

▾ Sentiment Analysis

By : Sampurna Dey

Linkedin: www.linkedin.com/in/sampurna-dey-060573129

Github: <https://github.com/sampurna-project>

Importing all Libraries and Dependencies

```
#importing the necessary libraries
import pandas as pd
import numpy as np
from wordcloud import WordCloud
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.ensemble import AdaBoostClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
import warnings
warnings.filterwarnings("ignore")

Threadsdata=pd.read_csv("/content/threads_reviews.csv")
Threadsdata
```

	source	review_description	rating	review_date
0	Google Play	Meh. Not the greatest experience on a Chromebo...	2	2023-07-08 14:18:24
1	Google Play	Pretty good for a first launch!! Its easy to u...	3	2023-07-19 20:52:48
2	Google Play	For a brand new app, it's very well optimized....	3	2023-07-06 23:03:11
3	Google Play	Great app with a lot of potential! However, th...	3	2023-07-10 00:53:25
4	Google Play	The app is good, but it needs a lot of functio...	3	2023-07-06 16:57:43
...
32905	App Store	This killed my dog. Mark zuckerburg strangled ...	1	2023-07-06 01:23:55
32906	App Store	Add Search and hashtag like Twitter !	1	2023-07-19 08:01:06
...	2023-07-17

```
import nltk
nltk.download('stopwords')
nltk.download('wordnet')
nltk.download('punkt')

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Unzipping corpora/stopwords.zip.
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt.zip.
True
```

```
#checking the datatypes
Threadsdata.dtypes
```

```
source          object
review_description  object
rating           int64
review_date      object
dtype: object
```

```
#checking for missing values
Threadsdata.isna().sum()

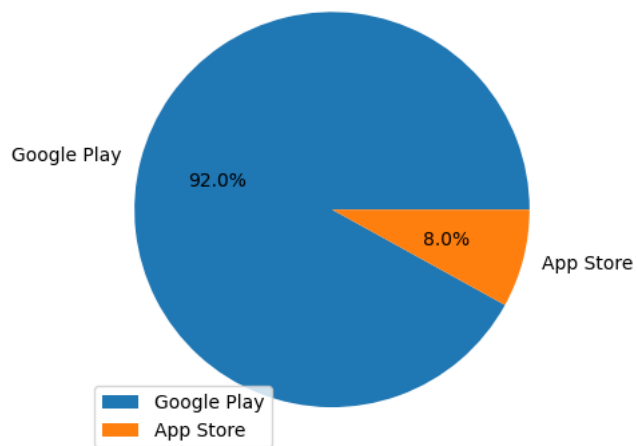
source          0
review_description  0
rating          0
review_date     0
dtype: int64
```

Data Visualization

```
Threadsdata['source'].value_counts()
```

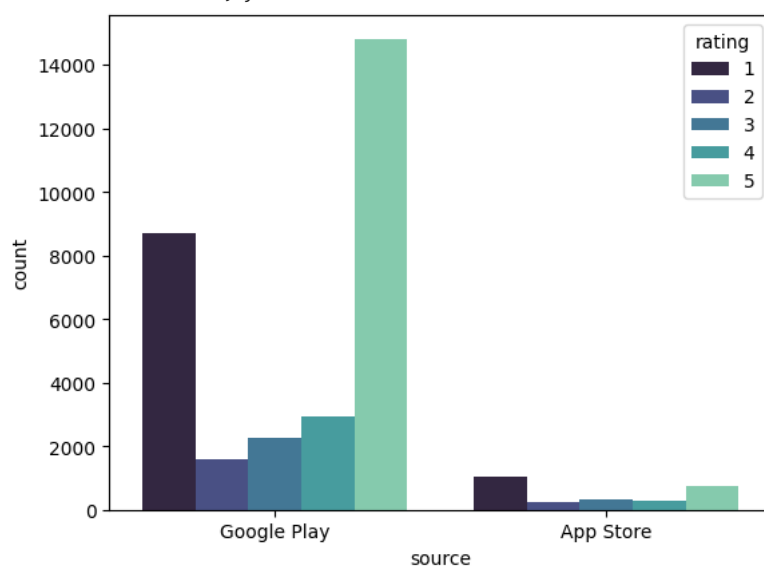
```
Google Play    30270
App Store      2640
Name: source, dtype: int64
```

```
#checking the source where review is more
y=Threadsdata['source'].value_counts()
labels=['Google Play','App Store']
plt.pie(y,labels=labels,autopct='%1.1f%%')
plt.legend(loc='lower left')
plt.show()
```

