

```

-----import libraries-----
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, TfidfVectorizer
import wordcloud
import matplotlib.pyplot as plt
from wordcloud import WordCloud, STOPWORDS

```

```

-----read dataset-----
tweet = pd.read_csv('Downloads/Elon_musk.csv')
tweet

```

Unnamed: 0		Text
0	1	@kunalb11 I'm an alien
1	2	@ID_AA_Carmack Ray tracing on Cyberpunk with H...
2	3	@joerogan @Spotify Great interview!
3	4	@gtera27 Doge is underestimated
4	5	@teslacn Congratulations Tesla China for amazi...
...	...	...
1994	1995	@flicnhvy True, it sounds so surreal, but the n...
1995	1996	@PPathole Make sure to read ur terms & con...
1996	1997	@TeslaGong @PPathole Samwise Gamgee
1997	1998	@PPathole Altho Dumb and Dumber is <U+0001F525...
1998	1999	Progress update August 28

```

-----drop-----
tweet1 = tweet.iloc[:,1:]
tweet1

```

	Text
0	@kunalb11 I'm an alien
1	@ID_AA_Carmack Ray tracing on Cyberpunk with H...
2	@joerogan @Spotify Great interview!
3	@gtera27 Doge is underestimated
4	@teslacn Congratulations Tesla China for amazi...

```

tweet1 = [Text.strip() for Text in tweet1.Text]

```

```
tweet1 = [Text for Text in tweet1 if Text]
tweet1
```

```
['@kunalb11 I am an alien',
 '@ID_AA_Carmack Ray tracing on Cyberpunk with HDR is next-level. Have you tried it?',
 '@joerogan @Spotify Great interview!',
 '@gtera27 Doge is underestimated',
 '@teslacr Congratulations Tesla China for amazing execution last year. Now on to the next for even more!!'
 'Happy New Year of the Ox! https://t.co/9WFKMYu2oj',
 'Frodo was the underdog,\nAll thought he would fail,\nHimself most of all. https://t.co/zGxJFDzzrM',
 '@Owensparks_ @flcnhvy @anonyx10 Haha thanks :)',
 '@flcnhvy @anonyx10 Indeed! Tweets definitely do not represent real-world time allocation.',
 'The most entertaining outcome is the most likely',
 '@GiveDirectly Just sent some',
 'Just agree to do Clubhouse with @kanyewest',
 'https://t.co/3rWE9uHSTS',
 '@geoffkeighley @UnrealEngine It's getting real',
 'Bought some Dogecoin for lil X, so he can be a toddler hodler',
 '@JoshManMode He definitely has issues, but the sentencing seems a bit high',
 '@freewalletorg Thanks for fixing',
 '@freewalletorg Please unlock my account',
 '@AstroJordy <U+0001F923><U+0001F923>',
 '@freewalletorg Any crypto wallet that won't give you your private keys should be avoided a
```

```
tweet2 = ".join(tweet1)
tweet2
```

```
'@kunalb11 I am an alien@ID_AA_Carmack Ray tracing on Cyberpunk with HDR is next-level. Have you tried it?@joerogan @Spotify
Great interview!@gtera27 Doge is underestimated@teslacr Congratulations Tesla China for amazing execution last year Now on to the
o the next for even more!!Happy New Year of the Ox! https://t.co/9WFKMYu2ojFrodo was the underdog,\nAll thought he would fai
l,\nHimself most of all. https://t.co/zGxJFDzzrM@Owensparks_ @flcnhvy @anonyx10 Haha thanks :)@flcnhvy @anonyx10 Indeed! Twee
ts definitely do not represent real-world time allocation.The most entertaining outcome is the most likely@GiveDirectly Just
sent someJust agree to do Clubhouse with @kanyewesthttps://t.co/3rWE9uHSTS@geoffkeighley @UnrealEngine It's getting realBoug
ht some Dogecoin for lil X, so he can be a toddler hodler@JoshManMode He definitely has issues, but the sentencing seems a bi
t high@freewalletorg Thanks for fixing@freewalletorg Please unlock my account@AstroJordy <U+0001F923><U+0001F923>This is true
power haha https://t.co/Fc9uhQsd70@freewalletorg Any crypto wallet that won't give you your private keys should be avoided a
```

```
nopunctweet = tweet2.translate(str.maketrans("", string.punctuation))
nopuncttweet
```

```
'kunalb11 I am an alienIDAACarmack Ray tracing on Cyberpunk with HDR is nextlevel Have you tried itjoerogan Spotify G
erviewgtera27 Doge is underestimatedteslacr Congratulations Tesla China for amazing execution last year Now on to the
r even moreHappy New Year of the Ox httpstco9WFKMYu2ojFrodo was the underdog,\nAll thought he would fail,\nHimself mos
httpstcozGxJFDzzrMOwensparks flcnhvy anonyx10 Haha thanks flcnhvy anonyx10 Indeed Tweets definitely do not represent
d time allocationThe most entertaining outcome is the most likelyGiveDirectly Just sent someJust agree to do Clubhous
anyewesthttpstco3rWE9uHSTSgeoffkeighley UnrealEngine It's getting realBought some Dogecoin for lil X so he can be a
hodlerJoshManMode He definitely has issues but the sentencing seems a bit highfreewalletorg Thanks for fixingfreewall
ease unlock my accountAstroJordy U0001F923U0001F923This is true power haha httpstcoFc9uhQsd70freewalletorg Any cryptc
that won't give you your private keys should be avoided at all costsfreewalletorg Your app sucksRT SpaceX NASA has s
Falcon Heavy to launch the first two elements of the lunar Gateway together on one mission httpstco3pwtajtourville
L3 RationalEtienne Adamklotz Once we can predict cash flow reasonably well Starlink will IPORationalEtienne Adamklotz
k is a staggeringly difficult technical amp economic endeavor However if we don't httpstco9Z8Ac6skqxRationalEtienne A
SpaceX needs to pass through a deep chasm of negative cash flow over the next year or httpstco7J1c92hdjlIDAACarmack
cost per ton of carbon sequestered net of value of any product made \n\nMust be scalable to g httpstcoXMyI7qWSgwAda
t's meant to be the same price in all countries Only difference should be taxes amp shippingtobyliiiiiiiiiii This is
for Earth but there may be some ideas that apply to Mars tooCryptoShrikar CoinDesk Tesla DanZPalmer U0001F923U0001F92
team will manage the 100M carbon capture prize httpstcofSw5IanL0rSRuhle Everyone at Tesla receives stock \n\nMy comp
tockoptions which I do not take off the table \n\nThat's what you're missingBack to work I go CapybaraSurfer Mattl
8 Does seem a bit highhitsALLrisky Doge appears to be inflationary but is not meaningfully so fixed of coins per unit
```

```
toktweet = word_tokenize(nopunctweet)
print(toktweet[0:100])
```

