

cosine

December 12, 2018

```
In [2]: '''Importing our modules'''
import re
import math

In [3]: '''to open and read our tweets to begin the preprocessing step'''
raw_text = open('tweets.csv').read()
tab_seperated = [item.split('\t') for item in raw_text.split('\n') if len(item.split('\t')) >= 4]

In [4]: data = {}
for i in range(len(tab_seperated)):
    '''extracting the filtered tweets and IDs'''
    data[tab_seperated[i][1]] = tab_seperated[i][4]

    '''Preprocessing and cleaning'''
    data_index = data.copy()
    for tweet in data_index.keys():
        data_index[tweet] = data_index[tweet].lower()
        data_index[tweet] = re.sub('\[newline\]', ' ', data_index[tweet])
        data_index[tweet] = re.sub('https?[\s]+', ' ', data_index[tweet])
        data_index[tweet] = re.sub('@#[^\s]+', ' ', data_index[tweet])
        data_index[tweet] = re.sub(r'[0-9][^\s]+', ' ', data_index[tweet])
        data_index[tweet] = re.sub(r'\w+\. [^\s]+', ' ', data_index[tweet])
        data_index[tweet] = re.sub(r'[a-zA-zöüß\s]', ' ', data_index[tweet])
        data_index[tweet] = re.sub(r'[^\w\s]', ' ', data_index[tweet])
    '''creating a term- frequency dictionary for the collection'''
    terms = {}
    for num,tweet in data_index.items():
        for word in tweet.split():
            if word in terms:
                terms[word] += 1
            else:
                terms[word] = 1

In [5]: '''create the weighted idf values of each term'''
idfDict = {}
N = len(data_index)
idfDict = terms.copy()
for word, val in idfDict.items():
    idfDict[word] = 1 +math.log10(N/float(val))

In [6]: '''This function takes two strings as input and computes their cosine similary score'''
def cosine(text1, text2):
    import math
    import re
    '''tokenizing the tweets and building a local word dictionary'''
    def wd(t1, t2):
        wd = {}
        t1 = t1.split()
        t2 = t2.split()
        for i in t1:
            wd[i] = 0.0
        for i in t2:
            wd[i] = 0.0
        return t1, t2, wd
    '''Building our weigthed TF vector '''
    def TF(wordDict, txt):
        tfDict = {}
        for word, count in wordDict.items():
            for i in txt:
                if word == i:
                    count +=1
            if count > 0:
                tfDict[word]=1+math.log10(count)
            else:
                tfDict[word] = 0.0
        return tfDict
```

```

'''Build our TF-IDF vectors using our IDF dictionary'''
def TF_IDF(tfs, idfs):
    tfidf = {}
    for word, val in tfs.items():
        tfidf[word] = val*idfs[word]
    return tfidf
'''calculate the normalizing length for a given vector'''
def length(tf):
    return math.sqrt(sum([(i**2) for i in list(tf.values())]))
'''dot product of two vectors'''
def dot(v1,v2):
    new_v = []
    for i in range(len(v1)):
        new_v.append(v1[i]*v2[i])
    return sum(new_v)

'''tokenize and build local word dictionary'''
txt1, txt2, wd = wd(text1, text2)

'''return a 0 for empty vectors'''
if len(txt1) == 0 or len(txt2) == 0:
    return 0

'''compute and populate the tf vectors'''
tf1 = TF(wd, txt1)
tf2 = TF(wd, txt2)

'''compute and populate the tf-idf vectors'''
tf_idf1 = TF_IDF(tf1, idfDict)
tf_idf2 = TF_IDF(tf2, idfDict)

'''compute the lengths of each vector'''
l1 = length(tf_idf1)
l2 = length(tf_idf2)

''' return 0 for vectors with length of 0'''
if l1 == 0 or l2 == 0:
    return 0

'''return rounded cosine similarity score'''
return round(dot(list(tf_idf1.values()), list(tf_idf2.values()))/(l1*l2),4)

```

In [7]: '''creates a list of tuples containing cosine score, id, and raw tweet sorted by score in descending order and restricted to the top 100 '''

```

def top_100(tweet):
    top = []
    for id, doc in data_index.items():
        if id != tweet:
            cos = cosine(data_index[tweet], doc)
            top.append((cos,id,data[id]))
    return sorted(top)[::-1][:100]

