```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import nltk
nltk.download('stopwords')

from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
from nltk.stem.porter import PorterStemmer

import string
import re
import textblob
from textblob import TextBlob
import os

from wordcloud import WordCloud, STOPWORDS
from wordcloud import ImageColorGenerator
import warnings
%matplotlib inline
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
```

```
#Read the JSON generated from the CLI command above and create a pandas dataframe
df = pd.read_csv(r'/content/flipkart.csv')
```

```
df.head(5)
```

	Unnamed: 0	Product_name	Review	Rating
0	0	Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600...	Best under 60k Great performanceI 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

```
df.to_csv()
```

```
',Unnamed: 0,Product_name,Review,Rating\n0,0,"Lenovo Ideapad Gaming 3 Ryzen 5
Hexa Core 5600H - (8 GB/512 GB SSD/Windows 11 Home/4 GB Graphics/NVIDIA GeFor
ce GTX 1650/120 Hz) 15ACH6 Gaming Laptop\xa0\xa0(15.6 inch, Shadow Black, 2.2
5$$kg kg)",Best under 60k Great performanceI got it for around 58500Battery b
ackup is bit low but thanks to rapid charger its very fast Display is Ok for
this price range Decent speakers with many customisation optionsVantage softw
are is so good for customisationOverall good performance till nowWill update
later if any problem occurs,5\n1,1,"Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core
5600H - (8 GB/512 GB SSD/Windows 11 Home/4 GB Graphics/NVIDIA GeForce GTX 165
```

```
df.shape
```

```
(2304, 4)
```

```
df.info
```

```
<bound method DataFrame.info of       Unnamed: 0                                       Product_name  \
0              0  Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600...
1              1  Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600...
2              2  Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600...
3              3  DELL Inspiron Athlon Dual Core 3050U - (4 GB/2...
4              4  DELL Inspiron Athlon Dual Core 3050U - (4 GB/2...
...          ...                                                ...
2299        2299  MSI 27 inch Full HD IPS Panel Monitor (PRO MP2...
2300        2300  MSI 27 inch Full HD IPS Panel Monitor (PRO MP2...
2301        2301  MSI 27 inch Full HD IPS Panel Monitor (PRO MP2...
2302        2302  MSI 27 inch Full HD IPS Panel Monitor (PRO MP2...
2303        2303  MSI 27 inch Full HD IPS Panel Monitor (PRO MP2...

                                             Review  Rating
0     Best under 60k Great performanceI 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 x 4 columns]>
```

```
df.value_counts()
```

```
Unnamed: 0  Product_name
Review
Rating
0           Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600H - (8 GB/512 GB SSD/Windows 11 Home/4 GB Graphics/NVIDIA
GeForce GTX 1650/120 Hz) 15ACH6 Gaming Laptop  (15.6 inch, Shadow Black, 2.25$$kg kg)  Best under 60k Great
performanceI got it for around 58500Battery backup is bit low but thanks to rapid charger its very fast Display is Ok
for this price range Decent speakers with many customisation optionsVantage software is so good for
customisationOverall good performance till nowWill update later if any problem occurs
5         1
1           Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600H - (8 GB/512 GB SSD/Windows 11 Home/4 GB Graphics/NVIDIA
GeForce GTX 1650/120 Hz) 15ACH6 Gaming Laptop  (15.6 inch, Shadow Black, 2.25$$kg kg)  Good perfomence...
5         1
1532        realme C25_Y (Glacier Blue, 128 GB)  (4 GB RAM)
My review After using 3 days Camera- 2/5Battery- 5/5 because of fast chargerDISPLAY-3/5 because no FHD+
DisplayPerformanc-4/5Back panel and build quality is weak☹☹
3         1
1533        realme C25_Y (Glacier Blue, 128 GB)  (4 GB RAM)
Please Don't Buy This Phone.This is totally waste of money.I was excited to see my phone 50 mp phone's camera but
camera is very bad.In this phone software is very different from any other realme phone.In this phone there is no one
app of photos and videos.In this phone settings is very different from any other realme phone.In this phone there is no
app lock and no hide apps settings.And there is no option of return in it, so that's why I could not even return this
phone.Please Don't ...         1         1
1534        realme C25_Y (Metal Grey, 128 GB)  (4 GB RAM)
Battery good,camera quality not goodPerformance is not good very slow work I am not satisfied with performance I don't
like this phone because its performance is not able to me.
1         1

..
768         Mi 5X 125.7 cm (50 inch) Ultra HD (4K) LED Smart Android TV with Dolby Atmos and Dolby Vision
Super awesome  mine is 55inch tv got delivered next day of Order but installation is very badd no response from
Flipkart. So finaly called mi store and they did it in same day. And coming to tv specs are awesome1. Sound is good but