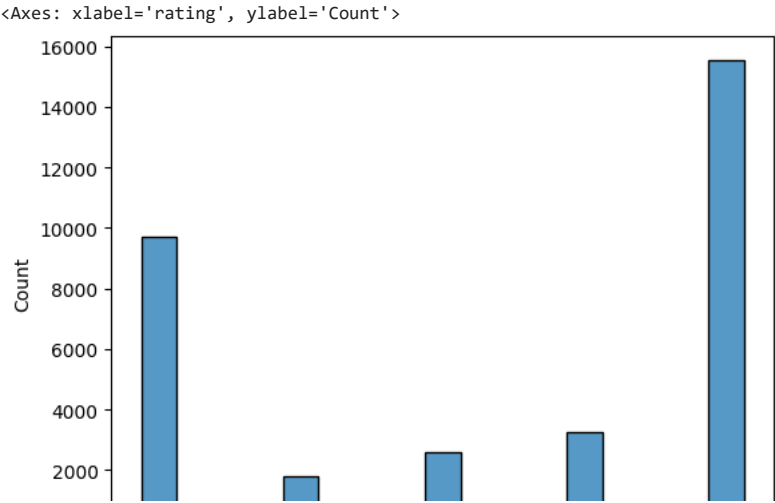


```
sns.countplot(x='source',data=Threadsdata,palette='mako',hue='rating')
```

<Axes: xlabel='source', ylabel='count'>



```
sns.histplot(x='rating',data=Threadsdata)
```



```
#clubbing rating 1 and 2 as '-1' , 3 as '0' and 4,5 as '1'
Threadsdata['rating']=Threadsdata['rating'].map({1:-1,2:-1,3:0,4:1,5:1})
Threadsdata
```

	source	review_description	rating	review_date
0	Google Play	Meh. Not the greatest experience on a Chromebo...	-1	2023-07-08 14:18:24
1	Google Play	Pretty good for a first launch!! Its easy to u...	0	2023-07-19 20:52:48
2	Google Play	For a brand new app, it's very well optimized....	0	2023-07-06 23:03:11
3	Google Play	Great app with a lot of potential! However, th...	0	2023-07-10 00:53:25
4	Google Play	The app is good, but it needs a lot of functio...	0	2023-07-06 16:57:43
...
32905	App Store	This killed my dog. Mark zuckerburg strangled ...	-1	2023-07-06 01:23:55
32906	App Store	Add Search and hashtag like Twitter !	-1	2023-07-19 08:01:06
...	2023-07-17

```
#dropping other columns
Threadsdata.drop(['source','review_date'],axis=1,inplace=True)
```

Threadsdata

	review_description	rating		
0	Meh. Not the greatest experience on a Chromebo...	-1		
1	Pretty good for a first launch!! Its easy to u...	0		
2	For a brand new app, it's very well optimized....	0		
3	Great app with a lot of potential! However, th...	0		
4	The app is good, but it needs a lot of functio...	0		
...		
32905	This killed my dog. Mark zuckerburg strangled ...	-1		
32906	Add Search and hashtag like Twitter !	-1		
32907	bad twister	-1		
32908	Yet another trash from Meta.	-1		
32909	Nothing special this app is just a copy of twi...	-1		