```

In [8]: '''Tweet selection from dictionary'''

```

for num, key in enumerate(data_index.keys()):
    if num == 0:
        print(num, data[key], '\n', key, data_index[key])

```

0 @knakatani @ChikonJugular @joofford @SteveBlogs11 <https://t.co/WHtaRYGNSY> says lifetime risk of cervical cancer in Japan is 1 in 100. That 965734992633565184 says lifetime risk of cervical cancer in japan is in that means hpv is endemic in japan and screening is

In [22]: '''Tweet selection from dictionary'''

```

print("Query:", data['965734992633565184'], '\n')
top = top_100('965734992633565184')
for i in top:
    print(i)

```

Query: @knakatani @ChikonJugular @joofford @SteveBlogs11 <https://t.co/WHtaRYGNSY> says lifetime risk of cervical cancer in Japan is 1 in 100.

(0.3896, '982642444943605760', "@beautyloveknow1 @VABVOX Yeah, cervical cancer is so rare that the lifetime risk of getting it is 1 in 161... (0.3413, '988887135699648512', "@vikki_r To prevent the cancers HPV causes?[NEWLINE]Lifetime risk of cervical cancer in US is 1 in 161 without (0.3389, '968982710222970880', "@RustyPee4 @williambolivar7 HPV 16 and 18 account for ~70% of cervical cancers, that's true.[NEWLINE]HPV 31, (0.3279, '968978184438915072', '@RustyPee4 @williambolivar7 No, alas, condoms don't reduce HPV risk much, the 1 in 100 lifetime risk already (0.3243, '1022596646591295488', "@LarisaAskew The lifetime risk of cervical cancer is 1 in 161, vs 1 in ~100,000 for HPV vaccine (see <https://> (0.3147, '991103844959506433', '@GeoffSchuler @MamaBear11011 @NeilTheMason @BillyMontana81 @EmmaGPaley @RyanMarino @LiamKav @StopVaxxedLies @ (0.2927, '976607630687813633', "@NancyARandazzo @chriscvsefalvay @SeattleiteLeo @and_kell @DrMichaelMaster @LaLaRueFrench75 The main risk fac (0.2589, '975868739605250049', "Contribution of HPV in neuroendocrine tumors from a series of 10,575 invasive cervical cancer cases": <https://>

(0.2554, '1017042589646639104', '@CBTweets7 @IntactCervix @starlinguk @laurabrennan091 @V_Andriukaitis It's complicated, but:[NEWLINE]1) smeas

(0.247, '983764707231203329', '@Wildlife2075 @DrMaurieMarkman HPV tests provide better protection than pap smears; that's why Australia has s

(0.2449, '966097477915967491', '@MandaMousePaws @hchk78 I'm interested in preventing cancer.[NEWLINE]Evidence is strong the HPV vaccine grea

(0.2344, '985156667829510145', '@gorman_garry @dublinwellwoman @IrishCancerSoc @HSEImm About 70% of invasive cervical cancers are due to the

(0.2319, '971150719981113344', '@Australia could become first country to eradicate cervical cancer. Free HPV vaccine program in schools has le

(0.2301, '1013861958414688261', 'Japan just surprised me')

(0.2284, '971296846617473024', '@Grantscheam @the_necrosis In Japan ist man schon soweit ;-))

(0.2221, '959612011242864640', '@PeterD_1996 @Anti_inequality @PaulaJaneB True that there is no evidence HPV is involved in prostate cancer.[

(0.2198, '960554806581407744', '@JayinWashington We know the risk of cervical cancer and other vaccine preventable diseases is significant.[N

(0.2195, '970615849588547584', 'Self Sampling Can Benefit Women At Risk Of Cervical Cancer - Anti Aging News <https://t.co/CMzhDkry7X>')

(0.2143, '979868405082808322', 'Too much testing is dangerous: Prostate cancer screening does NOT save lives, puts men at risk according to g

(0.2125, '982951510500753408', '@chifarayi @Tawand89 @SibandaSibbs @denvern3 The short answer:[NEWLINE]- ~99% of cervical cancer is caused by

(0.2112, '968976919608770560', '@RustyPee4 @williambolivar7 If you add the other HPV-caused cancers, lifetime risk is around 1 in 100.[NEWLIN

(0.2097, '988138302778822658', '@Mangan150 I am borderline obese in Japan.')