bass is not great2 picture quality is awesome 3 adaptive Brightness is not that good but it's fine u can enjoy it4
connectivity is fine ok ok5 tv design is awesome but u may feel light electric shocks on corners when u touch as it's
made of metal6. Patch wall...       4         1
769         Mi 5X 125.7 cm (50 inch) Ultra HD (4K) LED Smart Android TV with Dolby Atmos and Dolby Vision
Coming straight to the point, the tv feels a bit laggy at times. Switching apps takes time. The sound is not upto the
mark, tried all settings available to optimize but couldn't. You definitely have to use a soundbar for better
experience. I also noticed a bug, the tv automatically redirects to patchwall while playing YouTube videos. Happened
once.Another bug I saw, while switching from SD to HD channel, the auto-scalling cropped half of the screen for a
while. Switched it off and then on ag...  2         1
770         Mi 5X 125.7 cm (50 inch) Ultra HD (4K) LED Smart Android TV with Dolby Atmos and Dolby Vision
Great Buy.!I was in confusion after seeing the ratings given by people but still i went for it.And i must say the TV is
great and i am happy with it completely.I have 50 inch varient.Display- details are great. Red colour is a bit
overwhelming, but after a few tweaks in the Picture settings it is absolutely stunning. Watching cricket matches,
movies is absolute delight.you can increase the backlight and adjust contrast,brightness as per your liking.Sound-Sound
is great but i would reco...        5         1
771         Mi 5X 125.7 cm (50 inch) Ultra HD (4K) LED Smart Android TV with Dolby Atmos and Dolby Vision
Posting this review after using this TV for 8 days..1.Picture quality is just awesome. Every pixel is clear.4k 8k
videos just gives you awesome feel.2.Sound quality is better than previous MI 20w speakers, Sound is loud as this TV
having 40w speakers but I am not happy with bass of these speakers. Should have more bass with this price
range.3.Connectivity is good. I am using with 200mbps wifi. Remote is smooth and Google assistant is working
excellent.4. Patchwall and other features a...         5         1
2303        MSI 27 inch Full HD IPS Panel Monitor (PRO MP271(3PA2))  (Response Time: 5 ms, 75 Hz Refresh Rate)
Worth the money u spend for this monitor Great deal Using for cctv footage monitorWonderful built MSI brand which we
can trust for
5         1
Length: 2304, dtype: int64
```

```
plt.figure(figsize=(17, 5))
sns.heatmap(df.isnull(), cbar=True, yticklabels=False)
plt.xlabel("Column_Name", size=14, weight="bold")
plt.title("Places of missing values in column",size=17)
plt.show()
```


```
import plotly.graph_objects as go
Top_Location_Of_review= df['Review'].value_counts().head (10)
Top_Location_Of_review
```

```
Good
47
Nice
32
Nice product
27
Very good
15
-Camera □  quality is very poor...-Now days in this budget you can expect at least a case...
8
Screen awesome but speaker □  Too bad yr     Not satisfied  settings ad speaker
8
My review After using 3 days Camera- 2/5Battery- 5/5 because of fast chargerDISPLAY-3/5 because no FHD+
DisplayPerformanc-4/5Back panel and build quality is weak☹☹
8
Please Don't Buy This Phone.This is totally waste of money.I was excited to see my phone 50 mp phone's camera but
camera is very bad.In this phone software is very different from any other realme phone.In this phone there is no one
app of photos and videos.In this phone settings is very different from any other realme phone.In this phone there is no
app lock and no hide apps settings.And there is no option of return in it, so that's why I could not even return this
phone.Please Don't ...     8
I am very happy to purchase this product and it's display and camera quality is very good . I recommend you to buy this
product
8
Battery good,camera quality not goodPerformance is not good very slow work I am not satisfied with performance I don't
like this phone because its performance is not able to me.
8
Name: Review, dtype: int64
```

```
from nltk. corpus import stopwords
stop = stopwords.words('english')
df['Review'].apply(lambda x: [item for item in x if item not in stop])
df.shape
```

```
(2304, 4)
```

```
!pip install tweet-preprocessor
```

```
Collecting tweet-preprocessor
  Downloading tweet_preprocessor-0.6.0-py3-none-any.whl (27 kB)
Installing collected packages: tweet-preprocessor
Successfully installed tweet-preprocessor-0.6.0
```

```
#Remove unnecessary characters
punct  =  ['%','/',':','\\','&amp','&',';','?']

def remove_punctuations(text):
  for punctuation in punct:
    text = text.replace(punctuation,'')
  return text


df['Review'] = df['Review'].apply(lambda x: remove_punctuations(x))
```

```
#Drop tweets that has empty text fields
df['Review'].replace( '', np.nan, inplace=True)
df.dropna(subset=["Review"],inplace=True)
len(df)
```

```
2304
```