32910 rows × 2 columns

```
Threadsdata['rating'].unique()

array([-1,  0,  1])
```

```
Rev=Threadsdata.review_description
Rev
```

rev

```
0      Meh. Not the greatest experience on a Chromebo...
1      Pretty good for a first launch!! Its easy to u...
2      For a brand new app, it's very well optimized...
3      Great app with a lot of potential! However, th...
4      The app is good, but it needs a lot of functio...

...

32905   This killed my dog. Mark zuckerburg strangled ...
32906           Add Search and hashtag like Twitter !
32907                               bad twister
32908                               Yet another trash from Meta.
32909   Nothing special this app is just a copy of twi...
Name: review_description, Length: 32910, dtype: object
```

```
#remove special characters
Rev=Rev.str.replace('[^a-zA-Z0-9]+', " ")
Rev
```

```
0      Meh Not the greatest experience on a Chromeboo...
1      Pretty good for a first launch Its easy to use...
2      For a brand new app it s very well optimized H...
3      Great app with a lot of potential However ther...
4      The app is good but it needs a lot of function...

...

32905   This killed my dog Mark zuckerburg strangled h...
32906           Add Search and hashtag like Twitter
32907                               bad twister
32908                               Yet another trash from Meta
32909   Nothing special this app is just a copy of twi...
Name: review_description, Length: 32910, dtype: object
```

```
#plotting the word cloud of negative values
a=Threadsdata.loc[Threadsdata['rating']==-1][['review_description','rating']]
a
```

	review_description	rating
0	Meh. Not the greatest experience on a Chromebo...	-1
5	Currently, it's very challenging to use. It's ...	-1
6	I still don't want to see content from people ...	-1
7	Could be great if all pages loaded when you cl...	-1
10	Over all, the UI of this app is good. Using it...	-1
...
32905	This killed my dog. Mark zuckerburg strangled ...	-1
32906	Add Search and hashtag like Twitter !	-1
32907	bad twister	-1
32908	Yet another trash from Meta.	-1
32909	Nothing special this app is just a copy of twi...	-1

11522 rows × 2 columns

```
text_corpus = ' '.join(a['review_description'].values)
wordcloud = WordCloud(width=800, height=400, background_color='aqua').generate(text_corpus)
# Display the generated word cloud using matplotlib
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```



```
#plotting the word cloud of neutral values
```

```
b=Threadsdata.loc[Threadsdata['rating']==0][['review_description','rating']]
```

b

	review_description	rating
1	Pretty good for a first launch!! Its easy to u...	0
2	For a brand new app, it's very well optimized....	0
3	Great app with a lot of potential! However, th...	0
4	The app is good, but it needs a lot of functio...	0
8	I'm liking the concept! There is room for impr...	0
...
32857	Please add arabic language .	0
32862	Add gyphy - to reply 🧡 love new apps!	0
32864	i really just wanna say anything i want	0
32870	Threads is Twitter without a soul. It's not tr...	0
32875	kinda like this app, but pls put a gif search ...	0

2585 rows × 2 columns

```
text_corpus = ' '.join(b['review_description'].values)
```

```
wordcloud = WordCloud(width=800, height=400, background_color='white').generate(text_corpus)
```

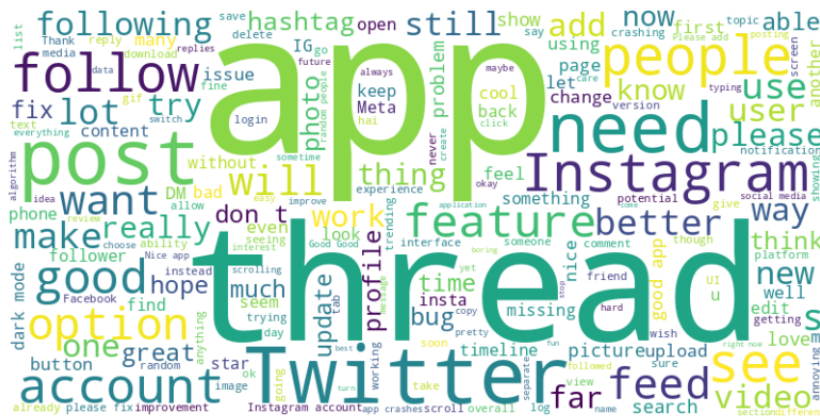
```
# Display the generated word cloud using matplotlib
```

