In [ ]:

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount
("/content/drive", force_remount=True).
```

In [ ]:

```
!pip install scikit-learn
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/publi
c/simple/
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.8/dist-packages (1.
0.2)
Requirement already satisfied: scipy>=1.1.0 in /usr/local/lib/python3.8/dist-packages (fr
om scikit-learn) (1.7.3)
Requirement already satisfied: numpy>=1.14.6 in /usr/local/lib/python3.8/dist-packages (f
rom scikit-learn) (1.21.6)
Requirement already satisfied: joblib>=0.11 in /usr/local/lib/python3.8/dist-packages (fr
om scikit-learn) (1.2.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.8/dist-pack
ages (from scikit-learn) (3.1.0)
```

In [ ]:

```
from sklearn.feature_extraction.text import TfidfVectorizer
import glob
import os
```

In [ ]:

```
data_dir = '/content/drive/MyDrive/tweet_files'
file_list = glob.glob(os.path.join(data_dir, '*.txt'))
```

In [ ]:

```
file_subset = file_list[0:1000]
```

In [ ]:

```
# Read the text from each file and store it in a list
texts = []
for filename in file_subset:
    with open(os.path.join('/content/drive/MyDrive/tweet_files', filename), 'r') as f:
        texts.append(f.read())
```

In [ ]:

```

```

In [ ]:

```
#print(file_subset[0:5])
texts[0:2]
```

Out[ ]:

```
['"Hey guys have you heard about this if didn\'t so must focus on it #JeepGrandCherokee #
OIIIIIIIO \n\'Simranshukla355\'',
 "'All the models are made in our own country , what a solid achievement @JeepIndia #Jeep
GrandCherokee \n'iAdityaSaxena_'"]
```

In [ ]:

```
vectorizer = TfidfVectorizer()
```

In [ ]:

```
vectorizer = TfidfVectorizer(stop_words=None)
```

In [ ]:

```
vectorizer = TfidfVectorizer(min_df=0.005)
```

In [ ]:

```
vectors = vectorizer.fit_transform(texts)
```

In [ ]:

```
from sklearn.metrics.pairwise import cosine_similarity
```

In [ ]:

```
similarities = cosine_similarity(vectors)
```

In [ ]:

```
print((similarities[26][97]))
```

```
0.0
```

In [ ]:

```
from sklearn.feature_extraction.text import CountVectorizer
count_vectorizer =CountVectorizer(stop_words='english', min_df=0.00005)
corpus2 = count_vectorizer.fit_transform(texts)
print(count_vectorizer.get_feature_names())
```

