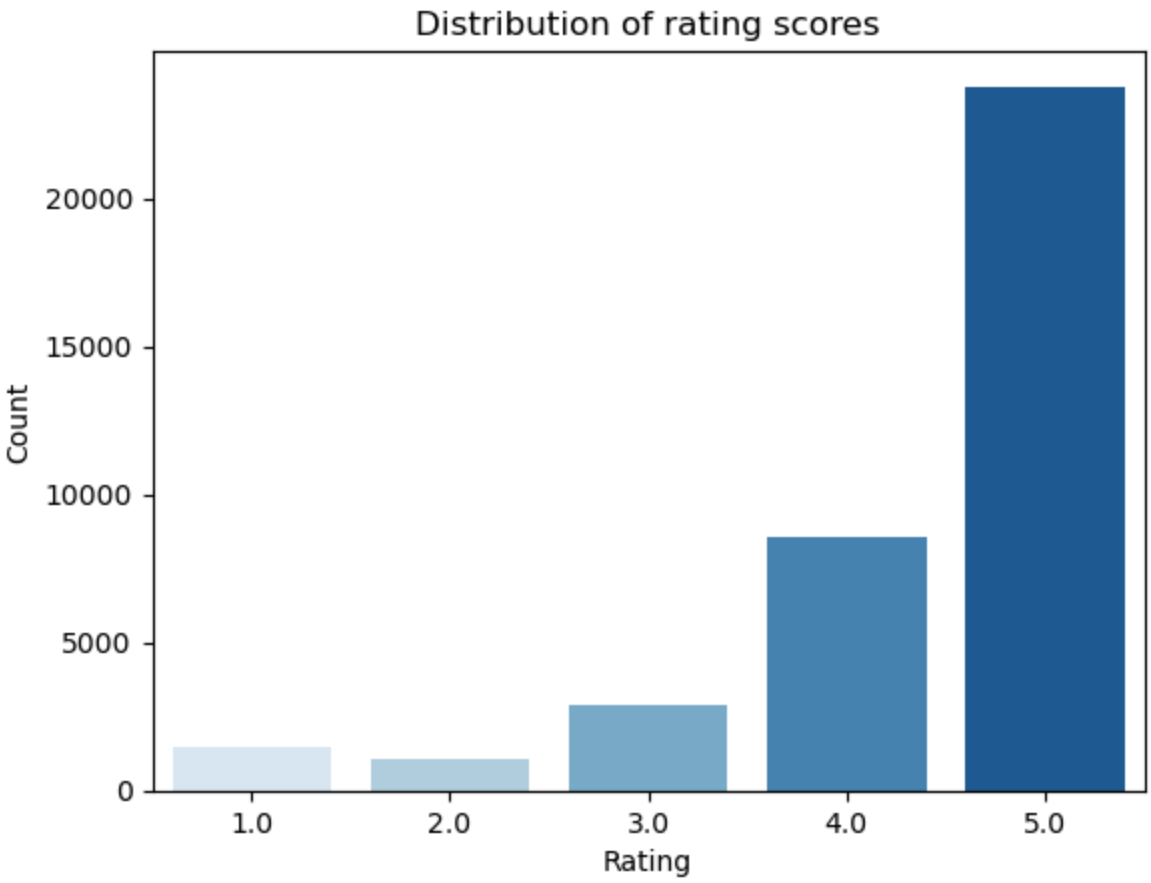
	reviews.text	reviews.rating
0	This product so far has not disappointed. My c...	5.0
1	great for beginner or experienced person. Boug...	5.0
2	Inexpensive tablet for him to use and learn on...	5.0
3	I've had my Fire HD 8 two weeks now and I love...	4.0
4	I bought this for my grand daughter when she c...	5.0

In [59]:
```
data["reviews.rating"].value_counts().sort_index(ascending=False)
```