[ 'kunalb11', 'I', 'an', 'alienIDAACarmack', 'Ray', 'tracing', 'on', 'Cyberpunk', 'with', 'HDR', 'is', 'nextlevel', 'Have', 'you', 'tried', 'itjoerogan', 'Spotify', 'Great', 'interviewgtera27', 'Doge', 'is', 'underestimatedteslacn', 'Congratulations', 'Tesla', 'China', 'for', 'amazing', 'execution', 'last', 'year', 'Now', 'on', 'to', 'the', 'next', 'for', 'even', 'moreHappy', 'New', 'Year', 'of', 'the', 'Ox', 'httpstcoz9WFKMYu2ojFrodo', 'was', 'the', 'underdoge', 'All', 'thought', 'he', 'would', 'fail', 'Himself', 'most', 'of', 'all', 'httpstcozGxJFDzrMOWenSparks', 'flcnhvy', 'anonyx10', 'Haha', 'thanks', 'flcnhvy', 'anonyx10', 'Indeed', 'Tweets', 'definitely', 'do', 'not', 'represent', 'realworld', 'time', 'allocationThe', 'most', 'entertaining', 'outcome', 'is', 'the', 'most', 'likelyGiveDirectly', 'Just', 'sent', 'someJust', 'agree', 'to', 'do', 'Clubhouse', 'with', 'ka

```
stopwords = stopwords.words('english')
```

stopwords

```
4]: ['I',  
      'me',  
      'my',  
      'myself',  
      'we',  
      'our',  
      'ours',  
      'ourselves',  
      'you',  
      "you're",  
      "you've",  
      "you'll",  
      "you'd",  
      'your',  
      'yours',  
      'yourself',  
      'yourselves',  
      'he',  
      'him',
```

```
nostoptokens = [word for word in toktweet if not word in stopwords]
```

```
print(nostoptokens[0:30])
```

```
['kunalb11', 'I', 'alienIDAACarmack', 'Ray', 'tracing', 'Cyberpunk', 'HDR', 'nextlevel', 'Have', 'tried', 'itjoerogan', 'S  
ify', 'Great', 'interviewgtera27', 'Doge', 'underestimatedteslacr', 'Congratulations', 'Tesla', 'China', 'amazing', 'executio  
n', 'last', 'year', 'Now', 'next', 'even', 'moreHappy', 'New', 'Year', 'Ox']
```

```
lowerwords = [Text.lower() for Text in nostoptokens]
```

```
print(lowerwords[0:16])
```

```
['kunalb11', 'i💎', 'alienidaacarmack', 'ray', 'tracing', 'cyberpunk', 'hdr', 'nextlevel', 'have', 'tried', 'itjoerogan', 'spot  
ify', 'great', 'interviewgtera27', 'doge', 'underestimatedteslacn']
```

```
ps = PorterStemmer()
```

```
stemmedtokens = [ps.stem(word) for word in lowerwords]
```

```
print(stemmedtokens[0:16])
```