```
plt.figure(figsize=(10, 5))
```

```
plt.imshow(wordcloud, interpolation='bilinear')
```

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

```
plt.show()
```



```
#plotting the word cloud of positive values
```

```
c=Threadsdata.loc[Threadsdata['rating']==1][['review_description','rating']]
```

C

	review_description	rating
9	Not bad! This is its first launch. There is st...	1
21	Firstly, I want to express my appreciation for...	1
22	It's not bad so far. A handful of unsolicited ...	1
29	Great App! Loved it from the verv first day al...	1

```
text_corpus = ' '.join(c['review_description'].values)
wordcloud = WordCloud(width=800, height=400, background_color='seagreen').generate(text_corpus)
# Display the generated word cloud using matplotlib
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```



STEMMING AND TOKENIZATION

```
from nltk.stem import PorterStemmer
from nltk.tokenize import word_tokenize

ps=PorterStemmer()
Rev=Rev.apply(lambda x:[ps.stem(i.lower()) for i in word_tokenize(x)]).apply(lambda y: " ".join(y))
Rev
```

```
0      meh not the greatest experi on a chromebook se...
1      pretti good for a first launch it easi to use ...
2      for a brand new app it s veri well optim howev...
3      great app with a lot of potenti howev there is...
4      the app is good but it need a lot of function ...
      ...
32905  thi kill my dog mark zuckerburg strangl him my...
32906          add search and hashtag like twitter
32907                      bad twister
32908                      yet anoth trash from meta
32909          noth special thi app is just a copi of twitter
Name: review description, Length: 32910, dtype: object
```

Removing Stopwords

```
from nltk.corpus import stopwords
st=stopwords.words('english')
Rev=Rev.apply(lambda x:[i for i in word_tokenize(x) if i.lower() not in st]).apply(lambda y: " ".join(y))
Rev
```

```

0      meh greatest experi chromebook seem custom pho...
1      pretti good first launch easi use self explana...
2      brand new app veri well optim howev miss quit ...
3      great app lot potenti howev lot need fix examp...
4      app good need lot function exampl search topic...
      ...
32905      thi kill dog mark zuckerburg strangl dog gone
32906      add search hashtag like twitter
32907      bad twister
32908      vet anoth trash meta

```

```
32909          noth special thi app copi twitter
Name: review_description, Length: 32910, dtype: object
```

```
Rev=Rev.apply(lambda x:[i for i in word_tokenize(x) if len(i)>2]).apply(lambda y:" ".join(y))
Rev
```

```
0      meh greatest experi chromebook seem custom pho...
1      pretti good first launch easi use self explana...
2      brand new app veri well optim howev miss quit ...
3      great app lot potenti howev lot need fix exampl...
4      app good need lot function exampl search topic...

...
32905      thi kill dog mark zuckerburg strangl dog gone
32906          add search hashtag like twitter
32907          bad twister
32908          yet anoth trash meta
32909          noth special thi app copi twitter
Name: review_description, Length: 32910, dtype: object
```