Out[59]:
```
reviews.rating
5.0    23774
4.0     8541
3.0     2902
2.0     1072
1.0     1438
Name: count, dtype: int64
```

In [61]:
```
# distribution of rating
sns.countplot(x = data['reviews.rating'], palette='Blues')

plt.title('Distribution of rating scores')
plt.xlabel('Rating')
plt.ylabel('Count')
plt.show()
```

```
C:\Users\AISWARYA\AppData\Local\Temp\ipykernel_21076\2890151192.py:2: FutureW
arning:

Passing `palette` without assigning `hue` is deprecated and will be removed i
n v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the sa
me effect.

  sns.countplot(x = data['reviews.rating'], palette='Blues')
```



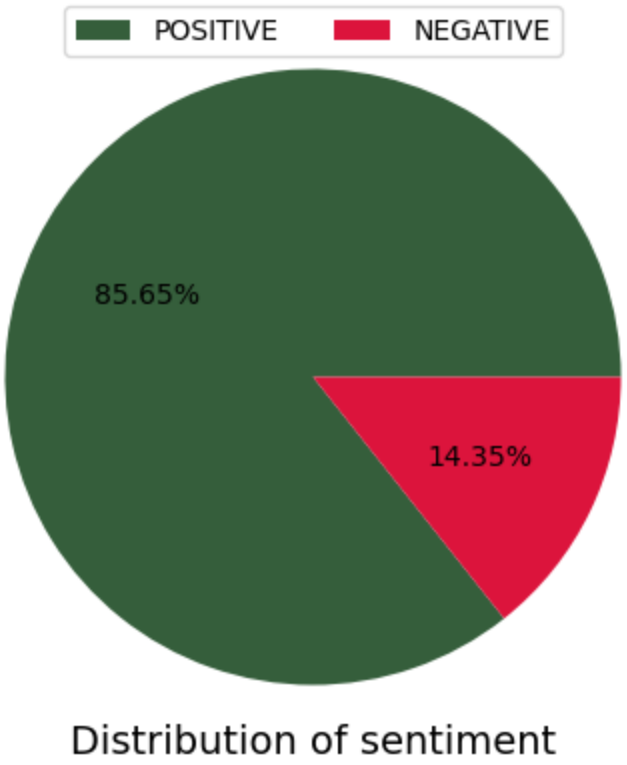
In [63]:
```
# we'll map ratings.

sentiment_score = {1: 0,
                   2: 0,
                   3: 0,
                   4: 1,
                   5: 1}
sentiment = {0: 'NEGATIVE',
             1: 'POSITIVE'}
data['sentiment_score'] = data ['reviews.rating'].map(sentiment_score)
data['sentiment'] = data['sentiment_score'].map(sentiment)
```

In [65]:
```
data.head()
```

Out[65]:

	reviews.text	reviews.rating	sentiment_score	sentiment
0	This product so far has not disappointed. My c...	5.0	1	POSITIVE
1	great for beginner or experienced person. Boug...	5.0	1	POSITIVE
2	Inexpensive tablet for him to use and learn on...	5.0	1	POSITIVE
3	I've had my Fire HD 8 two weeks now and I love...	4.0	1	POSITIVE
4	I bought this for my grand daughter when she c...	5.0	1	POSITIVE

