

```

import pandas as pd
import numpy as np
import string
import spacy
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from sklearn.feature_extraction.text import CountVectorizer
import wordcloud
import matplotlib.pyplot as plt
from wordcloud import WordCloud, STOPWORDS

```

```

data=pd.read_csv('Downloads/apple.txt',error_bad_lines=False)
data

```

	x
0	It's amazing..... battery last around 10 -12 h...
1	Finally a MacBook possession. Got it in 50k in...
2	Best Laptop under 60k. If you don't want any a...
3	Pros:1. Light weight and super fast response t...
4	"
5	I am very upset about the product as I checked...
6	Reason for 1 star is I ot an defective piece. ...
7	Got the Apple Macbook air on amazon discount s...
8	White line coming on screen in just three mont...
9	Bought this product recently on 5 April, was w...
10	Hi, Think thousand time before buying. One cal...
11	We buy macbook's expecting that it will lasts ...
12	I use this for music production and it works a...

```

data=[x.strip() for x in data.x]
data=[x for x in data if x]
data

```

```

["It's amazing..... battery last around 10 -12 hrs.....best laptop for students and coders",
"Finally a MacBook possession. Got it in 50k in Diwali sale. Technically good laptop.Good News.luxury product, niche technol
ogy, good battery life, reliable machine (other laptops crash and become defunct in 6-10 years), no anti-virus required, secu
rity of data and transactions much better. Terrific sense of possession & pride.Bad News. Problems will be there if you are s
witching from windows based system - very less space in hard disk, no CD drive, inability to transfer data from Mac to your e
xisting external hard disc unless you format it, apps are mostly paid and re unreasonably expensive (no free apps which are a
vailable otherwise on google play store, even the angry bird costs Rs 400 !!), all printers are not compatible (e.g. the most
economical MFD laser printer RICOH SP 111 can't be used), huge compatibility issues with Pages (MS Word) and Keynote (powerpc
int) unless you master it by working on these (still mostly the document and slides either do not open in windows environment
or have distortion issues). you need to spend extra for an external CD writer and tp buy a carry bag. Mac con not be connecte
d to most of the projectors unless you buy some connectors, which are expensive and theres no clarity which one to buy and fr
om where. you can't connect it to your TV, the ports are different and again theres no clarity - customer support, manuals or
help section are silent ! Marketing strategy for Indian market, if at all has been planned, has been very bad. If these criti
cal issues (and a few more not mentioned here) are handled properly, the Mac can sweep away other laptops in India. Presentl
y, it doesn't seem to be happening.Overall Verdict.If compatibility issues mentioned above do not bother you, go for it. Els
e, think hard. If its an emotional issue to own a long cherished Mac, then its a different thing. Go ahead and gradually you
would figure out most (but not all) of the above problems, like I did.",
"Best Laptop under 60k. If you don't want any accounting softwares like tally or busy then go for it, better than windows in
every term. I just love it's batteryv life with sleek look 🍌Now talking about MS officeIt's available on app store. free for

