AMAZON MARKET ANALYSIS - EXTERNAL ANALYSIS

With the help of this real time analysis we would like to make the analysis of amazon. This analysis will help us to understand the positive and as well as the negative reviews of amazon, this will help in listing down few aspects in the matrix evaluation.

In [1]:

```
import pandas as pd
```

In [2]:

```
import numpy as np
```

In [10]:

```
import seaborn as sns
```

In [42]:

```
!pip install spacy
```

```
Requirement already satisfied: spacy in c:\users\prathyu lachireddy\anaconda
3\lib\site-packages (3.2.1)
Requirement already satisfied: typer<0.5.0,>=0.3.0 in c:\users\prathyu lachi
reddy\anaconda3\lib\site-packages (from spacy) (0.4.0)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in c:\users\prathyu l
achireddy\anaconda3\lib\site-packages (from spacy) (3.3.0)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in c:\users\prathyu l
achireddy\anaconda3\lib\site-packages (from spacy) (2.0.6)
Requirement already satisfied: pathy>=0.3.5 in c:\users\prathyu lachireddy\a
naconda3\lib\site-packages (from spacy) (0.6.1)
Requirement already satisfied: numpy>=1.15.0 in c:\users\prathyu lachireddy
\anaconda3\lib\site-packages (from spacy) (1.19.2)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<1.9.0,>=1.7.4 in c:\us
ers\prathyu lachireddy\anaconda3\lib\site-packages (from spacy) (1.8.2)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in c:\users\prathyu
lachireddy\anaconda3\lib\site-packages (from spacy) (1.0.6)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in c:\users\prath
yu lachireddy\anaconda3\lib\site-packages (from spacy) (1.0.1)
Requirement already satisfied: setuptools in c:\users\prathyu lachireddy\ana
conda3\lib\site-packages (from spacy) (50.3.1.post20201107)
Requirement already satisfied: jinja2 in c:\users\prathyu lachireddy\anacond
a3\lib\site-packages (from spacy) (2.11.2)
Requirement already satisfied: blis<0.8.0,>=0.4.0 in c:\users\prathyu lachir
eddy\anaconda3\lib\site-packages (from spacy) (0.7.5)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in c:\users\prathyu lachi
reddy\anaconda3\lib\site-packages (from spacy) (2.0.6)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in c:\users\prathyu lachi
reddy\anaconda3\lib\site-packages (from spacy) (4.50.2)
Requirement already satisfied: thinc<8.1.0,>=8.0.12 in c:\users\prathyu lach
ireddy\anaconda3\lib\site-packages (from spacy) (8.0.13)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in c:\users\prathyu lac
hireddy\anaconda3\lib\site-packages (from spacy) (3.0.6)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in c:\users\prathyu l
achireddy\anaconda3\lib\site-packages (from spacy) (2.24.0)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.8 in c:\users\prathy
u lachireddy\anaconda3\lib\site-packages (from spacy) (3.0.8)
Requirement already satisfied: srsly<3.0.0,>=2.4.1 in c:\users\prathyu lachi
reddy\anaconda3\lib\site-packages (from spacy) (2.4.2)
Requirement already satisfied: packaging>=20.0 in c:\users\prathyu lachiredd
y\anaconda3\lib\site-packages (from spacy) (20.4)
Requirement already satisfied: wasabi<1.1.0,>=0.8.1 in c:\users\prathyu lach
ireddy\anaconda3\lib\site-packages (from spacy) (0.9.0)
Requirement already satisfied: click<9.0.0,>=7.1.1 in c:\users\prathyu lachi
reddy\anaconda3\lib\site-packages (from typer<0.5.0,>=0.3.0->spacy) (7.1.2)
Requirement already satisfied: smart-open<6.0.0,>=5.0.0 in c:\users\prathyu
lachireddy\anaconda3\lib\site-packages (from pathy>=0.3.5->spacy) (5.2.1)
Requirement already satisfied: typing-extensions>=3.7.4.3 in c:\users\prathy
u lachireddy\anaconda3\lib\site-packages (from pydantic!=1.8,!=1.8.1,<1.9.0,
>=1.7.4->spacy) (3.7.4.3)
Requirement already satisfied: MarkupSafe>=0.23 in c:\users\prathyu lachired
dy\anaconda3\lib\site-packages (from jinja2->spacy) (1.1.1)
Requirement already satisfied: chardet<4,>=3.0.2 in c:\users\prathyu lachire
ddy\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (3.0.
4)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in
c:\users\prathyu lachireddy\anaconda3\lib\site-packages (from requests<3.0.
0,>=2.13.0->spacy) (1.25.11)
```

```
Requirement already satisfied: idna<3,>=2.5 in c:\users\prathyu lachireddy\a
naconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (2.10)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\prathyu lachir
eddy\anaconda3\lib\site-packages (from requests<3.0.0,>=2.13.0->spacy) (202
0.6.20)
Requirement already satisfied: six in c:\users\prathyu lachireddy\anaconda3
\lib\site-packages (from packaging>=20.0->spacy) (1.15.0)
Requirement already satisfied: pyparsing>=2.0.2 in c:\users\prathyu lachired
dy\anaconda3\lib\site-packages (from packaging>=20.0->spacy) (2.4.7)
```