---

```
['kunalb11', 'i', 'alienidaacarmack', 'ray', 'trace', 'cyberpunk', 'hdr', 'nextlevel', 'have', 'tri', 'itjoerogan', 'spotify', 'great', 'interviewgtera27', 'doge', 'underestimatedteslacn']
```

---

```
nlp = spacy.blank('en')
doc = nlp("".join(nostoptokens))
print(doc[0:16])
```

```
kunalb11IalienIDAACarmackRaytracingCyberpunkHDRnextlevelHavetrieditjoeroganSpotifyGreatinterviewgtera27Dogeunderestimatec
acnCongratulationsTeslaChinaamazingexecutionlastyearNownextevenmoreHappyNewYearOxhttpstco9WFKMYu2ojFrodounderdogeAllthoughti
dfailHimselfhttpstcozGxJFDzrMOwenSparksflcnhvyanonyx10Hahathanksflcnhvyanonyx10IndeedTweetsdefinitelyrepresentrealworldtim
ocationTheentertainingoutcomelikelyGiveDirectlyJustsentsomeJustagreeClubhousekanyewesthttpstco3rWE9uHSTsgeoffkeighleyUnrealI
neIt'sgettingrealBoughtDogeecoinlilXtoddlerhodlerJoshManModeHedefinitelyissuessentencingseemsbithighfreewalletorgThanksfixir
ewalletorgPleaseunlockaccountAstroJordyU0001F923U0001F923ThistruepowerhahahttpstcoFc9uhQsd70freewalletorgAnycryptowalletwon
veprivatekeysavoidedcostsfreewalletorgYourappsucksRTSpaceXNASAselectedFalconHeavyLaunchfirsttwoelementslunarGatewaytogethern
issionhttpstco3pwtajtourvilleYesBLKMDL3RationalEtienneAdamklotzOncepredictcashflowreasonablywellStarlinkIPORationalEtienne
klotzStarlinkstaggeringlydifficulttechnicalampeconomicendeavorHoweverndonhttpstco9Z8Ac6skqxRationalEtienneAdamklotzSpaceXne
assdeepchasmnegativecashflownextyearorhttpstco7J1c92hdjlIDAACarmackLowestcostpertontcarbonsequesterednetvalueproductmadeMus
labeledhttpstcoXmyI7qwSgwAdamklotzIt
```

```
lemma = [token.lemma_ for token in doc]
print(lemma[0:25])
```

```
['kunalb11', 'I', 'alien', 'IDAACarmack', 'Ray', 'trace', 'Cyberpunk', 'HDR', 'nextlevel', 'have', 'joerogan', '
otify', 'great', 'interview', 'gtera27', 'Doge', 'underestimate', 'teslacn', 'Congratulations', 'Tesla', 'China', 'ar
ing', 'execution', 'year', 'now']
```