```

```
data="".join(data)
data
```

```
'It's amazing..... battery last around 10 -12 hrs.....best laptop for students and codersFinally a MacBook possession. Got it in 50k in Diwali sale. Technically good laptop.Good News.luxury product, niche technology, good battery life, reliable machine (other laptops crash and become defunct in 6-10 years), no anti-virus required, security of data and transactions much better. Terrific sense of possession & pride.Bad News. Problems will be there if you are switching from windows based system - very less space in hard disk, no CD drive, inability to transfer data from Mac to your existing external hard disc unless you format it, apps are mostly paid and re unreasonably expensive (no free apps which are available otherwise on google play store, even the angry bird costs Rs 400 !!), all printers are not compatible (e.g. the most economical MFD laser printer RICOH SP 111 can't be used), huge compatibility issues with Pages (MS Word) and Keynote (powerpoint) unless you master it by working on these (still mostly the document and slides either do not open in windows environment or have distortion issues). you need to spend extra for an external CD writer and tp buy a carry bag. Mac can not be connected to most of the projectors unless you buy some connectors, which are expensive and theres no clarity which one to buy and from where. you can't connect it to your TV, the ports are different and again theres no clarity - customer support, manuals or help section are silent ! Marketing strategy for Indian market, if at all has been planned, has been very bad. If these critical issues (and a few more not mentioned here) are handled properly, the Mac can sweep away other laptops in India. Presently, it doesn't seem to be happening.O verall Verdict.If compatibility issues mentioned above do not bother you, go for it. Else, think hard. If its an emotional issue to own a long cherished Mac, then its a different thing. Go ahead and gradually you would figure out most (but not all) of the above problems, like I did.Best Laptop under 60k. If you don't want any accounting softwares like tally or busy then go for it, better than windows in every term. I just love it's battery life with sleek look 🍌Now talking about MS officeit
```

```
trans=data.translate(str.maketrans("","",string.punctuation))
trans
```

```
'Its amazing battery last around 10 12 hrsbest laptop for students and codersFinally a MacBook possession Got it in 50k in Diwali sale Technically good laptopGood Newsluxury product niche technology good battery life reliable machine other laptops crash and become defunct in 610 years no antivirus required security of data and transactions much better Terrific sense of possession prideBad News Problems will be there if you are switching from windows based system very less space in hard disk no CD drive inability to transfer data from Mac to your existing external hard disc unless you format it apps are mostly paid and re unreasonably expensive no free apps which are available otherwise on google play store even the angry bird costs Rs 400 all printers are not compatible eg the most economical MFD laser printer RICOH SP 111 cant be used huge compatibility issues with Pages MS Word and Keynote powerpoint unless you master it by working on these still mostly the document and slides either do not open in windows environment or have distortion issues you need to spend extra for an external CD writer and tp buy a carry bag Mac can not be connected to most of the projectors unless you buy some connectors which are expensive and theres no clarity which one to buy and from where you cant connect it to your TV the ports are different and again theres no clarity c
```

```
word=word_tokenize(trans)
word
```

```
['Its',
 'amazing',
 'battery',
 'last',
 'around',
 '10',
 '12',
 'hrsbest',
 'laptop',
 'for',
 'students',
 'and',
 'codersFinally',
 'a',
 'MacBook',
 'possession',
 'Got',
 'it',
 '==']
```

```
stop=stopwords.words('english')
stop.append('the')
stopwords=[word for word in word if not word in stop]
stopwords
```

```
[ 'Its',
  'amazing',
  'battery',
  'last',
  'around',
  '10',
  '12',
  'hrsbest',
  'laptop',
  'for',
  'students',
  'and',
  'codersFinally',
  'a',
  'MacBook',
  'possession',
  'Got',
  'it',
  ...]
```

lower=[x.lower() for x in stopwords]

lower

```
[ 'its',
  'amazing',
  'battery',
  'last',
  'around',
  '10',
  '12',
  'hrsbest',
  'laptop',
  'students',
  'codersfinally',
  'macbook',
  'possession',
  'got',
  '50k',
  'diwali',
  'sale',
  'technically',
  'good',
  ...]
```

ps=PorterStemmer()

p = [ps.stem(word) for word in lower]