TFIDF VECTORISATION

```
from sklearn.feature_extraction.text import TfidfVectorizer
tf1=TfidfVectorizer()
data_vec=tf1.fit_transform(Rcv)
print(data_vec)
```

```
(0, 5207)      0.15107027859318817
(0, 4006)      0.10035493113397626
(0, 6722)      0.0972462960534521
(0, 11765)     0.10190070056219779
(0, 4582)      0.09006995803505909
(0, 818)       0.17640156906018106
(0, 4152)      0.18277132394750817
(0, 2498)      0.14590082712612149
(0, 10287)     0.13008881208148929
(0, 11255)     0.1367759733215397
(0, 1414)      0.17872155139664647
(0, 9976)      0.11153211777829317
(0, 11830)     0.10333757362094909
(0, 2989)      0.13275520426364992
(0, 2176)      0.16365901848925332
(0, 2459)      0.22335978334073403
(0, 6472)      0.14449952041105937
(0, 10698)     0.12275693038232713
(0, 6123)      0.1086552696873881
(0, 781)       0.046938259548761505
(0, 3668)      0.1187076445871143
(0, 3804)      0.14372863398250377
(0, 5799)      0.20044455249662121
(0, 7342)      0.22335978334073403
(0, 2399)      0.13612915119801525
:
:
(32904, 11083)      0.3635489084879956
(32905, 10193)      0.3984093272516867
(32905, 12217)      0.36431849565271374
(32905, 3210) 0.693663780253189
(32905, 6067) 0.29525445300150216
(32905, 6812) 0.19902851664332444
(32905, 4506) 0.2976556252268803
(32905, 10698)      0.10948145032558744
(32906, 376) 0.4650001125351025
(32906, 9471) 0.5432906890529202
(32906, 11145)      0.2767477991559158
(32906, 4776) 0.5472081378814833
(32906, 6472) 0.335535440794831
(32907, 11128)      0.9058422277265952
(32907, 1086) 0.42361522454619027
(32908, 10979)      0.5623300319980792
(32908, 12092)      0.521053865803039
(32908, 6998) 0.4270977123082578
(32908, 712) 0.4794531762182834
(32909, 10047)      0.6475785397733912
(32909, 2525) 0.4016426978554383
(32909, 7646) 0.4950998382971691
(32909, 11145)      0.2548897125418595
(32909, 10698)      0.26253435642146317
(32909, 781) 0.20076920665575368
```

```
y=Threadsdata['rating'].values
y
```

```
array([-1,  0,  0, ..., -1, -1, -1])
```

```
X_train,X_test,y_train,y_test=train_test_split(data_vec,y,test_size=0.2,random_state=1)
```

Model Building

```
sv=SVC()
dt=DecisionTreeClassifier()
rf=RandomForestClassifier()
ad=AdaBoostClassifier()
models=[sv,dt,rf,ad]
for model in models:
    print(model)
    model.fit(X_train,y_train)
    y_pred=model.predict(X_test)
    print(classification_report(y_test,y_pred))
```

```
SVC()
      precision    recall  f1-score   support

-1       0.79       0.74       0.76       2317
 0       0.43       0.02       0.05        532
 1       0.80       0.93       0.86       3733
```

```
   accuracy
macro avg   0.67       0.56       0.56       6582
weighted avg   0.76       0.79       0.76       6582
```

```
DecisionTreeClassifier()
      precision    recall  f1-score   support

-1       0.72       0.67       0.70       2317
 0       0.17       0.11       0.13        532
 1       0.78       0.86       0.82       3733
```

```
   accuracy
macro avg   0.56       0.55       0.55       6582
weighted avg   0.71       0.73       0.72       6582
```

```
RandomForestClassifier()
      precision    recall  f1-score   support

-1       0.74       0.76       0.75       2317
 0       0.22       0.02       0.03        532
 1       0.81       0.90       0.85       3733
```

```
   accuracy
macro avg   0.59       0.56       0.55       6582
weighted avg   0.74       0.78       0.75       6582
```

```
AdaBoostClassifier()
      precision    recall  f1-score   support

-1       0.77       0.54       0.63       2317
 0       0.26       0.06       0.10        532
 1       0.72       0.93       0.81       3733
```

```
   accuracy
macro avg   0.58       0.51       0.51       6582
weighted avg   0.70       0.72       0.69       6582
```

```
y_new=model.predict(tf1.transform(["Threads falls short due to its relentlessly grim and harrowing portrayal of a post apocalyptic world,
if y_new==-1:
    print("Negative")
if y_new==0:
    print("Neutral")
if y_new==1:
    print("Positive")
```

Negative

```
y_new=model.predict(tf1.transform(["Threads is an exceptional film that artfully weaves together powerful storytelling, outstanding perfc
if y_new==-1:
    print("Negative")
if y_new==0:
    print("Neutral")
if y_new==1:
    print("Positive")
```

Positive

End

+ Code

+ Text

✓ 0s completed at 10:01 AM