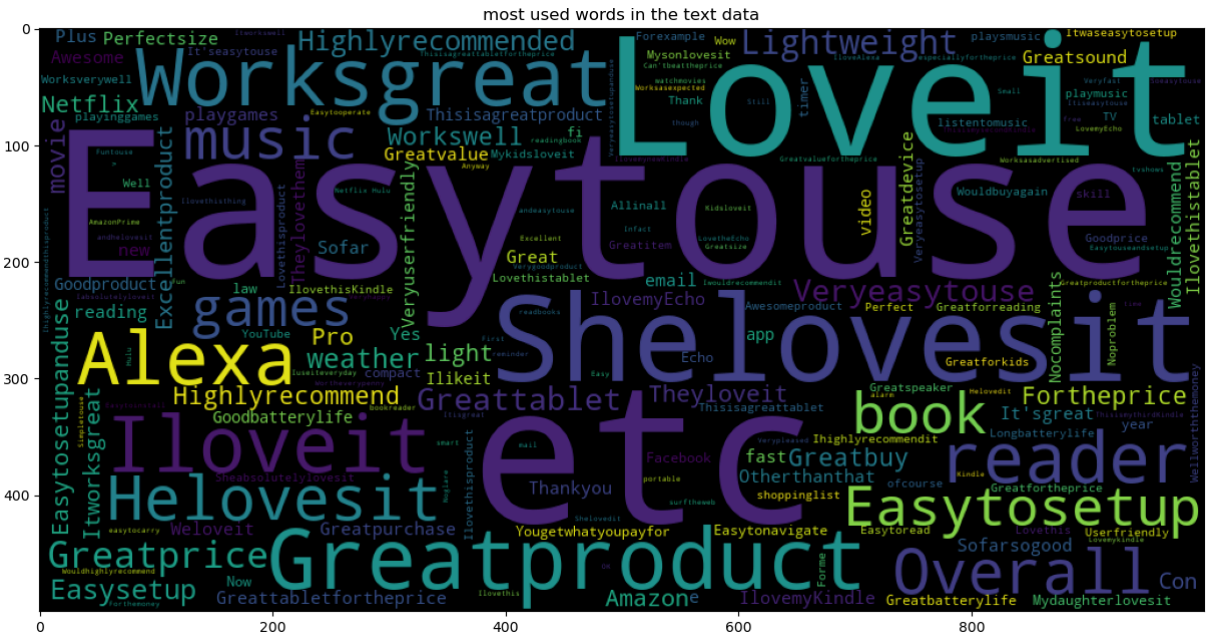
In [67]:
```
plt.figure(figsize = (5,5))

labels = ['POSITIVE', 'NEGATIVE']
colors = colors = ['#355E3B', '#DC143C']

plt.pie(data['sentiment'].value_counts(), autopct='%0.2f%%',colors=colors)

plt.title('Distribution of sentiment', size=14, y=-0.01)
plt.legend(labels, ncol=2, loc=9)
plt.show()
```



In [69]:
```
# getting all used words
all_words = pd.Series(' '.join(data['reviews.text']).split())
```

In [71]:
```
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
```

In [73]:
```
# plotting word cloud

wordcloud = WordCloud(width = 1000, height = 500).generate(''.join(all_words

plt.figure(figsize = (15,8) )
plt.imshow(wordcloud)
plt.title ('most used words in the text data')

plt.show()
```



In [45]:
```
data.dropna(inplace=True)

data.isnull().sum()
```

```
C:\Users\AISWARYA\AppData\Local\Temp\ipykernel_14556\446825523.py:1: SettingW
ithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy
  data.dropna(inplace=True)
```

Out[45]:
```
reviews.text       0
reviews.rating     0
sentiment_score    0
sentiment          0
dtype: int64
```

In [43]:
```
sentiment_score = {1: 0,

                   2: 0,

                   3: 0,
```

```
                    4: 1,

                    5: 1}


sentiment = {0: 'NEGATIVE',

             1: 'POSITIVE'}




# mapping

data['sentiment_score'] = data['reviews.rating'].map(sentiment_score)

data['sentiment'] = data['sentiment_score'].map(sentiment)


data.head()
```

```
C:\Users\AISWARYA\AppData\Local\Temp\ipykernel_14556\2508432333.py:23: Settin
gWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy
  data['sentiment_score'] = data['reviews.rating'].map(sentiment_score)
C:\Users\AISWARYA\AppData\Local\Temp\ipykernel_14556\2508432333.py:25: Settin
gWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy
  data['sentiment'] = data['sentiment_score'].map(sentiment)
```