---

```
vectorizer = CountVectorizer()
x = vectorizer.fit_transform(lemma)
```

```
print(vectorizer.vocabulary_)
```

```
{'kunalb11': 2672, 'alien': 229, 'idaacarmack': 2416, 'ray': 3580, 'trace': 4445, 'cyberpunk': 948, 'hdr': 16
evel': 3119, 'have': 1648, 'joerogan': 2594, 'spotify': 4068, 'great': 1588, 'interview': 2523, 'gtera27': 16
e': 1100, 'underestimate': 4613, 'teslacn': 4294, 'congratulations': 844, 'tesla': 4288, 'china': 726, 'amazi
n': 1303, 'year': 4877, 'now': 3158, 'happy': 1636, 'new': 3115, 'ox': 3274, 'httpstco9wfkmyu2oj': 1
o': 1473, 'underdoge': 4612, 'all': 232, 'think': 4365, 'fail': 1339, 'himself': 1698, 'httpstcozgxjfdzrm':
nsparks': 3271, 'flcnhvy': 1415, 'anonyx10': 281, 'haha': 1621, 'indeed': 2463, 'tweets': 4505, 'represent':
lworld': 3604, 'time': 4388, 'allocation': 234, 'the': 4339, 'entertaining': 1235, 'outcome': 3251, 'givedire
9, 'just': 2622, 'agree': 198, 'clubhouse': 768, 'kanyewest': 2631, 'httpstco3rwe9uhsts': 1780, 'geoffkeighle
'unrealengine': 4640, 'it': 2545, 'real': 3594, 'bought': 565, 'dogecoin': 1102, 'lil': 2749, 'toddler': 4412
r': 1708, 'joshmanmode': 2607, 'he': 1655, 'issue': 2543, 'sentencing': 3848, 'bit': 513, 'high': 1689, 'free
g': 1463, 'thank': 4331, 'fix': 1406, 'please': 3390, 'unlock': 4638, 'account': 150, 'astrojordy': 355, 'u00
1f923': 4576, 'this': 4367, 'true': 4475, 'power': 3432, 'httpstcofc9uhqsd7o': 1947, 'any': 290, 'crypto': 92
t': 4744, 'won': 4836, 'give': 1558, 'private': 3472, 'key': 2647, 'avoid': 413, 'cost': 885, 'your': 4888, '
'suck': 4172, 'rt': 3748, 'spacex': 4033, 'nasa': 3080, 'select': 3838, 'falcon': 1343, 'heavy': 1664, 'launc
element': 1196, 'lunar': 2828, 'gateway': 1522, 'mission': 2992, 'httpstco3pwt': 1778, 'ajtourville': 212, '
9, 'blkmdl3': 525, 'rationaletienne': 3577, 'adamklotz': 170, 'once': 3211, 'predict': 3447, 'cash': 668, 'fl
'starlink': 4092, 'ipo': 2533, 'staggeringly': 4082, 'difficult': 1056, 'technical': 4264, 'amp': 258, 'econo
1, 'endeavor': 1218, 'however': 1734, 'httpstco9z8ac6skqx': 1844, 'pass': 3310, 'deep': 999, 'chasm': 717, 'n
3102, 'httpstco7j1c92hdjl': 1820, 'lowest': 2812, 'ton': 4421, 'carbon': 654, 'sequester': 3855, 'net': 3107, '
t': 3488, 'make': 2853, 'must': 3067, 'scalable': 3798, 'httpstcoxmyi7qwsgw': 2332, 'mean': 2916, 'price': 34
```

```
print(vectorizer.get_feature_names()[50:80])
print(x.toarray()[50:80])
```

```

vectorizer = CountVectorizer(ngram_range=(1,3),analyzer='word',max_features=150)
x = vectorizer.fit_transform(tweet1)
print(vectorizer.get_feature_names())
print(x.toarray())

['0001f923', '0001f923 0001f923', 'about', 'actually', 'ajtourville', 'all', 'also', 'amp', 'an', 'and', 'are', 'as', 'at', 'austinbarnard45', 'be', 'best', 'better', 'but', 'by', 'can', 'cleantechnica', 'co', 'coming', 'could', 'crew', 'day', 'do', 'dragon', 'earth', 'engineers_feed', 'ercxspace', 'erdayastronaut', 'evafoxu', 'exactly', 'falcon', 'felixschlang', 'few', 'first', 'flcnhv', 'flight', 'for', 'from', 'get', 'go', 'good', 'great', 'haha', 'has', 'have', 'he', 'high', 'https', 'https co', 'id_aa_carmack', 'if', 'in', 'in the', 'is', 'is the', 'it', 'just', 'kristennetten', 'launch', 'lexfridman', 'like', 'long', 'lot', 'make', 'marcushousegame', 'maybe', 'me', 'mission', 'model', 'more', 'most', 'much', 'my', 'nasa', 'nasaspaceflight', 'need', 'new', 'next', 'no', 'not', 'now', 'of', 'of the', 'on', 'one', 'only', 'or', 'our', 'out', 'people', 'ppathole', 'pretty', 'probably', 'production', 'rationaletienne', 're', 'right', 'rt', 'rt spacex', 'should', 'should be', 'so', 'some', 'soon', 'space station', 'spacex', 'starship', 'still', 'sure', 'tesla', 'teslaownerss', 'teslarati', 'tesmanian com', 'than', 'that', 'the', 'there', 'thesheetztweetz', 'they', 'this', 'this is', 'time', 'to', 'to be', 'to the', 'tobyliiiiiiii', 'too', 'true', 'up', 'very', 'was', 'we', 'we re', 'well', 'what', 'wholemarsblog', 'will', 'will be', 'with', 'work', 'would', 'yeah', 'yes', 'you', 'your']
[[0 0 0 ... 0 0 0]
 [0 0 0 ... 0 1 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]]