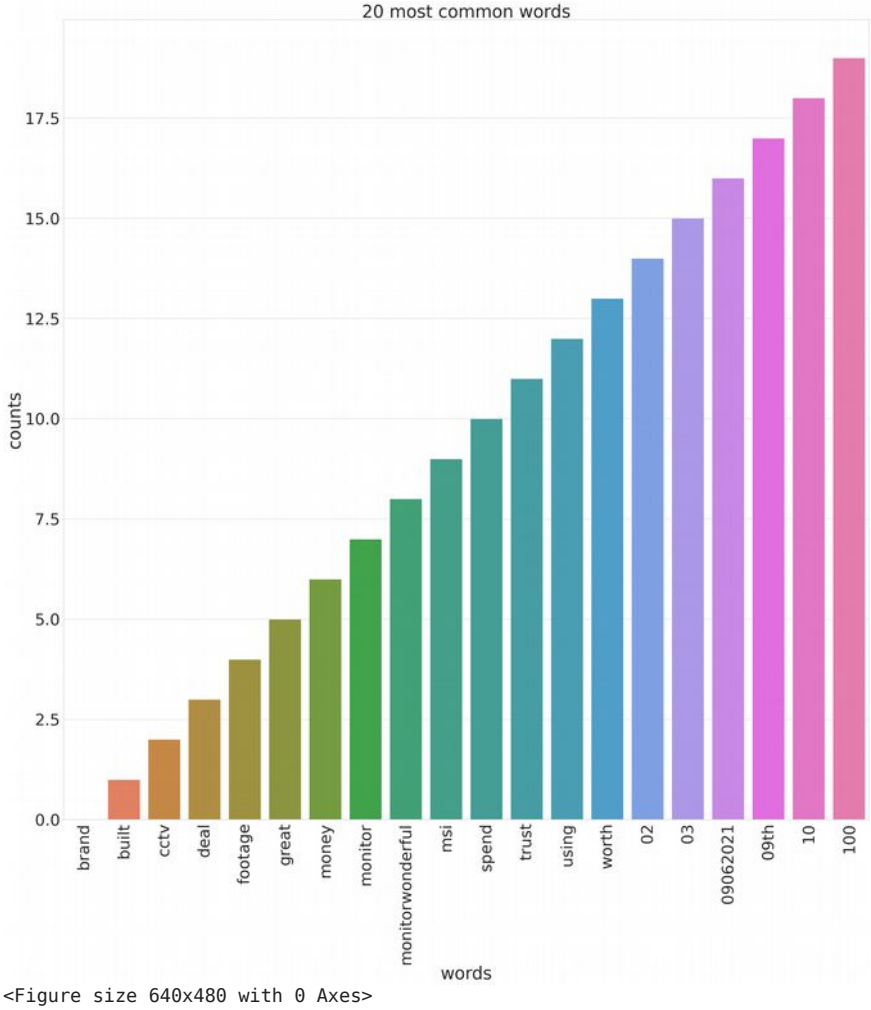
```
df = df.reset_index(drop=True)
df.head()
```