```
['000', '00iakki', '060', '0l', '11', '18', '1899netflix', '18th', '19', '1993', '1995jay
sharma', '1999indian', '1day', '1slamic', '2014', '2015', '2016', '2017', '2018', '2018b'
, '2020', '2021', '2022', '2023', '20x9', '22', '26', '270', '28500', '2amgossipp', '30',
'30500', '32', '3minutes', '40', '400', '400nm', '45abhumanyu', '4k', '4x4', '4xe', '50',
'5th', '60th', '63', '76', '77', '7897_verma', '99', '9mugdha', '_1butterflysly_', '_1sak
shi', '____meghna____', '__anniieee_____', '__bhoomika___', '__khanafroz', '__naazu__', '
__sapna_singh__', '__vishal03', '_aditi1', '_am_hr', '_badvibesx_', '_dxrathod', '_gaurav
k14', '_marvaaadi', '_nishi_mishra', '_okayybye', '_prakash1_', '_pspkfanboy_', '_rajkuma
r14', '_rudra1', '_secretbae', '_silpi_', '_swapnika', '_tweetsays', 'aa', 'aa_gye_aap',
'aabheer_9', 'aachal_joshi_', 'aadavan_10', 'aadhith_11', 'aagayi', 'aakashpal8151', 'aak
ashtrend', 'aamirs_fauji', 'aamirs_medhansh', 'aap', 'aapkepaap', 'aapko', 'aarah', 'aara
vsinghji', 'aathmikara_', 'aaur', 'aayegi', 'aayesha686', 'ab_d_17_', 'abbasansarid1', 'a
bdb_17', 'abeylo_e', 'abhi', 'abhi__1122', 'abhipsaa98', 'abhishekpathakk', 'ability', 'a
ble', 'absolute', 'absolutely', 'absurd', 'accept', 'accessible', 'acclaimed', 'according
', 'accumulate', 'achcha', 'achhi', 'achi', 'achieve', 'achievement', 'acting', 'actio',
'action', 'activities', 'activity', 'actor', 'actually', 'adani', 'adaniwilmar', 'adarshk
34919993', 'added', 'addition', 'aditi_yrr', 'adityajhaajphe2', 'admiration', 'admire', '
admiring', 'advance', 'advanced', 'adventure', 'adventures', 'advocat68155696', 'affordab
le', 'afraid', 'afshachoudhry4', 'aftab', 'agni_astra9', 'ago', 'ahead', 'ahmadkhank_', '
air', 'airajoseph11119', 'ajajay22', 'ajay', 'ajay778118', 'ajaydevgn', 'ajnabiprani', 'a
k_shita07', 'akf_kanha', 'akhil_prabha_01', 'akhil_yadav00', 'akki2g', 'aksar', 'akshara_
345', 'akshay', 'akshaye', 'akshaysraju', 'alcoholicshe', 'aliabhattachar7', 'alluarjun12
3_4', 'alot', 'alwayskotesh', 'ama', 'aman___rahman', 'aman_xdd', 'amansinghh72', 'amazeb
alls', 'amazi', 'amazing', 'amazingly', 'amazonfinds', 'america', 'american', 'amezing',
'amita_chaubey', 'amitbanat', 'amp', 'amrit_xdd', 'amritmohotsav', 'anamii_kaa', 'aneetar
oy2', 'angellradhika', 'angle', 'anika_jammu', 'anime', 'animedub', 'animefans', 'animesu
b', 'anjali_00001', 'anjali_v1', 'ankitababele', 'ankitabendre455', 'ankitsh15148708', 'a
nkurku53192153', 'ankushk36837793', 'ankuu_03', 'anmolsebi', 'announce', 'announced', 'an
othing', 'anpadh', 'ans', 'ansh__1144', 'anshika5780', 'anshitahu', 'anshu_v_sharma', 'an
shudolll', 'anshupandit8818', 'anthem', 'anticipated', 'anuj_thakur08', 'anupama6641', 'a
nupriyaojha68', 'anupriyartf', 'anuroy67', 'anusingh348', 'anuuux2s11', 'anvithakur54', '
anymore', 'aparna0055', 'apkepyaarmein', 'apne', 'app', 'apply', 'appointments', 'appreac
iting', 'appreciate', 'appreciated', 'approach', 'approaches', 'ar860_', 'aradhya2k', 'ar
adhya_jii', 'arautopartszone', 'area', 'areas', 'aren', 'arey', 'arid', 'arifari38568863'
, 'aritrasen45', 'arkarfly', 'army', 'arrey_yr', 'arrival', 'arrived', 'arshi7654574', 'a
rshi__tweets', 'arti300004k', 'arun4digit', 'arunsin011', 'asap', 'ashokku2000', 'asian08
7', 'ask', 'asked', 'assemb', 'assembled', 'assistance', 'assistant', 'assurer', 'atharv6
398', 'atheist_kanchaa', 'atlanta', 'attend', 'attract', 'attractive', 'attracts', 'au',
```

'audience', 'audiences', 'aur', 'ausveng', 'auto', 'automobile', 'automobiles', 'autos',
'available', 'avantika_ji', 'avantikathaku1', 'avi', 'avi7619', 'avinashsarma0', 'avneetk
aur02', 'avoid', 'avuthundi', 'awaited', 'awaiting', 'awarded', 'away', 'awe', 'awesome',
'ayesha_qureshi5', 'ayurveda', 'azadkumar94', 'baar', 'baba_jomandha', 'babug42', 'bad',
'badly', 'bae_wafa_', 'bahar', 'bahut', 'bakiye_1', 'ban', 'bang', 'banihal', 'bas', 'bas
e', 'based', 'basically', 'bat', 'bawaalxd', 'beast', 'beasts', 'beautiful', 'beauty', 'b
eek', 'beerus39887593', 'beg', 'begin', 'begins', 'behad', 'behtarin', 'being_krishn', 'b
eing_mohdasad', 'beingsam45', 'beingumairr', 'believe', 'bell', 'benchmark', 'best', 'bes
tand', 'betsythomasarun', 'better', 'betterment', 'bhaktchacha', 'bharatkumar4466', 'bhas
hkar_verma_', 'bhashker_ji', 'bhavani0_', 'bhiyo', 'bhp', 'bhule', 'bid', 'biden', 'big',
'bigger', 'biggest', 'bilaal_raza_', 'billy', 'bina', 'bio', 'birthday', 'biswajit_offf',
'bits', 'bittu_134', 'black', 'blackpink', 'blast', 'blind_patelbabu', 'blindly', 'blockb
uster', 'blue', 'bob', 'bobbybhagath', 'bobbygadu_local', 'bobyscaptain', 'bokings', 'bol
', 'book', 'booked', 'booking', 'bookingopen', 'bookings', 'books', 'bouchara', 'bought',
'box', 'brand', 'brands', 'bravocdjr', 'break', 'breaking', 'brian', 'brighten', 'brillia
nt', 'bring', 'bringing', 'brother', 'brotherhood', 'brought', 'buddy', 'built', 'bulid',
'bumper', 'bury', 'bus', 'busan', 'business', 'businessnews', 'buttons', 'buy', 'buying',
'buzz', 'ca', 'cabin', 'caledonia', 'called', 'came', 'cancel', 'cancelled', 'capability'
, 'capable', 'captivating', 'capture', 'car', 'carandbike', 'card', 'cards', 'career', 'c
arlelo', 'carleloindia', 'cars', 'carsofnewwest', 'carsofnewwestminster', 'carwale', 'cas
e', 'casereopenstomorrow', 'cast', 'cat', 'catch', 'caught', 'cbi', 'cbse_2020', 'cbt', '
cdjr', 'celebrating', 'central', 'centre', 'centres', 'ceo', 'cgcapitalnews1', 'chahta',
'chain', 'chala', 'chance', 'change', 'changing', 'channel', 'chapter', 'char', 'characte
r', 'characters', 'charging', 'charlie79187116', 'che', 'cheap', 'check', 'checkbrand2',
'checking', 'cheroke', 'cherokee', 'chesa', 'chetnapn64', 'chicken', 'child', 'childhood'
, 'children', 'china', 'chinki_here', 'chintusstar', 'chinuu356', 'chopramonika22', 'chri
s', 'christmas', 'chrome', 'chrysler', 'chuki', 'chunky', 'cinema', 'cinemas', 'circle',
'citizens', 'claim', 'class', 'classical', 'clever', 'clickhere', 'closed', 'cnbctv18news
', 'coaching', 'coalition', 'coldlikechill', 'collection', 'colour', 'come', 'comes', 'co
mfort', 'comfortable', 'coming', 'commence', 'commenced', 'commendable', 'comment', 'comm
enting', 'commits', 'compan', 'company', 'compass', 'complete', 'completely', 'concept',
'confess', 'confidence', 'conform', 'congress', 'connected', 'connectivity', 'consistentl
y', 'console', 'contact', 'content', 'continue', 'continued', 'control', 'controversial',
'convenience', 'cool', 'cool_bul_pandey', 'cool_chimtu', 'copper_fitting', 'copy', 'core'
, 'corey', 'cosmocb56988306', 'couldn', 'country', 'course', 'court', 'cousins', 'crafted
', 'craziestlaziest', 'crazy', 'crazyboy15_', 'crazyness', 'create', 'created', 'creating
', 'crew', 'cricmaulik45', 'crime', 'crimes', 'critics', 'crowd', 'crypto', 'csnt', 'cup'
, 'curaj', 'curiosities', 'curiosity', 'curious', 'custome', 'customer', 'customers', 'cu
te_6priya', 'cute_rashmika73', 'cuterina7', 'cwlaunch', 'cwnews', 'cybersecurity', 'dam',
'damn', 'damnn', 'damon_t_vd', 'danish1240', 'danish2h', 'das_bhuban0', 'das_jeniva', 'da
shboard', 'dashing', 'dat', 'date', 'day', 'days', 'deal', 'dealerships', 'deals', 'dear'
, 'debut', 'december', 'decent', 'decent_bhumi', 'deep', 'deepakk39945280', 'deepakparab2
40', 'deepakuswaha_', 'deepblue18_', 'deepu71535611', 'definite', 'definitel', 'definitel
y', 'dekh', 'dekhen', 'dekhna', 'dekhne', 'dekho', 'delay', 'delhiwalaboy', 'delighted',
'delip274586572', 'delivered', 'deliverie', 'deliveries', 'delivering', 'delivery', 'dema
nd', 'demonstrates', 'derek__17', 'desaipriya12', 'deserves', 'design', 'desperately', 'd
espite', 'deta', 'details', 'dev', 'dev_akash___', 'devgan', 'devgn', 'devgnkadeewana', '
devgon', 'devika__121', 'devoteerk', 'devrajs92118148', 'dhaara233', 'dhamal', 'dhanes_sh
arma', 'dharamveer546', 'dharamvir', 'dharmesh_jitiya', 'dher', 'dhfm__chintu', 'dhfpbnan
i', 'dhonistan10', 'dhrishyam', 'dickscottauto', 'did', 'didn', 'difference', 'different'
, 'digital', 'dil', 'dilseadian', 'dimpal_girl_', 'dino', 'dirhsyam', 'disappearan', 'dis
appearance', 'disappoints', 'discuss', 'disease', 'dishonesty', 'divided', 'diya', 'diyaa
snx313', 'diyamondal03', 'dkhna', 'dkumar_555', 'dodg', 'dodge', 'doing', 'doli_pandey',