Out[43]:

	reviews.text	reviews.rating	sentiment_score	sentiment
0	This product so far has not disappointed. My c...	5.0	1	POSITIVE
1	great for beginner or experienced person. Boug...	5.0	1	POSITIVE
2	Inexpensive tablet for him to use and learn on...	5.0	1	POSITIVE
3	I've had my Fire HD 8 two weeks now and I love...	4.0	1	POSITIVE
4	I bought this for my grand daughter when she c...	5.0	1	POSITIVE

In [135…
```
pip install --user -U nltk
```

```
Requirement already satisfied: nltk in c:\users\aiswarya\anaconda3\lib\site-p
ackages (3.9.1)
Requirement already satisfied: click in c:\users\aiswarya\anaconda3\lib\site-
packages (from nltk) (8.1.7)
Requirement already satisfied: joblib in c:\users\aiswarya\anaconda3\lib\site
-packages (from nltk) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in c:\users\aiswarya\anaconda3
\lib\site-packages (from nltk) (2024.9.11)
Requirement already satisfied: tqdm in c:\users\aiswarya\anaconda3\lib\site-p
ackages (from nltk) (4.66.5)
Requirement already satisfied: colorama in c:\users\aiswarya\anaconda3\lib\si
te-packages (from click->nltk) (0.4.6)
Note: you may need to restart the kernel to use updated packages.
```

In [139…
```
data=pd.read_csv(r"C:\Users\AISWARYA\Downloads\amazon.csv")

data
```

```
C:\Users\AISWARYA\AppData\Local\Temp\ipykernel_14556\4188227231.py:1: DtypeWa
rning: Columns (1,10) have mixed types. Specify dtype option on import or set
low_memory=False.
  data=pd.read_csv(r"C:\Users\AISWARYA\Downloads\amazon.csv")
```

Out[139...