	Unnamed: 0	Product_name	Review	Rating
0	0	Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600...	Best under 60k Great performanceI 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

```
from sklearn.feature_extraction. text import TfidfVectorizer, CountVectorizer
```

```
sns.set_style('whitegrid')
%matplotlib inline
stop=stop+['Virgin America' , 'San Francisco' , 'Boston' , 'New York', ' customer' ,'flight' , 'airline', 'San Diego' , 'Oak
def plot_20_most_common_words(count_data, count_vectorizer) :
  import matplotlib. pyplot as plt
  words = count_vectorizer.get_feature_names_out()
  total_counts = np. zeros(len(words))
  for t in count_data:
    total_counts = t.toarray()[0]

  count_dict = (zip(words, total_counts))
  count_dict = sorted(count_dict, key=lambda x:x[1],reverse=True)[0:20]
  words = [w[0] for w in count_dict]
  counts = [w[1] for w in count_dict]
  x_pos = np.arange(len(words))

  plt.figure(2, (40,40))
  plt.subplot(title = '20 most common words')
  sns. set_context('notebook',font_scale=4,rc={ 'lines.linewidth' :2.5})
  sns.barplot(x_pos, palette='husl')
  plt.xticks(x_pos, words, rotation=90)
  plt.xlabel('words')
  plt.ylabel('counts')
  plt.show()

count_vectorizer = CountVectorizer(stop_words=stop)
# Fit and transform the processed titles
count_data = count_vectorizer.fit_transform(df['Review'])
# print(count_vectorizer)
# print(count_data)
# Visualise the 20 most common words
plot_20_most_common_words(count_data,count_vectorizer)
plt.savefig('saved_figure.png')
```

```
<ipython-input-56-a75f82b9560e>:20: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed i
```