'dollychaturve5', 'dominate', 'don', 'donor', 'dor', 'dosti', 'doston', 'doubt', 'drama',
'dream', 'dressed', 'drishyam', 'drishyam2', 'drishyam2thisfriday', 'drive', 'driven', 'd
river', 'drivespark', 'driving', 'dropping', 'dsai_555', 'dub', 'duggunatkhat', 'dulalroy
_', 'duper', 'dushman', 'dutta_tupai1', 'dutybeyondborders', 'eager', 'eagerly', 'earned'
, 'earth', 'earthworm', 'easily', 'easy', 'edinburg', 'edition', 'education', 'eek', 'eff
ect', 'efficient', 'effort', 'ek', 'elder', 'elegance_45', 'elina385800', 'elonmusk', 'em
a_n_shah', 'emotion', 'emotions', 'emperor_offl', 'en', 'end', 'ends', 'enforced', 'engag
ement', 'engine', 'engrmht', 'enhancement', 'enjoy', 'enjoyable', 'enrollment', 'entertai
nment', 'enthralling', 'entire', 'entrance', 'entrepreneurs', 'entry', 'epf', 'epfo', 'ep
ic', 'eps', 'equally', 'equipped', 'era', 'escape', 'eternal_shivam', 'etsy', 'event', 'e
verybody', 'everytime', 'evidence', 'evolution', 'ex', 'exam', 'example', 'exc', 'excelle
nt', 'excited', 'excitement', 'excitiment', 'exciting', 'excitment', 'exhaust', 'exist',
'exited', 'exiting', 'exo', 'expect', 'expectations', 'expected', 'expensive', 'experienc
e', 'experts', 'explain', 'explore', 'express', 'extensive', 'extremely', 'eyes', 'fa', '
fabfeetfest', 'fabulous', 'face', 'faced', 'fact', 'factoryrepro', 'factoryreproductions'
, 'facts', 'fail', 'fairytal_girl', 'fam', 'family', 'famous', 'fan', 'fans', 'fantastic'
, 'far', 'farman0190', 'faruqazam___', 'fascinating', 'fast', 'fatima49721806', 'favor',
'favorite', 'favourite', 'fca', 'fdfs', 'feat', 'feature', 'features', 'featuring', 'febr
uary', 'feel', 'feeling', 'felt', 'fi', 'fifth', 'file', 'filling', 'film', 'filmed', 'fi

ims', 'filter', 'finallllly', 'finally', 'finger', 'finished', 'finnaly', 'fixed', 'flags
hip', 'fly', 'fo', 'focus', 'focused', 'folks', 'follow', 'force', 'forces', 'forefathers
', 'forget', 'forming', 'fortune', 'forvall', 'forward', 'fot', 'fours', 'fourth', 'fr',
'fred_eazy21', 'free', 'freinds', 'friday', 'friend', 'friends', 'frnd', 'fruit', 'fully'
, 'fun', 'furthermore', 'future', 'futuristic', 'gaga', 'gaitonde', 'game', 'games', 'gan
apathyjrfc', 'gandhi', 'gap', 'gatecrashing', 'gather', 'gauri3568', 'gautammazumdar0', '
gauthmi2', 'gaya', 'gays', 'ge', 'gen', 'generation', 'genuinely', 'gerat', 'getting', 'g
har', 'ghavanee', 'gill15987', 'gillnancy518', 'girl', 'girls', 'girlwidimple', 'gives',
'glad', 'glimpses', 'global', 'glorifying', 'gloss', 'gm', 'goa', 'goatxberg', 'god', 'go
ing', 'gold', 'gon', 'gone', 'gonna', 'goo', 'good', 'goosebumps', 'gorgeous', 'gorpooja1
63', 'got', 'government', 'gr', 'grab', 'grand', 'grandcherokee', 'grander', 'gratitude',
'great', 'greatest', 'grew', 'grey017_', 'griffinshubwi', 'gril', 'group', 'gu', 'guangzh
ou', 'gud_time_', 'gudiya101', 'gudiya_90', 'guidence', 'gujarat', 'gulafshaba', 'gulvind
ars', 'gun__gun_', 'gupta__pk', 'guru12355757', 'guy', 'guys', 'guyz', 'gya', 'gyus', 'ha
i', 'hain', 'hall', 'halls', 'halpatikishan46', 'hands', 'happened', 'happiness', 'happy'
, 'haraane', 'hardhtony', 'harsh_kumar56', 'haven', 'hawk', 'hayes', 'head', 'hear', 'hea
rd', 'heart', 'hearts', 'heat', 'heavily', 'hell_e_na', 'hella', 'hello', 'helped', 'help
ers_auto', 'helpf', 'hemandra120', 'hereisharsh', 'hero', 'hey', 'heyy_aditya_', 'heyypih
u', 'hi', 'high', 'higher', 'highly', 'highways', 'hilly', 'himanshisa', 'hindu', 'hindu_
boy77', 'hindu_sharma_', 'hit', 'ho', 'hoga', 'hoge', 'hokar', 'hold', 'holiday', 'hollyw
ood', 'home', 'hone', 'hook', 'hoon', 'hope', 'hopes', 'host', 'hosting', 'hot', 'hounded
', 'house', 'hsr_001_', 'httpsstrangerr', 'hu', 'huge', 'hugely', 'humes_cjdr', 'hun', 'h
unted', 'hurdle', 'hurray', 'hurry', 'husband', 'hy_m_new', 'hyderabad', 'hype', 'i_nisha
mis', 'i_pooja2', 'i_sher0', 'iadityasaxena_', 'iakashteja', 'iam', 'iam_darlingfan', 'ia
maditya_____', 'iamdhruv45', 'iamg2_0', 'iamraj6393', 'iamsunshiene', 'ibeingrashi1', 'ic
handanmaurya', 'icon', 'iconic', 'iconicraju', 'idhonifan', 'idlymeenkolamb', 'ig', 'igia
aviation', 'ignore', 'igolu97', 'im_anjuu', 'im_bablu01', 'im_kishore0', 'im_raghib1', 'i
m_reyansh', 'imishaambani32', 'immediate', 'immediately', 'immensely', 'imnamiita', 'impa
lak___', 'imparting', 'important', 'imported', 'impressed', 'imrajchoudhary7', 'imsuhail9
99', 'imvadapav', 'imvanu', 'including', 'increased', 'increasing', 'incredible', 'india'
, 'indian', 'indianarmy', 'indianarmypeoplesarmy', 'indianautos', 'indians', 'indmahi07',
'industry', 'info', 'inforezin', 'information', 'initiative', 'inni', 'innocent', 'innoce
ntaditi_', 'insane_mind12', 'inserts', 'inside', 'insists', 'inspiration', 'installed', '
installment', 'instant', 'instantrecruit_', 'int', 'inten', 'interesting', 'interior', 'i
ntriguing', 'introduce', 'introduces', 'introduction', 'introductory', 'investigating', '
ipalvi87', 'ipriya9696', 'iqrakhan657', 'iranjitnayak', 'iron4dome', 'isandeepdubey', 'is
e', 'isha__143', 'isha___980', 'ishaaxxa31912', 'ishagun1', 'ishika4_', 'iske', 'islam',
'isne', 'isoni740', 'iss', 'isse', 'isumitsingh49', 'isupportgemsofbollywood', 'it_pragya
', 'it_pritii', 'it_pujaa', 'its_adhrit', 'its_indrani1', 'its_khushi02', 'its_komal___',
'itsiopr', 'itsme_zoya', 'itsmekrishna__', 'itsmy_accountt', 'itsravitiwari7', 'itsrolexs
ir', 'itswaseem32', 'itsyoursg', 'itszapyi', 'itz_ivaan', 'itz_kavii', 'itz_maurya14', 'i
tz_me_kairav_', 'itz_renu', 'itzlaksita78950', 'itzz_mamta', 'itzz_monika', 'itzz_riddhii
', 'itzzz_jiya', 'ivishaltaneja', 'jaa', 'jaanviie', 'jaao', 'jahnviiiii62014', 'jainshar
ma6', 'jaisakrity', 'jaldi', 'janviji123', 'jao', 'jau', 'jaw', 'jaybuniverse2', 'jbdrst'
, 'jeap', 'jeep', 'jeeper', 'jeepers', 'jeepgladiator', 'jeepgram', 'jeepgrandcherokee',
'jeepindia', 'jeeponlyvegas', 'jeeprenegade', 'jeeprubicon', 'jeeps', 'jems83473587', 'je
ssicaaaahu', 'jessiemeranaam', 'jhaj42788657', 'jimmysoni89', 'jingle', 'jittujitendra45'
, 'job', 'jobs', 'jobseekerssa', 'joebiden', 'joint', 'jointheshahid', 'jonsnow1726', 'jo
urney', 'joyland', 'jp17', 'jpparihar4242', 'jrfc_vijay', 'jshubhangini5', 'juliesh563560
61', 'just', 'just_a_bird_1', 'jyada', 'jyotipa95799272', 'k9monish', 'ka', 'kaam', 'kabi
r', 'kailash64341035', 'kal', 'kamalpuroheet', 'kanakanubhav', 'kanathakur8077', 'kanchan
sales', 'kanikabandyopa2', 'kantara', 'kar', 'karan_e17', 'karanpa73498181', 'karanpatel2
6395', 'kardo', 'kare', 'karlo', 'karne', 'karo', 'karthikdarling_', 'kavita95881k', 'ke'
, 'kehndihoonsi', 'ket32121354', 'khanna', 'khasimvali9948', 'khud', 'khud_ki_fav', 'khul
', 'khush', 'khushal8001', 'khushal_rosh', 'ki', 'kill', 'kim_puja7', 'kind', 'kirana_rav
', 'kirtisi95', 'kittujadon1', 'kiye', 'klassy_womanya', 'knfilters', 'know', 'knowing',
'known', 'ko', 'koi', 'kosam', 'koshal__1', 'kousik_01', 'koyal_pandey', 'kr', 'krishna__
242', 'krishnauma245', 'krishtweetss', 'krithishetty345', 'kritika_rana2', 'kudos', 'kuma
r', 'kumar_rohan209', 'kumarbharath98', 'kumarmangat', 'kunalyadavhard1', 'kundrakaran23'
, 'kya', 'kyayaarshubhi', 'kyoyaa_1', 'la', 'lack', 'lailasi90623187', 'lakh', 'lakhans_i
ngh', 'lakhs', 'lalitadutta20', 'lalitak95592665', 'lalitakumari_', 'lame', 'land', 'lang
uage', 'languages', 'late', 'latest', 'launch', 'launchalert', 'launched', 'launches', 'l
aunching', 'laxmipriya34', 'layered', 'le', 'lead', 'leaders', 'leather', 'leaves', 'left
', 'legacy', 'legal_dealer_', 'legendary', 'lekar', 'leke', 'lekin', 'lemarque_1', 'lengt
hy', 'leomess_fc', 'lesson', 'let', 'lets', 'level', 'levels', 'lexihoward_x', 'liberally
', 'life', 'lifestyle', 'lifts', 'like', 'liked', 'liking', 'lilyroy123456', 'limited', '
line', 'link', 'list', 'listening', 'literally', 'live_mahima', 'lives', 'living', 'liye'
, 'll', 'lo', 'loaded', 'loading', 'loc', 'locally', 'locations', 'log', 'lokeshbangaram'
, 'long', 'look', 'looked', 'looking', 'lookout', 'looks', 'lost', 'lost_kanya', 'lot', '
lots', 'love', 'loved', 'lovejihaad', 'loves', 'loving', 'lucifer_the_m', 'luckyjunwar',
'luxurious', 'luxury', 'm0gll', 'ma', 'maanvi__x', 'mad', 'madhu02002', 'madhuyaar_', 'ma