```

```

vectorizer = TfidfVectorizer(ngram_range=(1,3),analyzer='word',max_features=150)
x = vectorizer.fit_transform(tweet1)
print(vectorizer.get_feature_names())
print(x.toarray())

['0001f525', '0001f5a4', '0001f923', '0001f923 0001f923', '20', '28delayslater', '60', 'about', 'absolutely', 'accurate', 'actually', 'after', 'ai', 'ajtourville', 'all', 'almost', 'also', 'although', 'am', 'amazing', 'amp', 'an', 'and', 'any', 'appreciated', 'archillect', 'are', 'as', 'astro_doug', 'astrobehnken', 'astrojordy', 'astronauts', 'at', 'at tesla', 'austinbarnard45', 'autopilot', 'back', 'bad', 'bar', 'be', 'be great', 'been', 'being', 'berlin', 'berniesanders', 'best', 'beta', 'better', 'between', 'big', 'bit', 'boca', 'bocachicagal', 'body', 'booster', 'boringcompany', 'both', 'build', 'but', 'but it', 'but we', 'by', 'c_bass3d', 'can', 'car', 'cargo', 'cars', 'caspar_stanley', 'cleantechnica', 'close', 'co', 'come', 'coming', 'coming soon', 'companies', 'confirmed', 'congratulations', 'cool', 'cost', 'could', 'crew', 'crew dragon', 'data', 'day', 'days', 'definitely', 'design', 'did', 'different', 'djsnm', 'do', 'doesnt', 'doge', 'doing', 'dojo', 'don', 'done', 'down', 'dragon', 'driving', 'droneship', 'early', 'earth', 'easy', 'electric', 'energy', 'engine', 'engineering', 'engineers_feed', 'engines', 'enn_nafnlaus', 'enough', 'ercxspace', 'erdayastronaut', 'erdayastronaut flcnhv', 'erdayastronaut rweb11742', 'erdayastronaut spacex', 'evafoxu', 'even', 'ever', 'exactly', 'fairing', 'falcon', 'falcon first', 'falcon first stage', 'far', 'fast', 'fe0f', 'feel', 'felixschlang', 'few', 'fire', 'first', 'first stage', 'flaps', 'flcnhv', 'flcnhv tesla', 'flight', 'flights', 'for', 'for sure', 'for the', 'from', 'from the', 'fsd', 'full', 'fully', 'fun', 'future', 'futurejurvetson', 'game', 'gerberkawasaki', 'get', 'giga', 'giga berlin', 'go', 'going', 'going to', 'good', 'great', 'grimezsz', 'had', 'haha', 'hard', 'hard to', 'has', 'have', 'he', 'heavy', 'here', 'high', 'hopefully', 'how', 'https', 'https co', 'human', 'icannot_enough', 'id_aa_carmack', 'if', 'if you', 'ihearttesla', 'important', 'improvements', 'in', 'in https', 'in https co', 'in the', 'into', 'is', 'is not', 'is the', 'it', 'it https', 'it https co', 'it is', 'it should', 'it was', 'it will', 'it will be', 'its', 'jack', 'janeideyeve', 'jgrano305', 'jgrano305 teslaratiteam', 'jgrano305 teslaratiteam teslarati', 'johnkrausphoto', 'johnnacridier1', 'josh_bickett', 'just', 'justpaulinelol', 'jwangark', 'know', 'kristennetten', 'la', 'land', 'landing', 'last', 'launch', 'legs', 'less', 'level', 'lexfridman', 'life', 'liftoff', 'like', 'likely', 'little', 'll', 'long', 'looks', 'lot', 'lot of', 'love', 'low', 'machine', 'made', 'major', 'make', 'makes', 'making', 'many', 'marcushousegame', 'mars', 'mass', 'may', 'maybe', 'me', 'michaelminalab', 'might', 'mission', 'mkbhd', 'model', 'money', 'month', 'months', 'more', 'most', 'most likely', 'mostly', 'much', 'my', 'nasa', 'nasaspaceflight', 'need', 'need to', 'nee

```

```

def cloud(wordcloud):
    plt.figure(figsize=(16,8))
    plt.imshow(wordcloud)

```

```
stopwords = STOPWORDS
```



```
stopwords.add('ppathole')
```

```
wordcloud = WordCloud(height=2000,width=3000,background_color='black',max_words=150)
```

```
plot_cloud(wordcloud)
```