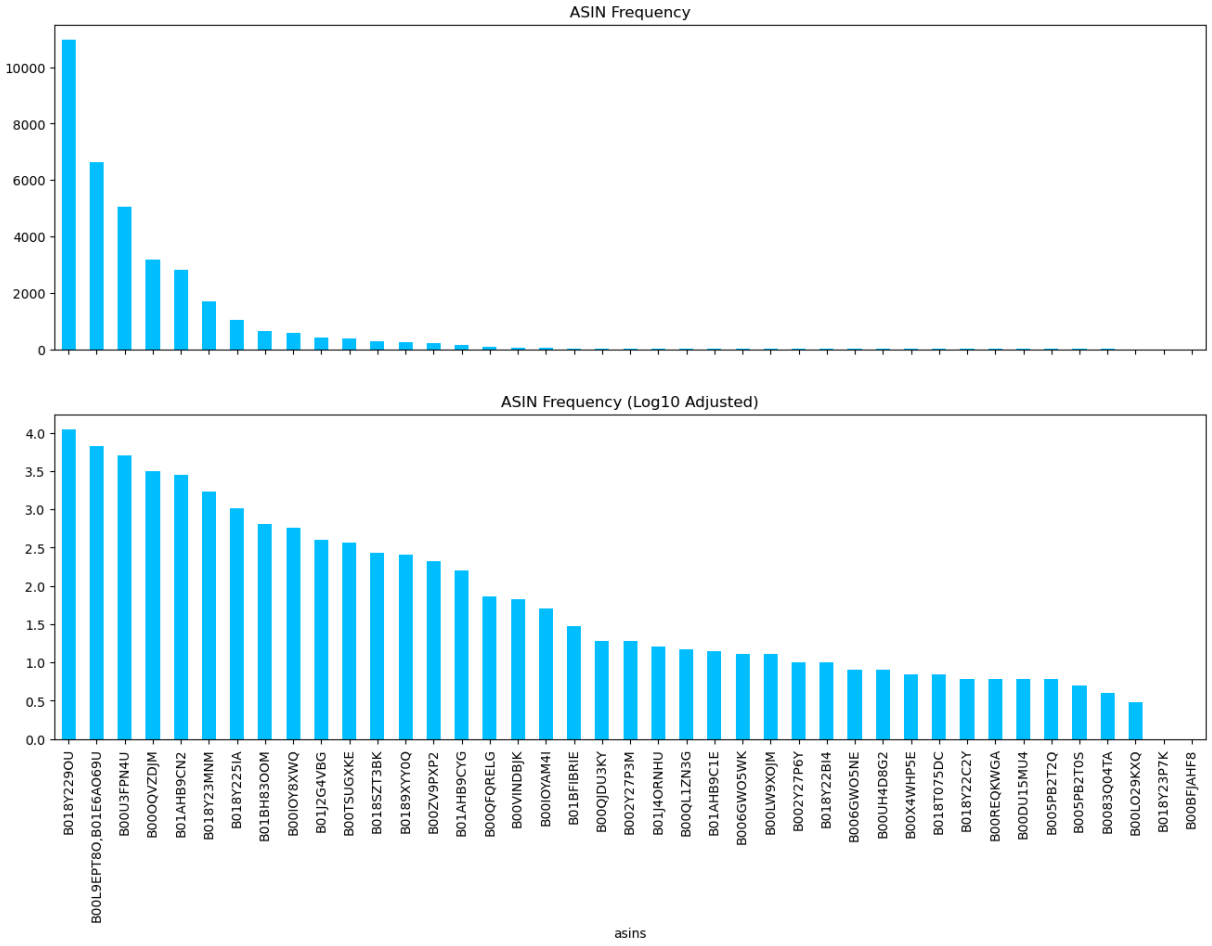
	id	name	asins	brand	categories	
0	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	8416671046
1	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	8416671046
2	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	8416671046
3	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	8416671046
4	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	8416671046
...	...	...	...	...	...	
34655	AVpfiBlyLJeJML43-4Tp	NaN	B006GWO5WK	Amazon	Computers/Tablets & Networking,Tablet & eBook ...	newama
34656	AVpfiBlyLJeJML43-4Tp	NaN	B006GWO5WK	Amazon	Computers/Tablets & Networking,Tablet & eBook ...	newama
34657	AVpfiBlyLJeJML43-4Tp	NaN	B006GWO5WK	Amazon	Computers/Tablets &	newama

	id	name	asins	brand	categories	
					Networking,Tablet & eBook ...	
34658	AVpfiBlyLJeJML43-4Tp	NaN	B006GWO5WK	Amazon	Computers/Tablets & Networking,Tablet & eBook ...	newama
34659	AVpfiBlyLJeJML43-4Tp	NaN	B006GWO5WK	Amazon	Computers/Tablets & Networking,Tablet & eBook ...	newama