dprince_05', 'magnificent', 'mahendr30662104', 'mahesh', 'mahesh_402', 'mahi567888', 'mah

imagujarat_', 'main', 'majni_ruma', 'majority', 'make', 'makers', 'making', 'malayalam', 'malayalis', 'malik_arshad_', 'mall', 'mamtabanjara1', 'man', 'manibhadraf', 'manisha_si77', 'manishar730', 'manishyadav1902', 'manjul_k1', 'mansa_0001', 'manshi1100', 'manshi_life', 'mansi1745', 'mansive68884732', 'mansivini1', 'manufactured', 'manvendra557', 'mark', 'market', 'marketing', 'markets', 'marque', 'marriedatfirstsight', 'mask__7', 'masterpiece', 'matter', 'mauka', 'maulikvadariya', 'mayur04899799', 'mayuripratap', 'mcha', 'md', 'me__jitendra', 'me__ritika', 'medicalpipelin2', 'meena99728364', 'mein', 'melodicdaksh', 'melt', 'member', 'memekidiwani_', 'memorable', 'memories', 'mention', 'meowwgirl_', 'meraabdulaisanahihai', 'mere', 'meri', 'meridian', 'metra', 'microways0903', 'mike', 'mileage', 'milti', 'mind', 'minsuro', 'mintyprabha', 'misha130_', 'mishka050', 'mishragudiya0', 'mishtiii05', 'miss', 'miss_riyaaa', 'missayeshaptel0', 'missed', 'missile', 'mistery', 'mitalithakur_', 'mo', 'mode', 'model', 'models', 'modern', 'modi', 'mohali', 'mohamed92171752', 'mohammadhaish', 'mohan1s2', 'mohanlal', 'mohdali010101', 'mohddanish67', 'mohi19love', 'mohini786', 'mohit89__', 'moli_awasthi0', 'moment', 'moments', 'mona79782', 'money', 'monikaa150', 'monikahoon', 'monikavarma2929', 'monimeranaam', 'monstrous', 'month', 'monutweets_', 'moon12435677', 'mopar', 'morethan_yours', 'morning', 'mother', 'motoroctane', 'motors', 'mov', 'movementm14', 'movi', 'movie', 'moviepratap', 'movies', 'moving', 'mrajput222', 'mrperfect175', 'mrrancho02', 'msrp', 'mu', 'muditsharma46', 'mukesha2024', 'mumbai', 'munna_bhaiya12', 'murderer', 'muskan18586669', 'muskan740', 'muskansaifi21', 'muslims', 'muthumuthura', 'myself_babu1', 'mystery', 'mysticaldimpl3', 'n18', 'n_shaikh77', 'na', 'naaziya____', 'nagar', 'nagarjunanene', 'nahaanejanahane', 'nahi', 'naina96401449', 'najar', 'najay', 'nall', 'nalso', 'nameiselsa', 'nameplate', 'nancygill89', 'nand', 'nandini1065', 'nandroid', 'nare', 'naren', 'narendrpsingh', 'nasty', 'national', 'nationalnaturopathyday', 'nationalpressday', 'natkhatsujal_', 'nato', 'navneet_yadav1', 'navyaa_a0315', 'navyadixit11', 'navyami91200', 'nayi', 'nb', 'nbecause', 'nbest', 'nbook', 'nbookings', 'nbrotherhood', 'nbsf', 'ncan', 'ncase', 'ncheck', 'ncherokee', 'ncompany', 'nd', 'ndefinitely', 'ndnt', 'ndon', 'ne', 'near', 'nearest', 'necessity', 'need', 'needs', 'neeraj1s1', 'neerajbangaa', 'neha_am_22', 'nehaaa_____', 'nehasharma8409', 'nehasi7376', 'netflix', 'neutral', 'neveryone', 'new', 'newgrandcherokee', 'news', 'newsbytesapp', 'newz', 'nfamily', 'nfeel', 'nfilm', 'nfirst', 'nfits', 'nfor', 'nget', 'nglad', 'ngo', 'ngoing', 'ngood', 'nguy', 'nguys', 'nhappiness', 'nhey', 'nhi', 'nhyped', 'ni', 'nice', 'nidhii_69', 'nidhikhemka6', 'nihal_thakur77', 'nihalrp_', 'nikhil_679', 'nikita72805521', 'nikki_000_', 'nin', 'nincredible', 'nirdosh', 'nis', 'nishantt29', 'nishatiwari__', 'nisne', 'nit', 'niteshkumar760', 'nitinmaratha90', 'nits', 'nitujha74589439', 'njustice', 'nkaise', 'nkyunki', 'nlook', 'nlooking', 'nloved', 'nm', 'nmuslims', 'nmust', 'nmy', 'nnew', 'nnow', 'nobitakisizuka', 'non', 'north', 'nostalgic_aatma', 'notch', 'notes', 'notice', 'nov', 'november', 'novemberbookings', 'npart2', 'npopcorn', 'npre', 'npure', 'nreally', 'nsafest', 'nsafety', 'nsal', 'nsheikh', 'nso', 'nstop', 'nsuper', 'nthank', 'nthat', 'nthe', 'nthis', 'nthrill', 'nthrilled', 'ntickets', 'nto', 'ntreak', 'nu', 'nukes', 'number', 'nupto', 'nust', 'nvery', 'nwaiting', 'nwatch', 'nwe', 'nwhat', 'nwill', 'nyou', 'nyour', 'nzvindonprime', 'obsessed', 'oewheelsllc', 'offer', 'offers', 'office', 'officer', 'ofl_sayeesha', 'og', 'ohh', 'oiiiiiiio', 'ok', 'oka', 'okayy', 'old', 'omg', 'ones', 'onl', 'online', 'open', 'opened', 'opening', 'opens', 'opinion', 'opportunity', 'order', 'organinsed', 'origin', 'oscar', 'ott_army', 'outside', 'outstanding', 'owner', 'owsome', 'oye_prishu', 'p2', 'pa75942333', 'pa_nkaj_tiwari', 'pack', 'pak', 'paki', 'pakistan', 'pakistani', 'pand201', 'pankajb_ana', 'pankajruhela109', 'pansare', 'pappuram_jani_0', 'par', 'parag_ki_friend', 'paramount', 'paranift', 'pariver143', 'parking', 'parkinglot', 'participate', 'particular', 'parul2fb', 'pasand', 'passenger', 'passengers', 'pateldepika00', 'patrickpinckney', 'pavmohanan44', 'payal_ll', 'payne', 'payne_edinburg', 'pcpjewellers', 'peak', 'peeche', 'peer', 'pen', 'pendel', 'pension', 'people', 'perfect', 'performance', 'personal', 'petrol', 'pgme', 'phenomenal', 'phir', 'phlegmaticsamar', 'photo', 'physics', 'piano', 'pick', 'picks', 'picture', 'pictures', 'piddiriddhi', 'piece', 'pikachu090909', 'pintuu0', 'pissful', 'plan', 'planing', 'play', 'played', 'plenty', 'plush', 'pm', 'pmmodi', 'podiyadav', 'pokedex603', 'poland', 'police', 'pondolem', 'pooja8120', 'poojavishnoi291', 'poojesh01', 'pools', 'poonam', 'popular', 'positive', 'possession', 'possible', 'post', 'postcard', 'postcards', 'poster', 'potentially', 'power', 'powered', 'powerful', 'ppf', 'pr', 'pragati', 'pragati0907', 'pragyahun', 'pragyamishra854', 'praise', 'praising', 'prakash_raj00', 'prakash_sn2', 'prashan5051', 'prashant29__', 'prashanth1394', 'praveen42046', 'praviny2_', 'pre', 'prebook', 'precaution', 'precious_ly_', 'preeshu_', 'preety88888', 'prefer', 'premium', 'preparing', 'presentation', 'presented', 'president', 'pressing', 'prettiee_22', 'pretty', 'previous', 'price', 'priced', 'prices', 'pride', 'primarily', 'prinkavirvani', 'prints', 'priority', 'priti_00001', 'private', 'priya', 'priya56646111', 'priya958870', 'priyankakrithi', 'priyankalubb', 'priyankasarka4', 'priyash274', 'problem', 'problems', 'problemsolve__', 'proceeding', 'produce', 'produces', 'producing', 'product', 'products', 'promising', 'promote', 'proof', 'prooving', 'proud', 'provided', 'prudhvi486', 'psrishti124', 'puddinj007', 'punjab', 'punjabadikudi', 'pushti5678', 'qatar', 'qhdposts', 'quadratrac', 'qualities', 'quality', 'quickly', 'quite', 'quotes', 'r108pankaj', 'r_i_n_y_a_36', 'raaj_____', 'raani_sneha1', 'radhikaredy', 'radhikavyas00', 'raghu_123_', 'raha', 'rahe', 'rahega', 'rahi', 'rahul', 'rahulku44780901', 'rahull_9', 'rainie', 'raishivani_', 'raj1_2', 'rajasthan', 'rajatlps2', 'rajatlunkad', 'rajdeep_desai_', 'rajivbanat3456', 'rajpriya45', 'ram', 'ramban', 'rameshr12137894', 'rameshv150', 'ramjeetkush', 'rampage', 'ranar3r', 'rao_eshika', 'rashmika_36', 'rashmika_khush',

'rasnmiravii', 'rasid312', 'rated', 'rathores053', 'ravipanat', 'ravikhemka7', 'ravirajput30', 'rawalparesh2', 'rawanirekha', 'rbi', 'reaction', 'reactions', 'readselective', 'ready', 'real', 'real_dasg', 'realising', 'really', 'realmanishyadav', 'reason', 'reasonable', 'rebeifan', 'rebeinlovee', 'rebelsalaar', 'rebuilt', 'record', 'records', 'recover', 'recruiting', 'recruitment', 'regard', 'regarding', 'regenerative', 'reko4156', 'relation', 'relatives', 'release', 'released', 'releasing', 'remained', 'remake', 'remarkable', 'remember', 'remembrance', 'renton', 'reopen', 'reopened', 'reopening', 'reopens', 'repairs', 'replaced', 'replica', 'replicawheels', 'reply', 'reports', 'reprise', 'republicans', 'rescue', 'reshma26153023', 'respect', 'response', 'restored', 'returns', 'retweet', 'revolution', 'rezervd', 'rha', 'riddhisai1', 'ride', 'ridesafely', 'riding', 'right', 'rihanchou3', 'rimiiii1318', 'rishabhlohiya35', 'rishikaa_02', 'rishithatweetss', 'ritikgoyalll', 'ritikkumar581', 'ritu_s00', 'rituraj4078', 'rivalry', 'riyachoudhary67', 'rizwankha76', 'rj13_se_hu_bc', 'rm', 'road', 'roader', 'roading', 'roads', 'rock', 'rockybh08429736', 'rohantweets__', 'rohit_mohanty01', 'rohithking12', 'rok', 'roland', 'rolisingh8077', 'ronakb2929', 'ronitvatsya', 'ronyy1241', 'rooting', 'roshni_rajput44', 'rowdyak36', 'royal_lover_999', 'royalstage20', 'rraja_k1', 'rs', 'rubijadaun', 'rubybennet20', 'ruchipr07', 'rudrap_ratap', 'rugged', 'ruhikum87776941', 'rule', 'ruled', 'rumigautam91200', 'run', 'running', 'rupadas92884041', 'ruparoy58878945', 'rushali986', 'rushed', 'russian', 'ruthikasri12', 'rutu77811', 'ryan1_da', 'saanu319j', 'saanviiii1', 