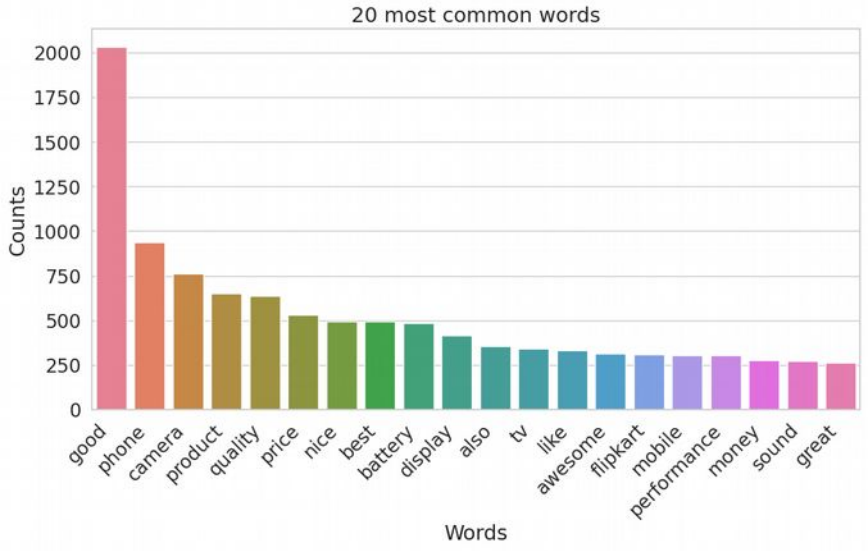



```
<Figure size 640x480 with 0 Axes>
```

```
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

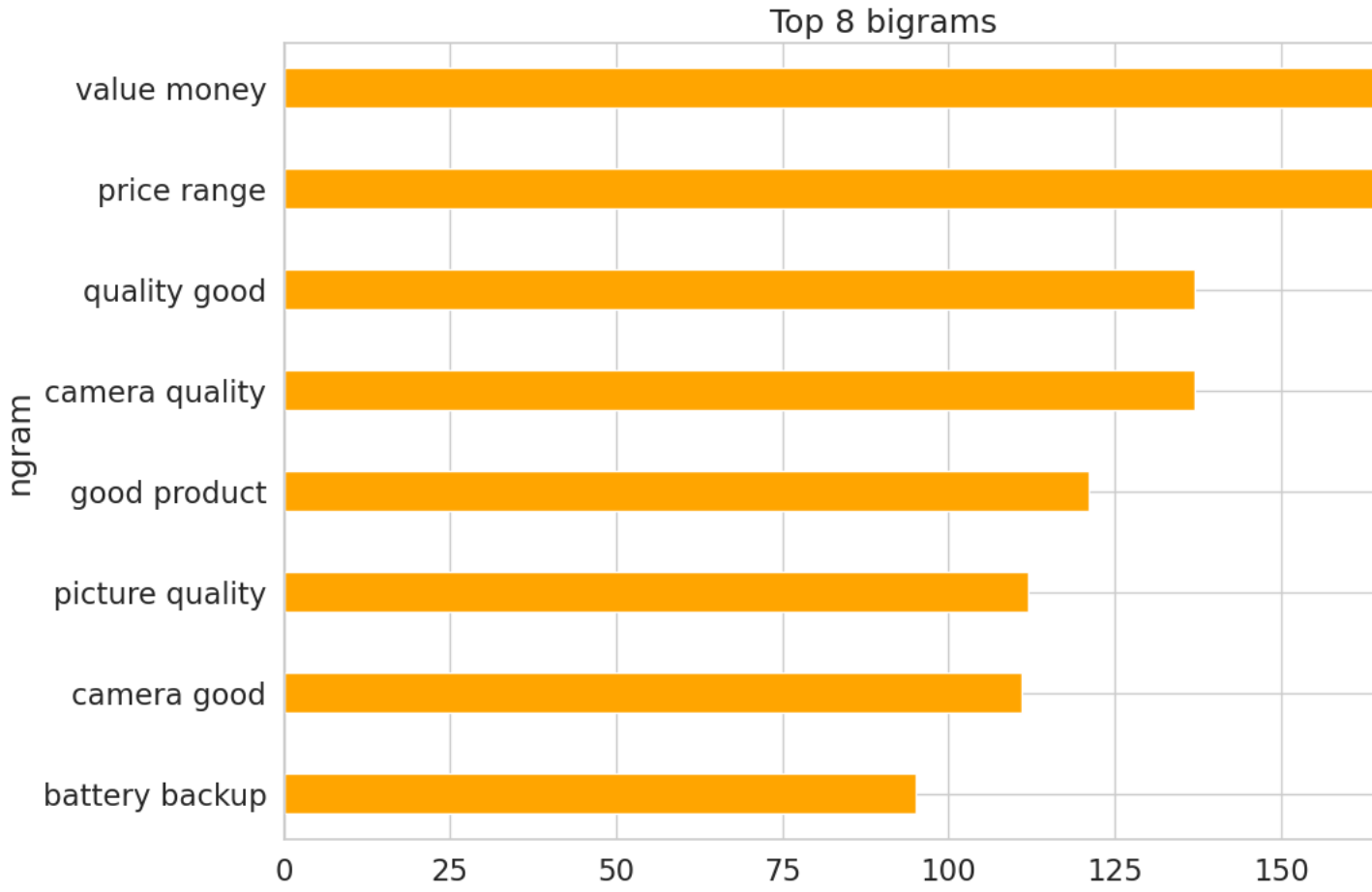
```
from sklearn.feature_extraction.text import CountVectorizer