34660 rows × 21 columns

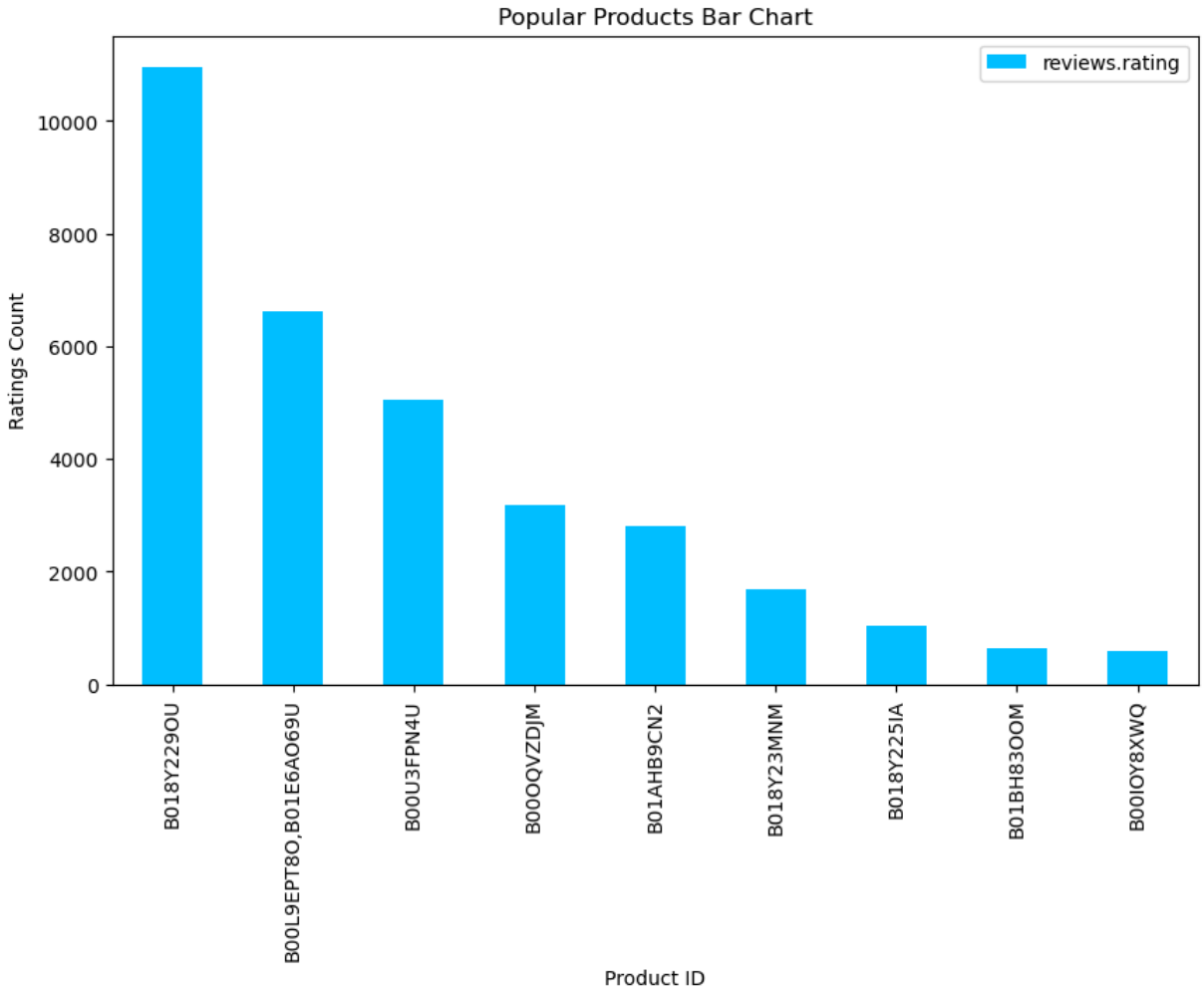
In [141…

```
colors = ['#00BFFF']

fig = plt.figure(figsize=(16,10))

ax1 = plt.subplot(211)

ax2 = plt.subplot(212, sharex = ax1)

data["asins"].value_counts().plot(kind="bar", ax=ax1, title="ASIN Frequency"

np.log10(data["asins"].value_counts()).plot(kind="bar", ax=ax2,
                                            title="ASIN Frequency (Log10



plt.savefig('ASIN_Frequency.png')

plt.show()
```



In [143…

```
popular_data = data.groupby("asins").filter(lambda x:x['reviews.rating'].cou

popular_products = pd.DataFrame(popular_data.groupby('asins')['reviews.ratin

popular_products = popular_products.sort_values('reviews.rating', ascending=


colors = ['#00BFFF', '#E6E6FA', '#8A2BE2', '#FFD700', '#DAA520', '#808080',


ax = popular_products.plot(kind = "bar", figsize = (10, 6), color = colors)


ax.set_xlabel("Product ID")

ax.set_ylabel("Ratings Count")

ax.set_title("Popular Products Bar Chart")


plt.savefig('popular_products_bar_chart.png')
```