p

```
[ 'its',
  'amazing',
  'battery',
  'last',
  'around',
  '10',
  '12',
  'hrsbest',
  'laptop',
  'students',
  'codersfinally',
  'macbook',
  'possession',
  'got',
  '50k',
  'diwali',
  'sale',
  'technically',
  'good',
  ...]
```

```
nlp=spacy.blank('en')
nlp
```

```
<spacy.lang.en.English at 0x2420c938970>
```

```
cv = CountVectorizer()
x = cv.fit_transform(p)
```

```
print(cv.vocabulary_)
```

```
{'it': 432, 'amaz': 74, 'batteri': 123, 'last': 463, 'around': 99, '10': 0, '12': 6, 'hrsbest': 393, 'laptop': 454, 'student': 768, 'codersfin': 185, 'macbook': 494, 'possess': 617, 'got': 361, '50k': 36, 'diwali': 255, 'sale': 687, 'technic': 784, 'good': 358, 'laptopgood': 457, 'newsluxuri': 549, 'product': 634, 'nich': 551, 'technolog': 785, 'life': 472, 'reliabl': 664, 'mac hin': 496, 'crash': 209, 'becom': 127, 'defunct': 232, '610': 46, 'year': 891, 'antiviru': 87, 'requir': 671, 'secur': 702, 'data': 220, 'transact': 826, 'much': 538, 'better': 132, 'terrif': 789, 'sens': 707, 'pridebad': 627, 'news': 548, 'problem': 631, 'switch': 777, 'window': 872, 'base': 121, 'system': 778, 'less': 469, 'space': 742, 'hard': 374, 'disk': 249, 'cd': 171, 'drive': 261, 'inabl': 408, 'transfer': 827, 'mac': 493, 'exist': 296, 'extern': 307, 'disc': 247, 'unless': 834, 'format': 335, 'app': 90, 'mostli': 532, 'paid': 590, 'unreason': 835, 'expens': 298, 'free': 338, 'avail': 107, 'otherwis': 583, 'googl': 360, 'play': 610, 'store': 761, 'even': 284, 'angri': 84, 'bird': 137, 'cost': 205, 'rs': 684, '400': 27, 'printer': 629, 'compact': 192, 'eg': 272, 'econom': 269, 'mfd': 519, 'laser': 462, 'ricoh': 680, 'sp': 741, '111': 4, 'cant': 163, 'use': 842, 'hugle': 394, 'issu': 429, 'page': 589, 'ms': 537, 'word': 878, 'keynot': 447, 'powerpoint': 619, 'master': 508, 'work': 881, 'still': 758, 'document': 256, 'slide': 727, 'either': 274, 'open': 575, 'environ': 283, 'distort': 253, 'need': 544, 'spend': 746, 'extra': 308, 'writer': 886, 'tp': 823, 'buy': 158, 'carri': 166, 'bag': 120, 'con': 193, 'connect': 197, 'projector': 641, 'connector': 198, 'there': 795, 'clariti': 181, 'one': 571, 'tv': 831, 'port': 615, 'differ': 243, 'custom': 216, 'support': 775, 'manual': 503, 'help': 382, 'section': 701, 'silent': 721, 'market': 505, 'strategi': 763, 'indian': 413, 'plan': 609, 'bad': 117, 'if': 399, 'critic': 212, 'mention': 517, 'handl': 371, 'properli': 642, 'sweep': 776, 'away': 110, 'india': 412, 'nresen
```

```
print(cv.get_feature_names()[0:10])
print(x.toarray()[0:10])
```

```
['10', '104', '10day', '11', '111', '1112', '12', '12000', '128', '128gb']
[[0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 ...
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]]
```

```
cv = CountVectorizer(ngram_range=(1,30),analyzer='word',max_features=30)
x = cv.fit_transform(p)
```

```
print(cv.vocabulary_)
```

```
{'batteri': 3, 'laptop': 13, 'macbook': 15, 'got': 10, 'good': 9, 'product': 18, 'year': 29, 'window': 26, 'mac': 14, 'cost': 6, 'use': 23, 'issu': 12, 'work': 27, 'need': 17, 'buy': 5, 'if': 11, 'go': 8, 'would': 28, 'want': 24, 'amazon': 1, 'warrant i': 25, 'servic': 21, 'replac': 20, 'qualiti': 19, 'month': 16, 'the': 22, 'appl': 2, 'air': 0, 'devic': 7, 'bought': 4}
```

```
print(x.toarray()[0:16])
```

```
['air', 'amazon', 'appl', 'batteri', 'bought', 'buy', 'cost', 'devic', 'go', 'good', 'got', 'if', 'issu', 'laptop', 'mac', 'mac book']
```

[illegible]

```
plt.imshow(wordcloud)
```

```
plot_cloud(wordcloud)
```