In [25]:

```
!pip install wordcloud
```

```
Collecting wordcloud
  Downloading wordcloud-1.8.1-cp38-cp38-win_amd64.whl (155 kB)
Requirement already satisfied: numpy>=1.6.1 in c:\users\prathyu lachireddy\a
naconda3\lib\site-packages (from wordcloud) (1.19.2)
Requirement already satisfied: pillow in c:\users\prathyu lachireddy\anacond
a3\lib\site-packages (from wordcloud) (8.0.1)
Requirement already satisfied: matplotlib in c:\users\prathyu lachireddy\ana
conda3\lib\site-packages (from wordcloud) (3.3.2)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.3 in
c:\users\prathyu lachireddy\anaconda3\lib\site-packages (from matplotlib->wo
rdcloud) (2.4.7)
Requirement already satisfied: certifi>=2020.06.20 in c:\users\prathyu lachi
reddy\anaconda3\lib\site-packages (from matplotlib->wordcloud) (2020.6.20)
Requirement already satisfied: python-dateutil>=2.1 in c:\users\prathyu lach
ireddy\anaconda3\lib\site-packages (from matplotlib->wordcloud) (2.8.1)
Requirement already satisfied: cycler>=0.10 in c:\users\prathyu lachireddy\a
naconda3\lib\site-packages (from matplotlib->wordcloud) (0.10.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\prathyu lachire
ddy\anaconda3\lib\site-packages (from matplotlib->wordcloud) (1.3.0)
Requirement already satisfied: six>=1.5 in c:\users\prathyu lachireddy\anaco
nda3\lib\site-packages (from python-dateutil>=2.1->matplotlib->wordcloud)
(1.15.0)
Installing collected packages: wordcloud
Successfully installed wordcloud-1.8.1
```

In [27]:

```
import wordcloud
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
```

In [6]:

```
amazon= pd.read_csv("E://Important//CIA 3//SM//cleaned_data.csv")
amazon
```

Out[6]:

	reviews.text	reviews.rating	sentiment_score	sentiment	text
0	This product so far has not disappointed. My c...	5.0	1	POSITIVE	product so far have not disappoint my child lo...
1	great for beginner or experienced person. Boug...	5.0	1	POSITIVE	great for beginner or experienced person buy a...
2	Inexpensive tablet for him to use and learn on...	5.0	1	POSITIVE	inexpensive tablet for him use learn on step u...
3	I've had my Fire HD 8 two weeks now and I love...	4.0	1	POSITIVE	i ve have my fire hd two week now i love table...
4	I bought this for my grand daughter when she c...	5.0	1	POSITIVE	i buy for my grand daughter when she come over...
...	...	...	...	...	...
37722	The battery is having more and more trouble ho...	2.0	0	NEGATIVE	battery have more more trouble hold charge i b...
37723	My daughter has had this tablet for almost 2 m...	3.0	0	NEGATIVE	my daughter have have tablet for almost month ...
37724	Very cheap and was not impressed at all never ...	1.0	0	NEGATIVE	very cheap be not impress at all never again
37725	Hard to use, Lots of ads, and Randomly closes ...	2.0	0	NEGATIVE	hard use lot of ad randomly close apps
37726	I wish it has some more of the apps from the p...	3.0	0	NEGATIVE	i wish have some more of apps from play store ...

37727 rows × 5 columns

In [9]:

```
amazon.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 37727 entries, 0 to 37726
Data columns (total 5 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   reviews.text     37727 non-null  object 
 1   reviews.rating   37727 non-null  float64
 2   sentiment_score  37727 non-null  int64  
 3   sentiment        37727 non-null  object 
 4   text             37726 non-null  object 
dtypes: float64(1), int64(1), object(3)
memory usage: 1.4+ MB
```

In [37]:

```
amazon.describe()
```

Out[37]:

	reviews.rating	sentiment_score
count	37727.000000	37727.000000
mean	4.382061	0.856548
std	1.008880	0.350537
min	1.000000	0.000000
25%	4.000000	1.000000
50%	5.000000	1.000000
75%	5.000000	1.000000
max	5.000000	1.000000

In [40]:

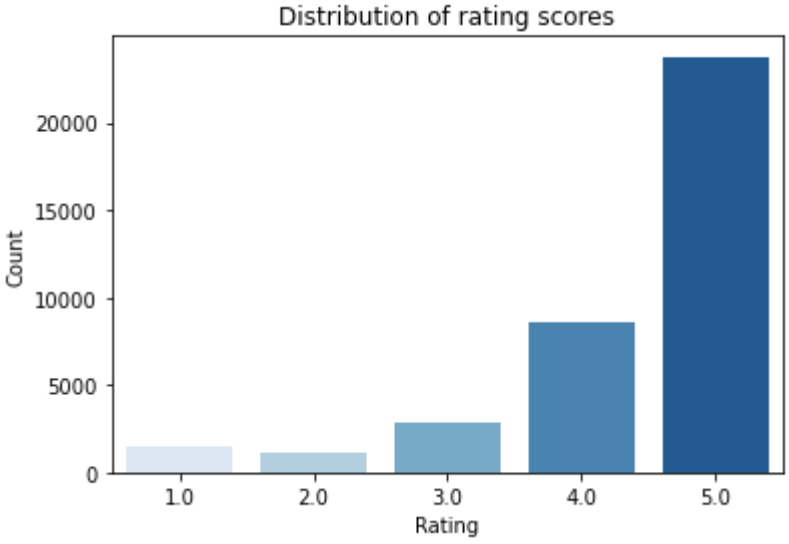
```
# explore some of the reviews
import random

n_samples = 5

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

```
REVIEW TEXT:
I like the kindle it is light and easy too sue battery life is good.

RATE:
3.0

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

REVIEW TEXT:
Very easy to set up with an Amazon prime account. Haven't had to pay anythin
g, great for small kids. Like the protective cover that goes on this device

RATE:
5.0

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

REVIEW TEXT:
If you enjoy reading you will not find a better device. Excellent price for
a product of this quality

RATE:
4.0

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

REVIEW TEXT:
This was an upgrade from an older model. It works better in daylight.

RATE:
4.0

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

REVIEW TEXT:
The apps I really want it won't let me download. I don't like that I HAVE to
sign up for "One Click" with a credit card.

RATE:
3.0

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

In [15]:

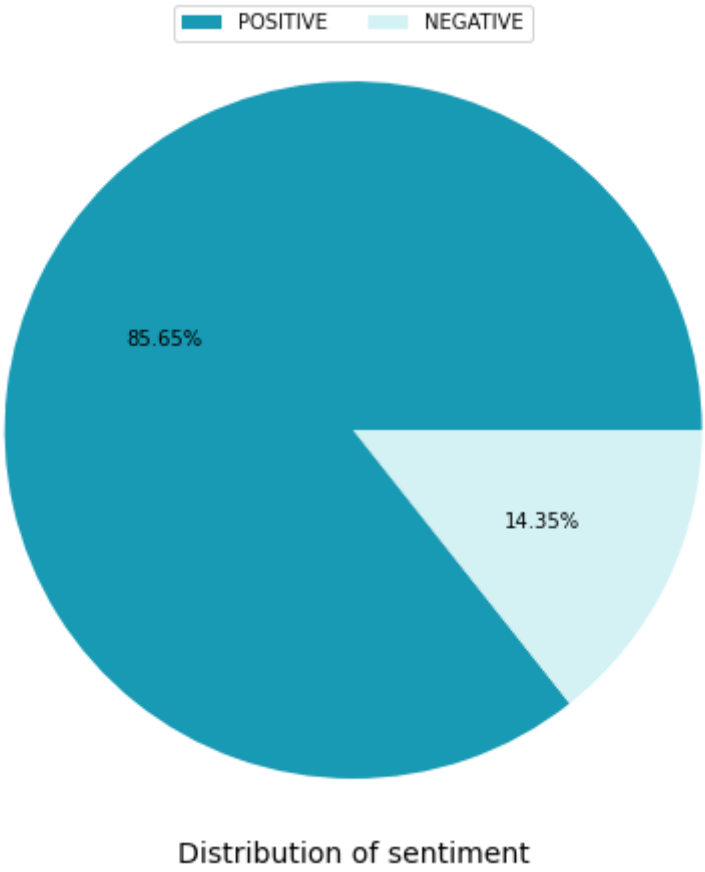
```
import matplotlib.pyplot as plt
# distribution of rating
sns.countplot(amazon['reviews.rating'], palette='Blues')

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

```
C:\Users\Prathyu Lachireddy\anaconda3\lib\site-packages\seaborn\_decorators.
py:36: FutureWarning: Pass the following variable as a keyword arg: x. From
version 0.12, the only valid positional argument will be `data`, and passing
other arguments without an explicit keyword will result in an error or misin
terpretation.
  warnings.warn(
```



In [17]:

```
# distribution of sentiment
plt.figure(figsize = (8, 8))

labels = ['POSITIVE', 'NEGATIVE']
colors = ['#189AB4', '#D4F1F4']
plt.pie(amazon['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()
```

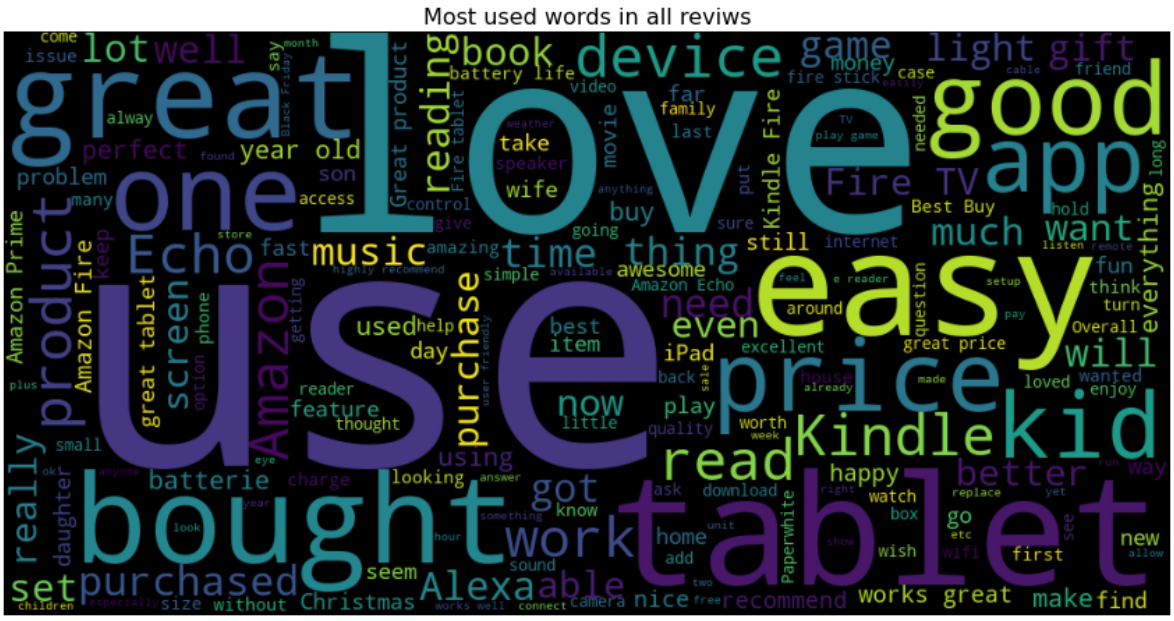


In [20]:

```
# get all used words
all_words = pd.Series(' '.join(amazon['reviews.text']).split())
```

In [28]:

```
# plot 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 all reviws", size=16)

plt.axis("off")
plt.show()
```



POSITIVE REVIEWS OF AMAZON

In [32]:

```
# get words used positive reivews
positiveWords = pd.Series(' '.join(amazon[amazon['sentiment']=='POSITIVE']['reviews.text'])
```

In [33]:

```
# plot word cloud
wordcloud = WordCloud(width = 1000, height = 500).generate(' '.join(positiveWords))

plt.figure(figsize=(15,8))

plt.imshow(wordcloud)
plt.title("Most used words in positive reviews", size=16)

plt.axis("off")
plt.show()
```

NEGATIVE REVIEWS OF AMAZON

In [35]:

```
# get words used negative reivews
negativeWords = pd.Series(' '.join(amazon[amazon['sentiment']=='NEGATIVE']['reviews.text'])
```

In [36]:

```
# plot word cloud
wordcloud = WordCloud(width = 1000, height = 500).generate(' '.join(negativeWords))

plt.figure(figsize=(15,8))

plt.imshow(wordcloud)
plt.title("Most used words in negative reviews", size=16)

plt.axis("off")
plt.show()
```

In [ ]:

```
```