'sab', 'sabhi', 'sabse', 'sach', 'safe', 'safety', 'safty', 'sahara', 'saheb', 'sahilsaraan', 'said', 'sainik', 'sakhsi417', 'sakshi_98t', 'sakshi_jii', 'sakta', 'sal_22_', 'sale', 'sales', 'salesman', 'salgaonkar', 'salgaonkars', 'salgonkar', 'salma1akhtar', 'salman41365044', 'salmanaarohi2', 'salmans30529639', 'sam', 'samarnyon', 'sameer8929', 'sameersabat001', 'samiimeranaam', 'samne', 'sanatanipratap', 'sanayaa007_', 'sanchit01_', 'sandeepbunny_vj', 'sandeepyogii', 'sandhiya_rajput', 'sandhyashukla34', 'sandyyk9', 'sangita_m779', 'sanjayjaat456', 'sanjugodara2929', 'sankiraman', 'sanskari_enough', 'sanskriti__7', 'santrampaljimaharaj', 'santrampaljiquotes', 'sanyasa72', 'sara', 'sara46125540', 'saraakhanna485', 'saraoffcl_x', 'sarcasticgaurab', 'sare', 'sarsasmicgirl', 'sarthika50', 'sarvirat', 'sastaheisenberg', 'sath', 'satin', 'satisfied', 'satyaku59343069', 'satyam38625036', 'saumyaktariya91', 'sauravbanat', 'save', 'saved', 'saving', 'savrav6d', 'saw', 'say', 'saying', 'says', 'scattered', 'scene', 'scheduled', 'scheme', 'school', 'scre', 'screen', 'screens', 'script', 'se', 'sea', 'seat', 'seats', 'second', 'secrets', 'security', 'seed', 'seeds', 'seeing', 'seeker', 'seemasi00571396', 'seen', 'segment', 'sekh_11', 'sekh_113', 'select', 'selling', 'senior', 'seq', 'sequel', 'serialchiller42', 'series', 'seriously', 'services', 'set', 'setting', 'seven', 'sexy', 'shaikh_riya_', 'shakshi0121', 'shalu99965', 'shamli', 'shamshukla097', 'shank_51', 'sharathpalle94', 'shardakhemka', 'share', 'sharma_niiva', 'sharmaabhishok', 'shashi_priya01', 'shehnaazlu', 'shehnaz84790754', 'shelli9675', 'shellijadon', 'shemaroo60yearsyoung', 'shemarooent', 'shersinghkhella', 'shilpas02281154', 'shimona sharma3', 'shishupalduve', 'shivani70638757', 'shivani9_', 'shivanikumari06', 'shivanya345', 'shiz__uka', 'shoes', 'shop', 'showing', 'shown', 'showroo', 'showroom', 'shraddhawalkar', 'shreyaa1430', 'shreyaaax8516', 'shreyansh_60', 'shris_ti_', 'shrrrrish', 'shruti_singh99', 'shrutit03466956', 'shubam__d', 'shubbuu', 'shubham3_india', 'shubhamkumarydv', 'shubhamsabat80', 'shubhjswll', 'shweta2point', 'shweta_987', 'siddharth6788', 'siddiquianayara', 'sif_pune', 'simply', 'simran9917', 'simranshukla355', 'sin', 'singh17_abhay', 'singha_9798', 'singhsahasa', 'sir', 'sirakashtic', 'siru_is_queen', 'sister', 'sitting', 'situation', 'sivangi99', 'skf_sunny', 'skr_fen', 'slave', 'slavery', 'slvg', 'sm_iconic', 'small', 'smart', 'smedleyburke1', 'smile', 'smitada9', 'snehu8001___', 'snigdhapandt', 'snow', 'soarbeamdigital', 'society', 'sofi0_1_', 'soldiers4indian', 'solid', 'solve', 'solved', 'somyarathi7', 'sonalika077', 'sonalshingh', 'sonam578089', 'sonamgrewal10', 'sonamgupta22', 'song', 'songs', 'sonisi6794', 'soniya9_', 'sonukumar9_', 'sonupanday22', 'sonurajput__', 'sonurathorarman', 'soo', 'soon', 'sophisticated', 'sorabhchoudhry', 'soul', 'sounds', 'south', 'souvik__d', 'sowmyasri01', 'space', 'spec', 'special', 'specialist', 'specifically', 'spectacular', 'speed', 'splash', 'sports', 'spread', 'spring', 'sree_nikhil9999', 'srinagar', 'srinivasrc143', 'srinumb18', 'srt', 'srt8', 'stage', 'stakes', 'standard', 'standout', 'start', 'started', 'starting', 'statement', 'stay', 'steevyscofield', 'stellantis', 'step', 'steve_rogers25', 'stock', 'stop', 'stories', 'story', 'straight', 'strength', 'students', 'stunning', 'stupendous', 'style', 'stylish', 'stylish_allu55', 'sub', 'subalku87101906', 'subconsciously', 'success', 'successes', 'suchita3000', 'suchitasen4', 'sudhans97562905', 'suganakayal', 'suman_its1', 'suman_its144', 'sumanku85668691', 'sumisom3', 'sumitrathod58', 'summer', 'sunena___', 'sunilku58705103', 'sunny____k', 'super', 'superadian', 'superb', 'superhit', 'superstar', 'supirgori', 'support', 'surajjjaat665', 'surbhis1100', 'sure', 'surely', 'sureshot', 'sureshpatel7718', 'sureshrathi980', 'surojit55426172', 'surprise', 'suruchika7', 'suspense', 'suv', 'suvs', 'svagarwal29', 'swapna', 'swarupxyz', 'sweetypriya_20', 'sweta2576', 'switches', 'tabu', 'taejin', 'tag', 'taimoor777k', 'tak', 'taki', 'taking', 'talented', 'talk', 'talking', 'talks', 'tanurajput56', 'tanya__32', 'tarakrajesh9', 'tareekh', 'tarrrock_twin', 'taruvatha', 'tathaga48814316', 'teach', 'teacher', 'team', 'technical', 'technique', 'technir700', 'technol', 'technology', 'teens', 'tejran', 'tejv_eer', 'tell', 'telling', 'tensionhai', 'terminated', 'terrain', 'test', 'tgdealsofficial', 'th', 'tha', 'thank', 'thankful', 'thanks', 'tharapoojan', 'the_real__gamer', 'thea', 'theater', 'theaters', 'theatre', 'theatres', 'theatrical', 'thedominant_o', 'theminskashi', 'themintingm', 'therethrough', 'thi', 'thia', 'thing', 'things', 'think', 'thinking', 'thisizzrocky', 'thiss', 'thomas_tom12', 'threaters

', 'thrill', 'thrilled', 'thriller', 'thrilling', 'thrown', 'thunderthorr', 'ticket', 'ti
ckets', 'tied', 'till', 'tillubolthe', 'time', 'timesnow', 'tk', 'tl', 'tntimesdrive', 't
oday', 'toh', 'toiauto', 'tommorow', 'tomo', 'tomor', 'tomorrow', 'ton', 'tongue', 'tons'
, 'torq', 'total', 'totally', 'touchalove', 'touchalove12', 'touching', 'toy', 'traffic',
'trail', 'trailer', 'trailhawk', 'training', 'transferred', 'travel', 'traverse', 'tree',
'tremendous', 'trends', 'trinity_reload', 'trip', 'triumphs', 'trivikram_1', 'true', 'tru
ely', 'truly', 'trust', 'truth', 'try', 'ttarakitha', 'ttl', 'tu', 'tulikapandey14', 'tum
haraba', 'tune', 'tuned', 'turbo', 'turbocharged', 'ture', 'turn', 'turns', 'tweeps', 'tw
eet', 'tweets', 'tweetshaina_', 'tweetslove143', 'twisha_upadhyay', 'twists', 'type', 'ty
son', 'udaydere', 'udyotp', 'ukrainian', 'ultimate', 'umakuma50845437', 'un____conscious'
, 'underestimate', 'understand', 'undoubtedly', 'unfashionable', 'unique', 'uniquechrysle
r', 'uniquedevsharma', 'unit', 'universal', 'university', 'unlike', 'unmatched', 'unrival
led', 'unveiling', 'unveils', 'upcoming', 'updated', 'updates', 'upgrade', 'ur', 'ureleve
nn', 'urvashi91200', 'use', 'usha01_', 'using', 'vachesindhi', 'vaishali9636', 'vanavil69
69', 'vandnamishra374', 'vanella_cup', 'vanshika_f', 'variant', 'various', 'varmarajuu73'
, 'varsetile', 'varunpachauri13', 'vaseemakram485', 've', 'vechile', 'vehicle', 'vehicles
', 'venture', 'verma__000', 'version', 'vibesof3am', 'vice', 'victims', 'video', 'vihangu
pta15_', 'vij', 'vijay', 'vijaykumartj1', 'vikram_e18', 'vikramsingh170', 'vikrant_soni16
', 'vinay_ambiger', 'vinitamanju96', 'vinniewooo', 'vinodgu63597924', 'vintage_babu', 'vi
ralbake', 'virattlove18', 'vishalokhanna', 'visheshji2022', 'vishuu112', 'visit', 'vivek_
da1', 'viyabharathi73', 'vkbro4567', 'vldmiraditynath', 'vlog', 'wa', 'wacth', 'wait', 'w
aited', 'waiting', 'wala', 'wali', 'wanna_hot_', 'want', 'wao', 'wapis', 'waseems01990387
', 'waste', 'wasting', 'wat', 'watch', 'watched', 'watching', 'way', 'web', 'weekend', 'w
eird_exx', 'welcome', 'went', 'wh', 'wheeler', 'wheels', 'whic', 'wi', 'wide', 'wilmar',
'win', 'winning', 'winter', 'wireless', 'wisely', 'wishes', 'wishing', 'witness', 'witnes
sed', 'wo', 'woahh', 'wocharlog1234', 'wohhoo', 'wohi', 'wohooo', 'won', 'wonderful', 'wo
od', 'wooow', 'words', 'work', 'working', 'world', 'worth', 'wouu', 'wow', 'woww', 'wowww
', 'wrangler', 'write', 'wt', 'wtf_kajal', 'wtfpoojaa', 'x80', 'x82', 'x84', 'x85', 'x85m
erry', 'x87', 'x87_', 'x89', 'x8a', 'x8d', 'x8e', 'x8f', 'x90', 'x91', 'x92', 'x93', 'x94
', 'x96', 'x974', 'x98', 'x98grand', 'x98no', 'x98our', 'x99', 'x99ll', 'x99m', 'x99s', '
x99t', 'x99ve', 'x9a', 'x9b', 'x9c', 'x9cclass', 'x9cextensive', 'x9d', 'x9e', 'x9f', 'xa
0vijay', 'xa1', 'xa2rahul', 'xa4', 'xa5', 'xa6', 'xa8', 'xa9', 'xaashimax', 'xab', 'xad',
'xae', 'xaf', 'xb0', 'xb0_', 'xb0s', 'xb2', 'xb3', 'xb5', 'xb6', 'xb6jeep', 'xb7', 'xb8',
'xb9', 'xb9599', 'xb977', 'xba', 'xbc', 'xbd', 'xbe', 'xbf', 'xc2', 'xc3', 'xc4', 'xe0',
'xe2', 'xef', 'xf0', 'xhanqs_', 'xmartamit_', 'ya', 'yaad', 'yabhavani333', 'yah', 'yashr
jpt27', 'yatika1027', 'ye', 'year', 'years', 'yes', 'yeshpalsin', 'ylt19', 'youlittle__gi
rl', 'younger', 'yours_priya32', 'youth', 'youthclub_jammu', 'youtube', 'yuri_hojo', 'zay
n_rz_01', 'zeenewsenglish', 'zelensky', 'zoyakhanmam', 'zptvofficial']