In [160…
```
data = data[['reviews.text', 'reviews.title', 'reviews.rating', 'reviews.doR

data.head()
```

Out[160…

	reviews.text	reviews.title	reviews.rating	reviews.doRecommend
0	This product so far has not disappointed. My c...	Kindle	5.0	True
1	great for beginner or experienced person. Boug...	very fast	5.0	True
2	Inexpensive tablet for him to use and learn on...	Beginner tablet for our 9 year old son.	5.0	True
3	I've had my Fire HD 8 two weeks now and I love...	Good!!!	4.0	True
4	I bought this for my grand daughter when she c...	Fantastic Tablet for kids	5.0	True

In [162…
```
data.isnull().sum()
```

Out[162…

```
reviews.text              1
reviews.title             6
reviews.rating           33
reviews.doRecommend     594
dtype: int64
```

In [164…

```
rows_to_drop = []


for index, row in data.iterrows():
    if pd.isnull(row['reviews.doRecommend']):
        if row['reviews.rating'] > 3:
            data.at[index, 'reviews.doRecommend'] = True
        elif row['reviews.rating'] < 3:
            data.at[index, 'reviews.doRecommend'] = False
        elif row['reviews.rating'] == 3:
            rows_to_drop.append(index)
    if pd.isnull(row['reviews.doRecommend']) and pd.isnull(row['reviews.rati
        rows_to_drop.append(index)


data.drop(rows_to_drop, inplace=True)


data
```

```
C:\Users\AISWARYA\AppData\Local\Temp\ipykernel_14556\1522960176.py:27: Settin
gWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy
  data.drop(rows_to_drop, inplace=True)
```

Out[164…

	reviews.text	reviews.title	reviews.rating	reviews.doRecommend
0	This product so far has not disappointed. My c...	Kindle	5.0	True
1	great for beginner or experienced person. Boug...	very fast	5.0	True
2	Inexpensive tablet for him to use and learn on...	Beginner tablet for our 9 year old son.	5.0	True
3	I've had my Fire HD 8 two weeks now and I love...	Good!!!	4.0	True
4	I bought this for my grand daughter when she c...	Fantastic Tablet for kids	5.0	True
...	...	...	...	...
34654	This is exactly like any other usb power charg...	Not Necessary	1.0	False
34656	Amazon should include this charger with the Ki...	Should be included	1.0	False
34657	Love my Kindle Fire but I am really disappoint...	Disappointing Charger	1.0	False
34658	I was surprised to find it did not come with a...	Not worth the money	1.0	False
34659	to spite the fact that i have nothing but good...	as with everyone else	1.0	False

34598 rows × 4 columns

In [166…
```
nltk.download('stopwords')

nltk.download('punkt')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\AISWARYA\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\AISWARYA\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```

Out[166… `True`

In [178…
```
y = data['reviews.rating'].values - 1

y
```

Out[178… `array([4., 4., 4., ..., 0., 0., 0.])`

In [182…
```
xgb = XGBClassifier(learning_rate = 0.2 , n_estimators = 200)
```

In [196…
```
title = input("Enter your review title: ")

text = input("Enter your review: ")

recommend = input("Do you recommend this product? (yes/no): ").strip().lower


recommend = 1 if recommend == "yes" else 0


combined_input = f"{title} {text}"

l1 = []

clean = re.sub('[^a-zA-Z]',' ',combined_input)

clean = clean.lower()

clean = word_tokenize(clean)

clean = [ps.stem(word) for word in clean if word not in stop_words ]

clean = ' '.join(clean)

new_review_text = cv.transform([clean])

new_review_recommend = np.array([recommend]).reshape(1, -1)

new_review = hstack([new_review_text, new_review_recommend])

res = naive.predict(new_review)


print(f"Predicted Review Rating: {int(res[0] + 1)}")
```

```
---------------------------------------------------------------------------
LookupError                               Traceback (most recent call last)
Cell In[196], line 21
     17 clean = re.sub('[^a-zA-Z]',' ',combined_input)
     19 clean = clean.lower()
---> 21 clean = word_tokenize(clean)
     23 clean = [ps.stem(word) for word in clean if word not in stop_words ]
     25 clean = ' '.join(clean)