(0.2085, '975521139186429953', '@iBlali Cancer')

(0.2085, '963199970055741445', '@kxantwinz cancer')

(0.208, '982628545154330624', '@realPaulHaye @christufton You forgot these facts: [NEWLINE]- 50% of all cervical cancers are not detected by

(0.2063, '982629304604311554', '@realPaulHaye @christufton - ~99% of cervical cancer is caused by HPV, <https://t.co/MDsxagPIJ3> [NEWLINE]- vac

(0.2056, '984163305827831813', '@moorehn Inflammation is key in cancer, if not the cause of some')

(0.2042, '999660387057446917', '\$SGMO anti tau would lower risk of cancer as well ! <https://t.co/HdDIhbgVVs>')

(0.2041, '989339479109353472', '@lifebiomedguru @jaggakatja @toppediatrician @mcfunny @kevinault @viva_lala @Just4TheCause @Joshmazer2018 @P

(0.2012, '982748295276650497', '@beautyloveknow1 That's right, the vaccine only lowers the risk of cervical cancer greatly - by about 70% - n

(0.1983, '98602805977168896', 'US public may lack understanding of cancer screening - Reuters <https://t.co/iFnEXmZK5>')

(0.1983, '986028029699862529', 'US public may lack understanding of cancer screening - Reuters <https://t.co/aXZk0MwUrv>')

(0.1961, '98015706688962560', '@Benardean2 @komonews There's a movement to offer HPV vaccinations to boys. Big difference.[NEWLINE][NEWLINE

(0.1958, '991432747414622208', 'IM SO PISSED MY WIFI IS NOT WORKING')

(0.1941, '992337486360862720', 'Japan not fearful of robots stealing jobs, deputy chief says - Information by <https://t.co/I1J9Mvgc2f> <https://t.co/I1J9Mvgc2f>

(0.1924, '976504460078804993', 'Der Tunnel der Liebe liegt in Japan [U+1F33A] [NEWLINE] [NEWLINE] #Japan #love #Travelers <https://t.co/TezChYKMg>

(0.1871, '984174853845495808', '@chifarayi To prevent cancer and save lives.[NEWLINE] [NEWLINE] Vaccination is the only tool we have at the mom

(0.1867, '961933803575848960', 'Lung Cancer Screening: Updated Guidelines From CHEST <https://t.co/joqEo81bpz>')

(0.1854, '983053084866691073', '@YengwayokaKomo @chifarayi @SibandaSibbs @Tawand89 @denvern3 About 99% of cervical cancer is caused by HPV, v

(0.1823, '959119033818853378', 'The future is here and now! Bladder cancer has the highest lifetime treatment cost per patient of all cancers

(0.1816, '993928144288665601', 'I'm even more nervous than in 2015. And that means A LOT.')

(0.1715, '1011300674620022784', '@ReltheBelle @broodingYAhero You act like that means anything...')

(0.169, '991751677194719232', 'THAT'S WHAT IT IS ABOUT, IF YOU'RE NOT SCARED OF DOING IT THEN DON'T IF YOU ARE IT MEANS IT MATTERS <https://t.co/I1J9Mvgc2f>

(0.1685, '985860979392049153', 'Seroprevalence and risk factors of Hepatitis E virus infection in cancer patients in eastern China. <https://t.co/I1J9Mvgc2f>

(0.1663, '991773325306290176', '@taehnyungie WELL, sad is not so good, but happy is very good! :0')

(0.1659, '1014260978299654144', '@Xynxa @FuDePanda @Lnkas Well that is just sad then and not Worldchampion-worthy')

(0.1657, '974681253223059456', '#Cisplatin is a #chemotherapy #medication used to treat a number of cancers. This includes testicular #cancer

(0.1654, '984048170085822464', '@drphil_urology Right! The interesting fact is that the risk of recurrence is not getting less over time in s

(0.165, '98924167616593921', '@AshAngelArt @SeaOfThieves No that means you have autism')