```
/usr/local/lib/python3.8/dist-packages/sklearn/utils/deprecation.py:87: FutureWarning: Fu
nction get_feature_names is deprecated; get_feature_names is deprecated in 1.0 and will b
e removed in 1.2. Please use get_feature_names_out instead.
  warnings.warn(msg, category=FutureWarning)
```

In [ ]:

```
import pandas as pd
df= pd.DataFrame(corpus2.toarray(),columns=count_vectorizer.get_feature_names())
df2 = pd.DataFrame(cosine_similarity(df, dense_output=True))
df2.head()
```

```
/usr/local/lib/python3.8/dist-packages/sklearn/utils/deprecation.py:87: FutureWarning: Fu
nction get_feature_names is deprecated; get_feature_names is deprecated in 1.0 and will b
e removed in 1.2. Please use get_feature_names_out instead.
  warnings.warn(msg, category=FutureWarning)
```

Out[ ]:

	0	1	2	3	4	5	6	7	8	9	...	990	991
0	1.000000	0.133631	0.400892	0.267261	0.098058	0.375000	0.144338	0.250000	0.250000	0.400892	...	0.000000	0.000000
1	0.133631	1.000000	0.142857	0.142857	0.209657	0.133631	0.308607	0.133631	0.133631	0.285714	...	0.000000	0.000000
2	0.400892	0.142857	1.000000	0.428571	0.104828	0.534522	0.154303	0.267261	0.400892	0.428571	...	0.119523	0.000000
3	0.267261	0.142857	0.428571	1.000000	0.104828	0.267261	0.154303	0.267261	0.267261	0.285714	...	0.119523	0.000000
4	0.098058	0.209657	0.104828	0.104828	1.000000	0.294174	0.113228	0.098058	0.098058	0.209657	...	0.000000	0.240192

5 rows × 1000 columns

In [ ]:

```
df2
```

Out[ ]:

	0	1	2	3	4	5	6	7	8	9	...	990	9
0	1.000000	0.133631	0.400892	0.267261	0.098058	0.375000	0.144338	0.250000	0.250000	0.400892	...	0.000000	0.0000
1	0.133631	1.000000	0.142857	0.142857	0.209657	0.133631	0.308607	0.133631	0.133631	0.285714	...	0.000000	0.0000
2	0.400892	0.142857	1.000000	0.428571	0.104828	0.534522	0.154303	0.267261	0.400892	0.428571	...	0.119523	0.0000
3	0.267261	0.142857	0.428571	1.000000	0.104828	0.267261	0.154303	0.267261	0.267261	0.285714	...	0.119523	0.0000
4	0.098058	0.209657	0.104828	0.104828	1.000000	0.294174	0.113228	0.098058	0.098058	0.209657	...	0.000000	0.2401
...	...	...	...	...	...	...	...	...	...	...	...	...	...
995	0.000000	0.000000	0.000000	0.000000	0.240192	0.000000	0.000000	0.102062	0.102062	0.000000	...	0.000000	0.3333
996	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.111803	0.0000
997	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.098058	0.000000	0.104828	...	0.175412	0.0000
998	0.000000	0.000000	0.000000	0.000000	0.230769	0.000000	0.000000	0.098058	0.000000	0.104828	...	0.175412	0.2401
999	0.000000	0.000000	0.000000	0.000000	0.222375	0.000000	0.000000	0.094491	0.000000	0.101015	...	0.084515	0.2314

1000 rows × 1000 columns

In [ ]:

```
from sklearn.feature_extraction.text import CountVectorizer
count_vectorizer = CountVectorizer(stop_words='english')
count_vectorizer = CountVectorizer()
sparse_matrix = count_vectorizer.fit_transform(texts)
doc_term_matrix = sparse_matrix.todense()
df = pd.DataFrame(doc_term_matrix,columns=count_vectorizer.get_feature_names())
df.head()
```

In [ ]:

```
from sklearn.metrics.pairwise import cosine_similarity
dj=pd.DataFrame(cosine_similarity(df, dense_output=True))
dj.head()
```

Out[ ]:

	0	1	2	3	4	5	6	7	8	9	...	990	991
0	1.000000	0.062622	0.257248	0.259281	0.054233	0.210042	0.148522	0.313112	0.206835	0.269069	...	0.047565	0.137505
1	0.062622	1.000000	0.091287	0.069007	0.288675	0.149071	0.263523	0.133333	0.110096	0.358057	...	0.151911	0.000000
2	0.257248	0.091287	1.000000	0.283473	0.079057	0.408248	0.072169	0.182574	0.226134	0.294174	...	0.069338	0.000000
3	0.259281	0.069007	0.283473	1.000000	0.059761	0.154303	0.054554	0.138013	0.113961	0.148250	...	0.052414	0.000000
4	0.054233	0.288675	0.079057	0.059761	1.000000	0.193649	0.045644	0.057735	0.047673	0.124035	...	0.000000	0.126773

5 rows × 1000 columns