File ~\anaconda3\Lib\site-packages\nltk\tokenize\__init__.py:142, in word_tok
enize(text, language, preserve_line)
    127 def word_tokenize(text, language="english", preserve_line=False):
    128     """
    129     Return a tokenized copy of *text*,
    130     using NLTK's recommended word tokenizer
   (...)
    140     :type preserve_line: bool
    141     """
--> 142     sentences = [text] if preserve_line else sent_tokenize(text, lang
uage)
    143     return [
    144         token for sent in sentences for token in _treebank_word_token
izer.tokenize(sent)
    145     ]

File ~\anaconda3\Lib\site-packages\nltk\tokenize\__init__.py:119, in sent_tok
enize(text, language)
    109 def sent_tokenize(text, language="english"):
    110     """
    111     Return a sentence-tokenized copy of *text*,
    112     using NLTK's recommended sentence tokenizer
   (...)
    117     :param language: the model name in the Punkt corpus
    118     """
--> 119     tokenizer = _get_punkt_tokenizer(language)
    120     return tokenizer.tokenize(text)

File ~\anaconda3\Lib\site-packages\nltk\tokenize\__init__.py:105, in _get_pun
kt_tokenizer(language)
     96 @functools.lru_cache
     97 def _get_punkt_tokenizer(language="english"):
     98     """
     99     A constructor for the PunktTokenizer that utilizes
    100     a lru cache for performance.
   (...)
    103     :type language: str
    104     """
--> 105     return PunktTokenizer(language)

File ~\anaconda3\Lib\site-packages\nltk\tokenize\punkt.py:1744, in PunktToken
izer.__init__(self, lang)
   1742 def __init__(self, lang="english"):
   1743     PunktSentenceTokenizer.__init__(self)
-> 1744     self.load_lang(lang)

File ~\anaconda3\Lib\site-packages\nltk\tokenize\punkt.py:1749, in PunktToken
```

```
izer.load_lang(self, lang)
   1746 def load_lang(self, lang="english"):
   1747     from nltk.data import find
-> 1749     lang_dir = find(f"tokenizers/punkt_tab/{lang}/")
   1750     self._params = load_punkt_params(lang_dir)
   1751     self._lang = lang

File ~\anaconda3\Lib\site-packages\nltk\data.py:579, in find(resource_name, p
aths)
    577 sep = "*" * 70
    578 resource_not_found = f"\n{sep}\n{msg}\n{sep}\n"
--> 579 raise LookupError(resource_not_found)

LookupError:
**********************************************************************
  Resource punkt_tab not found.
  Please use the NLTK Downloader to obtain the resource:

  >>> import nltk
  >>> nltk.download('punkt_tab')

  For more information see: https://www.nltk.org/data.html

  Attempted to load tokenizers/punkt_tab/english/

  Searched in:
    - 'C:\\Users\\AISWARYA/nltk_data'
    - 'C:\\Users\\AISWARYA\\anaconda3\\nltk_data'
    - 'C:\\Users\\AISWARYA\\anaconda3\\share\\nltk_data'
    - 'C:\\Users\\AISWARYA\\anaconda3\\lib\\nltk_data'
    - 'C:\\Users\\AISWARYA\\AppData\\Roaming\\nltk_data'
    - 'C:\\nltk_data'
    - 'D:\\nltk_data'
    - 'E:\\nltk_data'
**********************************************************************
```

In [ ]: