```
!pip install textblob
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: textblob in /usr/local/lib/python3.10/dist-packages (0.17.1)
Requirement already satisfied: nltk>=3.1 in /usr/local/lib/python3.10/dist-packages (from textblob) (3.8.1)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages (from nltk>=3.1->textblob) (8.1.3)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from nltk>=3.1->textblob) (1.2.0)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.10/dist-packages (from nltk>=3.1->textblob) (2022.10.31)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from nltk>=3.1->textblob) (4.65.0)
```

```
import pandas as pd
import numpy as np
from textblob import TextBlob
from wordcloud import WordCloud
import re
import matplotlib.pyplot as plt
import seaborn as sns


import nltk
from nltk.stem.snowball import SnowballStemmer
from nltk.corpus import stopwords
import spacy
nlp = spacy.load("en_core_web_sm")
```

```
df=pd.read_csv('flipkart.csv')
```

```
df
```

	Unnamed: 0	Product_name	Review	Rating
0	0	Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600...	Best under 60k Great performancel got it for a...	5
1	1	Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600...	Good perfomence...	5
2	2	Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600...	Great performance but usually it has also that...	5
3	3	DELL Inspiron Athlon Dual Core 3050U - (4 GB/2...	My wife is so happy and best product 🤞 😘	5
4	4	DELL Inspiron Athlon Dual Core 3050U - (4 GB/2...	Light weight laptop with new amazing features,...	5
...	...	...	...	...
2299	2299	MSI 27 inch Full HD IPS Panel Monitor (PRO MP2...	Great display, accurate colours at this price ...	5
2300	2300	MSI 27 inch Full HD IPS Panel Monitor (PRO MP2...	Superb monitor first brought 1 used for 2 mont...	5
2301	2301	MSI 27 inch Full HD IPS Panel Monitor (PRO MP2...	Awesome	5
2302	2302	MSI 27 inch Full HD IPS Panel Monitor (PRO MP2...	Only one issue with adapter	5

```
df=df[['Review', 'Rating']]
df
```

	Review	Rating
0	Best under 60k Great performancel got it for a...	5
1	Good perfomence...	5
2	Great performance but usually it has also that...	5
3	My wife is so happy and best product 🤞 😘	5
4	Light weight laptop with new amazing features,...	5
...	...	...
2299	Great display, accurate colours at this price ...	5
2300	Superb monitor first brought 1 used for 2 mont...	5
2301	Awesome	5
2302	Only one issue with adapter	5
2303	Worth the money u spend for this monitor Great...	5

2304 rows × 2 columns

```
def cleanReviews(text):
```

```
    text = re.sub('@[A-Za-z0-9_]+', '', text)
    text = re.sub('#','',text)
    text = re.sub('https?:\/\/\S+', '', text)
    text = re.sub('\n',' ',text)
    text = re.sub(r'www\S+', " ", text)
    text = re.sub(r'\.|/|:|-', " ", text)
    text = re.sub(r'[^\w\s]','',text)
    return text
df['cleanedReviews'] = df['Review'].apply(cleanReviews)
df.head()
```

```
    <ipython-input-7-2b1c8fe99d07>:10: SettingWithCopyWarning:
    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/stable/user_guide/indexing.html#ret
      df['cleanedReviews'] = df['Review'].apply(cleanReviews)
```
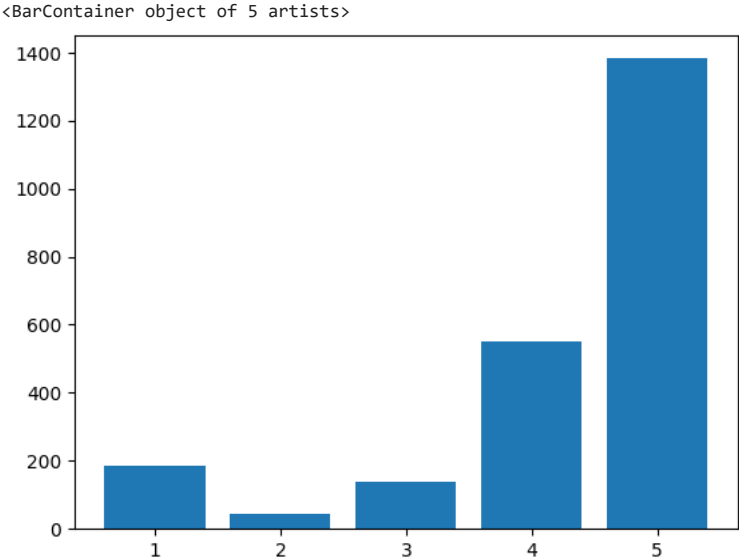
	Review	Rating	cleanedReviews
0	Best under 60k Great performancel got it for a...	5	Best under 60k Great performancel got it for a...
1	Good perfomence...	5	Good perfomence
2	Great performance but usually it has also that...	5	Great performance but usually it has also that...
3	My wife is so happy and best product ✌️ 😘	5	My wife is so happy and best product
4	Light weight laptop with new amazing features,...	5	Light weight laptop with new amazing features ...

```
df1=df[['cleanedReviews','Rating']]
df1
```

	cleanedReviews	Rating
0	Best under 60k Great performancel got it for a...	5
1	Good perfomence	5
2	Great performance but usually it has also that...	5
3	My wife is so happy and best product	5
4	Light weight laptop with new amazing features ...	5
...	...	...
2299	Great display accurate colours at this price r...	5
2300	Superb monitor first brought 1 used for 2 mont...	5
2301	Awesome	5
2302	Only one issue with adapter	5
2303	Worth the money u spend for this monitor Great...	5

2304 rows × 2 columns

```
x=[5,4,1,3,2]
plt.bar(x,df['Rating'].value_counts())
```

```
    <BarContainer object of 5 artists>
```



```
def getAnalysis(rating):
```

```python
    if rating<3:
        return 'Negative'
    elif rating ==3:
        return 'Neutral'
    else:
        return 'Positive'


df1['Analysis'] = df1['Rating'].apply(getAnalysis)

df1
```

	cleanedReviews	Rating	Analysis
0	Best under 60k Great performanceI got it for a...	5	Positive
1	Good perfomence	5	Positive
2	Great performance but usually it has also that...	5	Positive
3	My wife is so happy and best product	5	Positive
4	Light weight laptop with new amazing features ...	5	Positive
...	...	...	...
2299	Great display accurate colours at this price r...	5	Positive
2300	Superb monitor first brought 1 used for 2 mont...	5	Positive
2301	Awesome	5	Positive
2302	Only one issue with adapter	5	Positive
2303	Worth the money u spend for this monitor Great...	5	Positive

2304 rows × 3 columns

```python
def getSubjectivity(text):
    return TextBlob(text).sentiment.subjectivity


def getPolarity(text):
    return TextBlob(text).sentiment.polarity
df1['Subjectivity'] = df1['cleanedReviews'].apply(getSubjectivity)
df1['Polarity'] = df1['cleanedReviews'].apply(getPolarity)

df1
```

	cleanedReviews	Rating	Analysis	Subjectivity	Polarity
0	Best under 60k Great performanceI got it for a...	5	Positive	0.472424	0.438788
1	Good perfomence	5	Positive	0.600000	0.700000
2	Great performance but usually it has also that...	5	Positive	0.666667	0.183333
3	My wife is so happy and best product	5	Positive	0.650000	0.900000
4	Light weight laptop with new amazing features ...	5	Positive	0.763636	0.534091
...	...	...	...	...	...
2299	Great display accurate colours at this price r...	5	Positive	0.691667	0.600000
2300	Superb monitor first brought 1 used for 2 mont...	5	Positive	0.666667	0.625000
2301	Awesome	5	Positive	1.000000	1.000000
2302	Only one issue with adapter	5	Positive	1.000000	0.000000
2303	Worth the money u spend for this monitor Great...	5	Positive	0.425000	0.550000

2304 rows × 5 columns

```python
def getAnalysis(score):
    if score<0:
        return 'Negative'
    elif score ==0:
        return 'Neutral'
    else:
        return 'Positive'


df1['Analysis2'] = df1['Polarity'].apply(getAnalysis)

df1
```

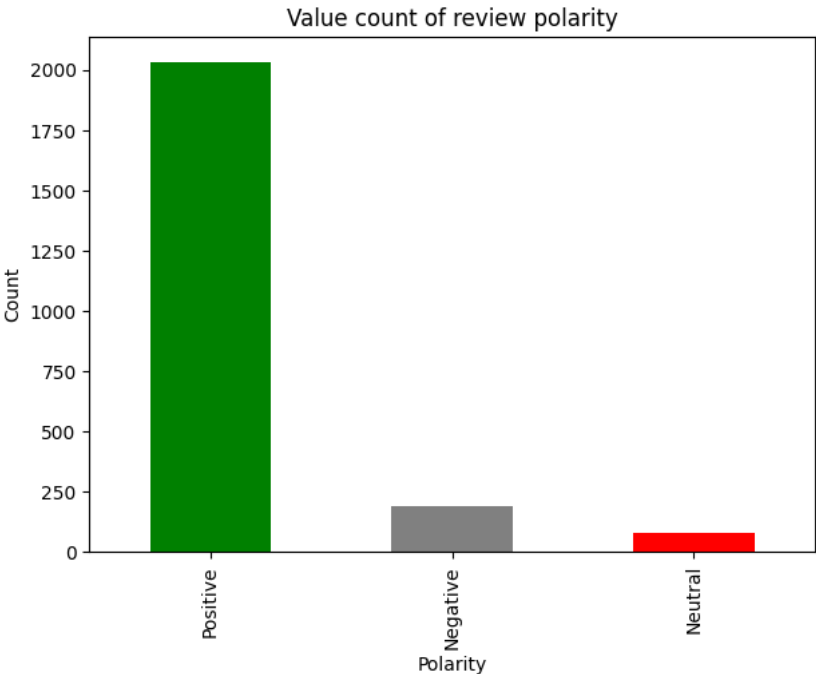|	cleanedReviews	Rating	Analysis	Subjectivity	Polarity	Analysis2
0	Best under 60k Great performancel got it for a...	5	Positive	0.472424	0.438788	Positive
1	Good perfomence	5	Positive	0.600000	0.700000	Positive
2	Great performance but usually it has also that...	5	Positive	0.666667	0.183333	Positive
3	My wife is so happy and best product	5	Positive	0.650000	0.900000	Positive
4	Light weight laptop with new amazing features ...	5	Positive	0.763636	0.534091	Positive
...	...	...	...	...	...	...
2299	Great display accurate colours at this price r...	5	Positive	0.691667	0.600000	Positive
2300	Superb monitor first brought 1 used for 2 mont...	5	Positive	0.666667	0.625000	Positive
2301	Awesome	5	Positive	1.000000	1.000000	Positive
2302	Only one issue with adapter	5	Positive	1.000000	0.000000	Neutral
2303	Worth the money u spend for this monitor Great...	5	Positive	0.425000	0.550000	Positive

```
df1['Analysis'].value_counts()
```

```
Positive    1934
Negative     230
Neutral      140
Name: Analysis, dtype: int64
```
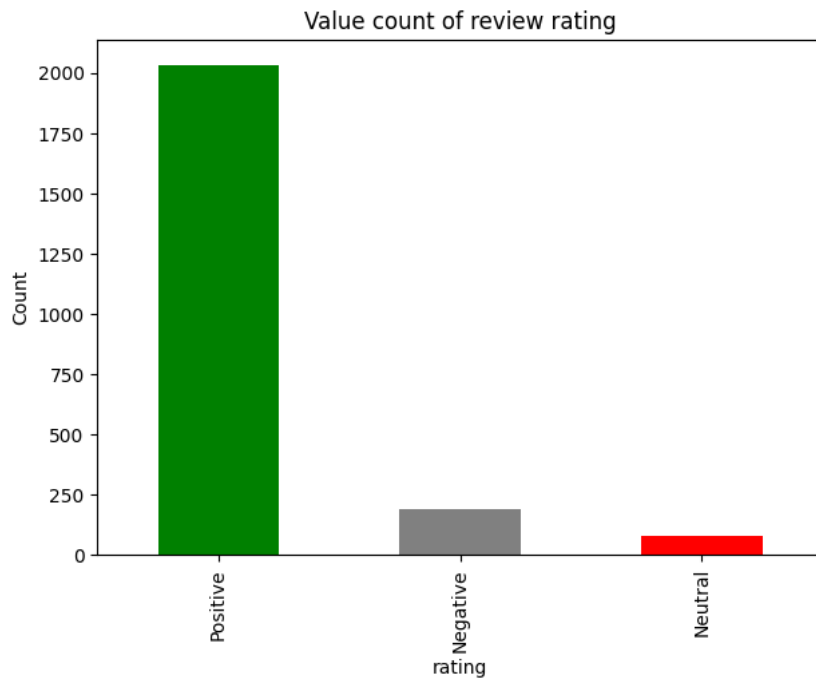
```
df1['Analysis2'].value_counts()
```

```
Positive    2035
Negative     190
Neutral       79
Name: Analysis2, dtype: int64
```

```
fig = plt.figure(figsize=(7,5))
color = ['green','grey','red']
df1['Analysis2'].value_counts().plot(kind='bar',color = color)
plt.title('Value count of review polarity')
plt.ylabel('Count')
plt.xlabel('Polarity')
plt.grid(False)
plt.show()
```



```
fig = plt.figure(figsize=(7,5))
color = ['green','grey','red']
df1['Analysis2'].value_counts().plot(kind='bar',color = color)
plt.title('Value count of review rating')
plt.ylabel('Count')
plt.xlabel('rating')
plt.grid(False)
plt.show()
```

```python
fig = plt.figure(figsize=(7,5))
color = ['green','grey','red']
df1['Analysis2'].value_counts().plot(kind='bar',color = color)
plt.title('Value count of review rating')
plt.ylabel('Count')
plt.xlabel('rating')
plt.grid(False)
plt.show()
```
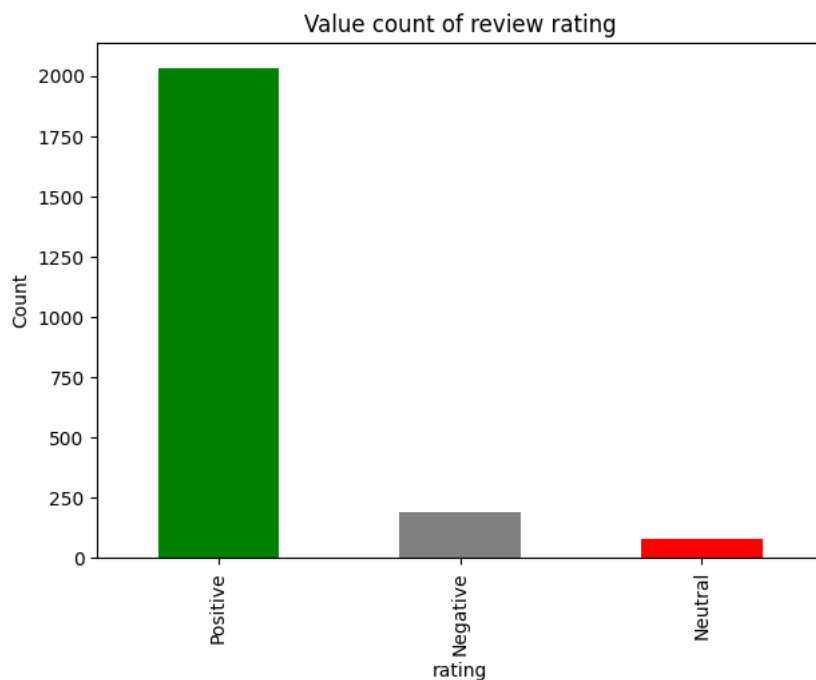


```python
fig = plt.figure(figsize=(7,7))
colors = ('green', 'grey', 'red')
wp={'linewidth':2, 'edgecolor': 'black'}
tags=df1['Analysis'].value_counts()
explode = (0.1,0.1,0.1)
tags.plot(kind='pie', autopct='%1.1f%%', shadow=True, colors=colors,
          startangle=90, wedgeprops=wp, explode=explode, label='')
plt.title('Distribution of ratings')
```
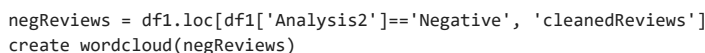
```
Text(0.5, 1.0, 'Distribution of ratings')
```

Distribution of ratings
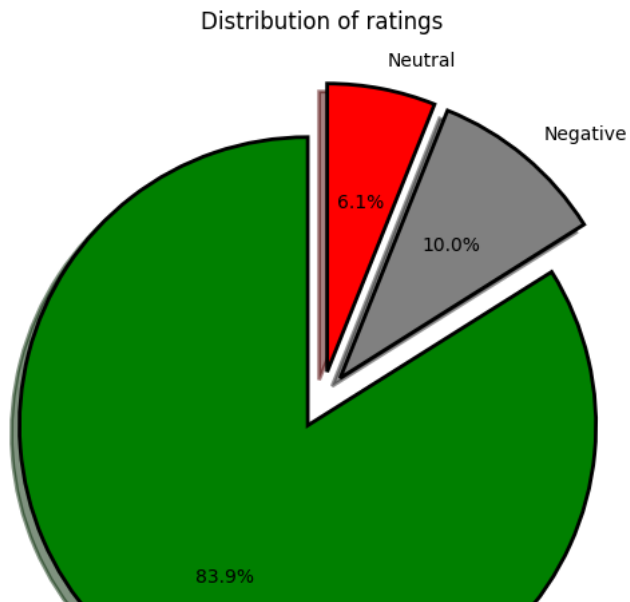

```python
plt.figure(figsize=(5,5))
for i in range(0,df1.shape[0]):
    plt.scatter(df1['Polarity'][i],df1['Subjectivity'][i], color='blue')
plt.title('Sentiment Analysis')
plt.xlabel('Polarity')
plt.ylabel('Subjectivity')
plt.show()
```
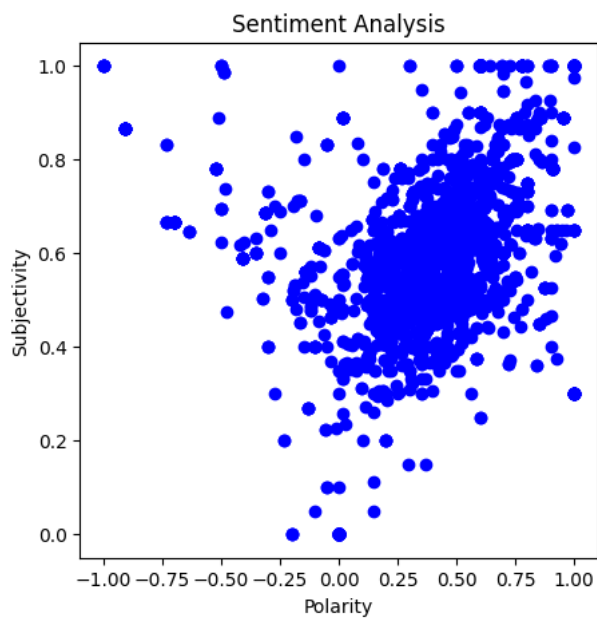


```python
def create_wordcloud(text):
    allWords = ' '.join([x for x in text])
    wordCloud = WordCloud(background_color='white', width=800, height=500, random_state=21, max_font_size=130).generate(allWords)
    plt.figure(figsize=(20,10))
    plt.imshow(wordCloud)
    plt.axis('off')
    plt.show()


posReviews = df1.loc[df1['Analysis2']=='Positive', 'cleanedReviews']
create_wordcloud(posReviews)
```

```
negReviews = df1.loc[df1['Analysis2']=='Negative', 'cleanedReviews']
create_wordcloud(negReviews)
```

✓ 3s   completed at 12:10 PM   ● ✕