sns.set_style('whitegrid')
%matplotlib inline

stop = stop + ['Virgin America', 'San Francisco', 'Boston', 'New York', 'customer', 'flight', 'airline', 'San Diego', 'Oaklan

def plot_20_most_common_words(count_data, count_vectorizer):
    words = count_vectorizer.get_feature_names_out()
    total_counts = np.zeros(len(words))

    for t in count_data:
        total_counts += t.toarray()[0]

    count_dict = dict(zip(words, total_counts))
    count_dict = sorted(count_dict.items(), key=lambda x: x[1], reverse=True)[:20]

    words = [w[0] for w in count_dict]
    counts = [w[1] for w in count_dict]

    x_pos = np.arange(len(words))

    plt.figure(figsize=(12, 6))
    sns.set_context('notebook', font_scale=1.5)
    sns.barplot(x=x_pos, y=counts, palette='husl')
    plt.title('20 most common words')
    plt.xticks(x_pos, words, rotation=45, ha='right')
    plt.xlabel('Words')
    plt.ylabel('Counts')
    plt.show()


count_vectorizer = CountVectorizer(stop_words=stop)
count_data = count_vectorizer.fit_transform(df['Review'])

# Visualize the 20 most common words
plot_20_most_common_words(count_data, count_vectorizer)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/feature_extraction/text.py:409
  warnings.warn(
<ipython-input-26-431feb3fb6de>:28: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed i

  sns.barplot(x=x_pos, y=counts, palette='husl')
```



```
import cufflinks as cf
cf.go_offline()
cf.set_config_file(offline=False, world_readable=True)

def get_top_n_bigram(corpus, n=None) :
  vec = CountVectorizer(ngram_range=(2, 4), stop_words="english").fit(corpus)
  bag_of_words = vec.transform(corpus)
  sum_words = bag_of_words.sum(axis=0)
  words_freq =[(word, sum_words[0, idx]) for word, idx in vec.vocabulary_.items()]
  words_freq =sorted(words_freq, key = lambda x: x[1], reverse=True)
  return words_freq[:n]

common_words = get_top_n_bigram(df['Review'] , 8)
mydict={}
for word, freq in common_words:
  bigram_df = pd.DataFrame(common_words,columns = ['ngram', 'count'])

bigram_df.groupby( 'ngram' ).sum()['count'].sort_values(ascending=False).sort_values().plot.barh(title = 'Top 8 bigrams',col
```

```
<Axes: title={'center': 'Top 8 bigrams'}, ylabel='ngram'>
```




```
def get_subjectivity(text):
  return TextBlob(text).sentiment.subjectivity
def get_polarity(text):
  return TextBlob(text).sentiment.polarity


df['subjectivity']=df[ 'Review'].apply(get_subjectivity)
df[ 'polarity' ]=df[ 'Review'].apply(get_polarity)
df.head()
```

	Unnamed: 0	Product_name	Review	Rating	subjectivity	polar
0	0	Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600...	Best under 60k Great performancel got it for a...	5	0.472424	0.438
1	1	Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600...	Good perfomence...	5	0.600000	0.700
2	2	Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600...	Great performance but usually it has also that...	5	0.666667	0.183
3	3	DELL Inspiron Athlon Dual Core 3050U - (4 GB/2...	My wife is so happy and best product 😘	5	0.650000	0.900
4	4	DELL Inspiron Athlon Dual Core 3050U - (4 GB/2...	Light weight laptop with new amazing features,...	5	0.763636	0.534

5. Sentiment Analysis

```
df['textblob_score'] =df['Review'].apply(lambda x: TextBlob(x).sentiment.polarity)
df
```

	Unnamed: 0	Product_name	Review	Rating	subjectivity	polarity	textblob_score	textbl
0	0	Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600...	Best under 60k Great performancel got it for a...	5	0.472424	0.438788	0.438788	
1	1	Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600...	Good perfomence...	5	0.600000	0.700000	0.700000	
2	2	Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600...	Great performance but usually it has also that...	5	0.666667	0.183333	0.183333	
3	3	DELL Inspiron Athlon Dual Core 3050U - (4 GB/2...	My wife is so happy and best product 😘	5	0.650000	0.900000	0.900000	
4	4	DELL Inspiron Athlon Dual Core 3050U - (4 GB/2...	Light weight laptop with new amazing features,...	5	0.763636	0.534091	0.534091	
...	...	...	...	...	...	...	...	
2299	2299	MSI 27 inch Full HD IPS Panel Monitor (PRO MP2...	Great display, accurate colours at this price ...	5	0.691667	0.600000	0.600000	
2300	2300	MSI 27 inch Full HD IPS Panel Monitor (PRO MP2...	Superb monitor first brought 1 used for 2 mont...	5	0.666667	0.625000	0.625000	
2301	2301	MSI 27 inch Full HD IPS Panel Monitor (PRO MP2...	Awesome	5	1.000000	1.000000	1.000000	
2302	2302	MSI 27 inch Full HD IPS Panel Monitor (PRO MP2...	Only one issue with adapter	5	1.000000	0.000000	0.000000	
2303	2303	MSI 27 inch Full HD IPS Panel Monitor (PRO MP2...	Worth the money u spend for this monitor Great...	5	0.425000	0.550000	0.550000	

2304 rows × 8 columns

```
neutral_threshold=0.05
```

```
df['textblob_sentiment']=df[ 'textblob_score'].apply(lambda c:'positive' if c >= neutral_threshold else ('Negative' if c <=
df
```

	Unnamed: 0	Product_name	Review	Rating	subjectivity	polarity	textblob_score	textbl
0	0	Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600...	Best under 60k Great performancel got it for a...	5	0.472424	0.438788	0.438788	
1	1	Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600...	Good perfomence...	5	0.600000	0.700000	0.700000	
2	2	Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600...	Great performance but usually it has also that...	5	0.666667	0.183333	0.183333	
3	3	DELL Inspiron Athlon Dual Core 3050U - (4 GB/2...	My wife is so happy and best product 😘	5	0.650000	0.900000	0.900000	
4	4	DELL Inspiron Athlon Dual Core 3050U - (4 GB/2...	Light weight laptop with new amazing features,...	5	0.763636	0.534091	0.534091	
...	...	...	...	...	...	...	...	
2299	2299	MSI 27 inch Full HD IPS Panel Monitor (PRO MP2...	Great display, accurate colours at this price ...	5	0.691667	0.600000	0.600000	
2300	2300	MSI 27 inch Full HD IPS Panel Monitor (PRO MP2...	Superb monitor first brought 1 used for 2 mont...	5	0.666667	0.625000	0.625000	
2301	2301	MSI 27 inch Full HD IPS Panel Monitor (PRO MP2...	Awesome	5	1.000000	1.000000	1.000000	
2302	2302	MSI 27 inch Full HD IPS Panel Monitor (PRO MP2...	Only one issue with adapter	5	1.000000	0.000000	0.000000	
2303	2303	MSI 27 inch Full HD IPS Panel Monitor (PRO MP2...	Worth the money u spend for this monitor Great...	5	0.425000	0.550000	0.550000	

2304 rows × 8 columns

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

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

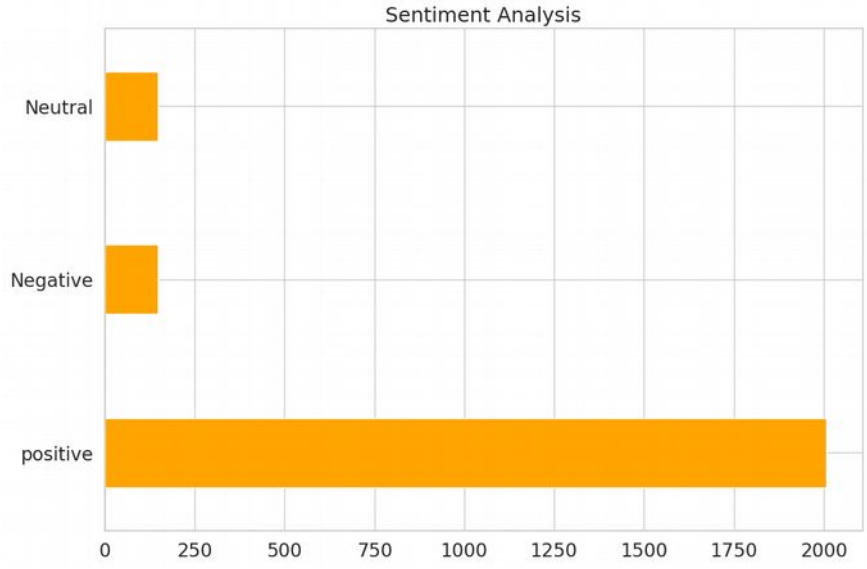
2304 rows × 3 columns

```
textblob_df["textblob_sentiment"].value_counts()
```

```
positive    2008
Negative     148
Neutral      148
Name: textblob_sentiment, dtype: int64
```

```
textblob_df["textblob_sentiment"].value_counts().plot.barh(title = 'Sentiment Analysis',color='orange' , width=.4, figsize=(
```

```
<Axes: title={'center': 'Sentiment Analysis'}>
```




```
df_positive=textblob_df[textblob_df['textblob_sentiment']=='positive' ]
```

```
df_very_positive=df_positive[df_positive['Rating']>3]
```

```
df_negative=textblob_df[textblob_df['textblob_sentiment']=='Negative' ]
```

```
df_negative
```

	Review	textblob_sentiment	Rating
31	Every thing is just fine but the tn display is...	Negative	5
39	Display so bad quality and wifi connectivity i...	Negative	1
65	Very bad experience	Negative	1
132	Giving this review after using for 7days*Slik ...	Negative	4
171	Phone Is Value For Maney ..But Camera Is Bad C...	Negative	1
...	...	...	...
2174	I am very upset because monitor is coming with...	Negative	2
2175	1 year back my lenovo order not nice very bad ...	Negative	1
2199	HDMI CABLE VERY POOR QUALITY	Negative	1
2214	as expected from dell	Negative	5
2284	Pros - Slim border line style. Lightweight co...	Negative	3

148 rows × 3 columns

```
df_neutral=textblob_df[textblob_df['textblob_sentiment']=='Neutral' ]
```

```
df_neutral
```

	Review	textblob_sentiment	Rating
38	nice product everything is okay but battery po...	Neutral	3
52	Im statisfied .. valueble money	Neutral	5
60	nyc laptop	Neutral	4
64	This price range it's super…..	Neutral	5
67	Nyce product	Neutral	4
...	...	...	...
2198	Value for money monitor with IPS panel.	Neutral	5
2236	awesom ips led monitor	Neutral	5
2240	pesa vasool, dont think twice go for it.btw i ...	Neutral	5
2294	The device is really nice and stand is sturdy ...	Neutral	5
2302	Only one issue with adapter	Neutral	5

148 rows × 3 columns

```
from wordcloud import WordCloud, STOPWORDS
from PIL import Image
```

```
#Creating the text variable
positive_tw =" ".join(t for t in df_very_positive.Review)
# Creating word _ cloud with text as argument in . generate() rtpthod
word_cloud1 = WordCloud(collocations = False, background_color = 'white') .generate(positive_tw)
# Display the generated Word Cloud
plt. imshow(word_cloud1, interpolation='bilinear')
plt.axis('off')
plt.show()
```


```
#Creating the text variable
negative_tw =" ".join(t for t in df_negative.Review)
# Creating word _ cloud with text as argument in . generate() rtpthod
```

(

```
    word_cloud2 = WordCloud(collocations = False, background_color = 'white') .generate(negative_tw)
    # Display the generated Word Cloud
    plt. imshow(word_cloud2, interpolation='bilinear')
    plt.axis('off')
    plt.show()
```