(0.1638, '995280813024407552', '@nbc is just doing that is so I am not angry at them any more for canceling #Hannibal... and yeah, fuck damn

(0.1628, '991426100650758147', '@voiceswrestling I'm so frustrated with a lot of things going on in Big Japan that I'm about to throw the tow

(0.1627, '989475550543695872', '@gerpux @jamaloh_james The entire point of this is that in Japan women's wrestling is strong enough to run th

(0.1623, '990130115660124160', '@josephwrio @RandyEBarnett should absolutely step in. What is different about this case, is that the state sa

(0.1592, '975753551392616449', '@RMult @SmallcageGG @psiangelic @Anna3344 Sorry I'm not a native speaker...minor cancer means it's only a sma

(0.1582, '970860343848136704', '@RustyPee4 @kevinault @KittenKoder @ProAntiVaxxer @Plasticdoe @TheFrankmanMN @chriscvsefalvay @FreeBeeCee @Ge

(0.1569, '1000449596936646656', '@Shakarlicious I wish i could gain some weight XD My metabolism is working TOO well XD')

(0.1564, '966458088491569152', '@CrudeCSGO @MichaelMartin @Stun_Empire Japan, prepare your anus')

(0.155, '988561012495335425', '@alexstv Esse ich hier in Japan häufiger als normale Nudeln [U+1F605]')

(0.154, '1001242477104418816', 'working on some graphic design, fiddling around... [NEWLINE] well not happy yet [U+1F610] [U+1F914] [U+1F603] <https://t.co/I1J9Mvgc2f>

(0.1538, '993578386156027904', '@TheGoddessSara well, that's why i'm not scared :P')

(0.1527, '967909667409223681', '@suns243 ranked is so cancer lol')

(0.1526, '977334147088728064', 'i have obs installed as well yeah.. so some kind of working camera etc is the main think i need and waiting 2

(0.1526, '959237268031528960', '@JayinWashington Vaccines don't really have politics. Let's disagree about one thing at a time :-)[NEWLINE][

(0.1524, '959574201328852992', 'Kebbi First Lady Donated Breast Cancer Screening Kits To 500 Women <https://t.co/MCWyw4sXS5> <https://t.co/zFLX0>

(0.1524, '1023895789825937408', 'Tomorrow is my last working day in this company. Kinda happy kinda sad.')

(0.1523, '1002276602003279872', 'I will love you for a lifetime [U+2764] i'll do anything for you [U+2764] [U+2764] #beaglesofinstagram #Beag

(0.1505, '975519024351084544', '@eshypresso as a cancer,')

(0.1494, '1000821535131750400', 'Mothers hold their #childs hand for a moment, and their #heart for a lifetime.[NEWLINE] #Mother #Haitianmom #

(0.149, '1010444113974571008', 'He seemed to respect Japan as well. But Japanese now are very nervous because they fear that the US might onl

(0.1487, '989486224313331712', '@sunnytheslaya @europaraiders again...127 has a japanese member why are yall so shocked about the debuting in

(0.1479, '1008160677947510785', 'my bf could pick me up. If not I'd have just walked 40 mins.. at like 1 AM. I mean I'm not really scared her

(0.1473, '1013846065550970880', 'I'd be happy if Japan actually made it. Happy & very surprised. [NEWLINE] #BELJPN #WorldCup')

(0.1469, '999801434819395584', 'Fuck cancer. :(')

(0.1468, '992550875133460480', '@ThobiasO @rovoca @REWE_Supermarkt Immer noch besser als in Japan, wo es nur 1-Finger-Kitkat davon gibt für 4

(0.1458, '994992508206755840', 'The #tomato is #alive... [NEWLINE] [NEWLINE] #tomate #leben #life #love #frühling #spring #liebe #plant #pflanzen

(0.1458, '975546597172969477', '#gratitude is the #best #medicine #meditation #nature #hiking #love #joy # [U+2764] [U+FE0F] #newearth... <https://t.co/I1J9Mvgc2f>

(0.1447, '967461629930635270', 'C-sections and #gut #bacteria increase risk of #childhood #obesity <https://t.co/5x09IVwh7g>')

(0.1437, '987843221991690240', '@Madonna That is really sad!')

(0.1434, '989388484837179393', 'not happy...')

(0.1429, '974793779914797056', '@ChristineJaniuk @ChrissieCat1225 @joegooding @RealJamesWoods Oh, I'm not claiming any expertise. I'm just r

(0.1426, '991675431299506176', ':/ :/ :/ Stop fucking around and give him new songs in Japan, Avex. When your artist is so frustrated that he

(0.1423, '985128889436524544', '@VegaVoid Dann ist ja gut. An das Screening hab ich nämlich auch gedacht.')

(0.1418, '980116502954131456', '@JohnJoedotcom Do you believe that HPV vaccine injures more people than HPV-caused cancers do? [NEWLINE] [NEWLI

(0.1416, '98200915831349248', 'What about cancer')

```
(0.1406, '1007376491934617601', 'ASTHMA IS THAT YOU https://t.co/WLAssA3cfk')
(0.1402, '962431501001347077', '@locobutterfly @heather42077 @expatback @KnowHPV Ah, so if the HPV vaccine prevents a million cancers, and ca
(0.1392, '982490978798985216', 'Es ist so seltsam und komisch wieder in Deutschland zu sein .. [U+1F648][U+1F62D] Ich vermisse Japan jetzt sch
(0.1392, '959522237501100032', 'Obesity and Cancer Risk (Fact Sheet) - Oncology Nurse Advisor https://t.co/nLtl0jS3k4')
(0.1388, '921846269186633477', 'Fuyuhiko is trying to cosplay as Solid Snake and it's not working h alp")
(0.1388, '989289298754179073', 'Fuyuhiko is trying to cosplay as Solid Snake and it's not working h alp")
(0.1388, '987828422553501721', 'Fuyuhiko is trying to cosplay as Solid Snake and it's not working h alp")
(0.1388, '980524042519240704', 'Fuyuhiko is trying to cosplay as Solid Snake and it's not working h alp")
(0.1388, '971494544016662533', 'Fuyuhiko is trying to cosplay as Solid Snake and it's not working h alp")
(0.1388, '968584116479381504', 'Fuyuhiko is trying to cosplay as Solid Snake and it's not working h alp")
(0.1388, '968221730299883521', 'Fuyuhiko is trying to cosplay as Solid Snake and it's not working h alp")
(0.1388, '967832916876181505', 'Fuyuhiko is trying to cosplay as Solid Snake and it's not working h alp")
(0.1388, '961385436001243137', 'Fuyuhiko is trying to cosplay as Solid Snake and it's not working h alp")
(0.1388, '959143160411762688', 'Fuyuhiko is trying to cosplay as Solid Snake and it's not working h alp")
(0.1388, '1013172924990263296', 'Fuyuhiko is trying to cosplay as Solid Snake and it's not working h alp")
(0.1388, '1003060036527165440', 'Fuyuhiko is trying to cosplay as Solid Snake and it's not working h alp")
```

```
In [10]: '''Tweet selection from dictionary'''
for num, key in enumerate(data_index.keys()):
    if num == 8:
        print(num, data[key], '\n', key, data_index[key])
```

```
8 @rip_tear Tumors?[NEWLINE]Dat leg is straight mccain
965481507266727936 tumors dat leg is straight mccain
```

```
In [23]: print("Query:", data['965481507266727936'], '\n')
top = top_100('965481507266727936')
for i in top:
    print(i)
```

```
Query: @rip_tear Tumors?[NEWLINE]Dat leg is straight mccain
```

```
(0.412, '974741740807737344', 'Leg' auf[NEWLINE]#KayasGroßeKinoshow")
(0.299, '960300385738031105', '@Lengrad du und straight als ob')
(0.2476, '960119588309688320', 'Did McCain's Brain Tumor Come From Cellphone Use? https://t.co/0u4HkeKHFZ via @mercola")
(0.2446, '997875345767444480', 'HE HAS TUMORS???)
(0.2438, '988902240260886528', 'Trump is obese, not large.[NEWLINE][NEWLINE]Get it straight. https://t.co/pj8omIV710')
(0.2432, '976800329839898625', 'Show a little more leg, honey [U+1F495][NEWLINE].[NEWLINE].[NEWLINE].#leg #woman #sculpture #clay #c
(0.2394, '962780166966726656', '@the_necrosis @KieferGabriele @ChristianNbg Was soll dat heißen? [U+1F64A]')
(0.2272, '995409237324034049', '@mounira_ber [U+062A][U+0642][U+0631][U+064A][U+0628][U+0627][U+0643][U+0644][U+0627][U+0644] Solid tumors.
(0.2208, '986434399804841992', 'Here is to walking funny on Wednesday's. Love me some leg day!!![NEWLINE]#bigjoeyd #josephdesanze #love... ht
(0.2206, '967544965357162496', 'region for seven straight https://t.co/hXVihEwb8p https://t.co/1G30G0VUAg')
(0.215, '1010970068560293890', '@waterparks His leg hair just scared tf out of me')
(0.2114, '966713114770231296', '@AtypistAnne @Proud2BeAut Jeetje, dat is ernstig!')
(0.2013, '975807026034028545', "[U+1F3AE] Horror ist durch? Dat war's! [U+1F534] JETZT LIVE https://t.co/LlBy519XSG #150Follower #rulg #monta
(0.201, '1005476423828758528', 'I'm[NEWLINE][U+1F535]Straight[NEWLINE][U+1F535] GAY[NEWLINE][U+1F518] Afraid I just "blue" myself https://t.
(0.1948, '959868538617032705', '@BILD_Politik McCain ist Teil des tiefen Staates. Sein Tumor hat seinen Charakter noch einmal verändert, zum
(0.194, '970811133899755520', '@jamescharles I m not gay.I am straight but as long as i support others its okay[NEWLINE]Love is love')
(0.1934, '998598663365693441', '*slightly nervous laugh* i can't really think straight[NEWLINE]#YOONGI #BTS #BTSBBMAs https://t.co/emhMPILqZ0
(0.1825, '1010208281640488961', 'DAT HAM WA UNS VERDIENST!!! [U+2605] BOLOGNA.ITALY [U+1F355][U+1F44C][U+1F3FB][U+2600][U+FE0F][U+1FEE][U+1F
(0.1804, '982761886629408769', '@Crisisart @GermanLetsPlay Manu als Dat Adam Mitglied? [U+1F914]')
(0.1764, '974971929370210309', 'sucks sex anal tumors https://t.co/xtF5WZTkcg')
(0.1761, '989601618059583488', 'Ich bin kinda pissed dat Trixi gehen muss #GNTM')
(0.1757, '999768260995371008', 'Can Cannabis Fight Tumors? https://t.co/xtiWzrT91M')
(0.1755, '966458539450527744', '@CrudeCSGO @_MichaelMartin @Stun_Empire This straight up looks like Logan's mom lol')
(0.1745, '999797912065409024', 'Dat is Kunst. [NEWLINE]Wieso willst du nicht deine Karriere als Künstler starten? https://t.co/PSsvyqjdJe')
(0.1721, '986017684293865473', 'Take one bite and then have your leg amputated bc of Diabetes https://t.co/WdiKhYTmX4')
(0.1661, '1000167376682475520', '@AnfieldRd96 If we end up winning the CL my leg is ready to get tattooed with the Liverbird.[NEWLINE][NEWLIN
(0.1652, '967112502621491204', '@Rose_Blue87 Ja precies. Ik hoop dat dat voor gaat.')
(0.1612, '1001612325957271553', 'Ich leg mich wieder hin und versuche einzuschlafen. Ist schließlich nicht so, als hätte ich eine andere Wahl
(0.1607, '975057864204177413', 'anal tumors musclenudegirls https://t.co/ZdE20o0JKh')
(0.1577, '967752350667624448', '@jibs1990 @FunnyLADude @AyyitsJaay_ @Gangstabromie @Dande_Lion55 @Sms179812 @vincebeaulieu_ @Ericof_ @Mr_GBM
(0.157, '994992508206755840', 'The #tomato is #alive...[NEWLINE][NEWLINE]#tomate #leben #life #love #frühling #spring #liebe #plant #pflanzen
(0.157, '975546597172969477', '#gratitude is the #best #medicine #meditation #nature #hiking #love #joy #[U+2764][U+FE0F] #newearth... https:
(0.1556, '990513826075496448', 'Just leg pressed 3x165kg (363 pounds) and I'm hella happy because my maximum used to be 125kg (275 pounds) [U
(0.1548, '995228913386209288', '#THC shrinks tumors!! That's a fact!![NEWLINE]#LegalizeIt #medicalmarijuana #cannabis https://t.co/dHGh3jGEyn
(0.1541, '960291458883371008', 'Cancer 'vaccine' is eliminating tumors in mice, researchers found - https://t.co/lw2qZeTMqS https://t.co/DUCF
(0.1541, '960291434036277248', 'Cancer 'vaccine' is eliminating tumors in mice, researchers found - https://t.co/lw2qZeTMqS https://t.co/EiNX
(0.1536, '968582346315268096', 'Mal was für die Freunde des schwarzen Tumors. #Impfen #Impfgegner https://t.co/epq9JJUWd')
(0.1535, '995776984214097922', 'its actually pretty sad but ive been laughing at this for 5 minutes straight https://t.co/bNuwmzWERb')
(0.1504, '1014634232415703040', '-VCing or just straight up and getting an asthma attack right after it')
(0.1494, '959932821669376002', 'Ich bin stolz auf Jenny. Dat is nh süße zuckerschnute und I'm sorry aber sie ist hübscher als die Katzenberge
(0.1494, '1003563779391655937', '@jeffreycrash @GadSaad I really thought it was satire. How could you say anything like that with a straight
(0.1482, '960986531451277319', '@E8Emma @dertwerst @YourAnonCentral They should, because it's the therapy. You don't get therapy for a missin
(0.1473, '1013206622401941504', 'Lol dat Moment wenn man von nem 18ner als erstes heim muss weil noch 15:(')
```

(0.1472, '1006305358388883458', '@aNekoInSpace @EinKruemelchen @GermanLetsPlay Ich will auch Wörter verstehen! Und so tuhl sprechen können. S
(0.1466, '960301934715523072', 'Omg no I can perform sql injections straight in the os [U+1F4BB][U+1F601] https://t.co/z7fP8014Pu')
(0.1441, '960908349440188417', 'Straight Fire [U+1F525][U+1F525] The Red Embroidered Sweatshirt. Available Online. https://t.co/PjDPQUQNZ3[NEWLINE]
(0.1429, '998741903662346242', '@LawyerNoChills "Karma should fucking kick you right into your grave!!"[NEWLINE][NEWLINE]He was angry enough
(0.1415, '993189900236226561', '@samittie Not terrified. I need it. I am can't breathe straight until the day it dropps.")
(0.1413, '986306607209504768', '@Sequisha I have a thrombosis in my left leg, a pulmonary embolism, a pneumonia and fever. The best thing abo
(0.1413, '1014656795938377729', '@Shepherdrdog1970 @Dragonkiss1977 @rich752913078 Dat schreibste jetzt hundertmal ab. [NEWLINE]Und zwar als Th
(0.1396, '989565340081442816', 'Kanye makes me so mad. I promise he aint straight in the head. Only thing I'll respect him for is his shoes.
(0.1389, '1017774831071387648', '@HaznoTV @IslandsOfNyne True dat. But I'm just scared that will take like a year just like BugG')
(0.1387, '994294436480897024', 'Daydream came on and I straight out screamed bc I was so surprised... why am I like this again?')
(0.1387, '984484514876592130', 'pictures of anal tumors girls fucking guys with a https://t.co/01RUa9siGU')
(0.1384, '967115675998347265', 'Ist jetzt schon Silber! Und wenn die Russen Asthma haben, ist dat ne goldene.')
(0.1383, '967120345244356608', '@Tattoomara Oh. Dat was wel een fout gifje.')
(0.138, '1008348722340737025', '"women shouldnt have leg hair" haha nice try you fucking piece of shit you're just afraid mine'll be longer t
(0.1377, '994003017606168582', '@cccele4eva @2AStars @DavidDMiller5 @bb158 @bearerofmeaning @GHeadicus @sj9640 @carlfidelio @LawannaTurner3 @
(0.1372, '982152801915551747', 'Cancer VACCINE eliminates tumors in mice[NEWLINE][NEWLINE]https://t.co/mT6azVvXJO')
(0.1372, '960566346923806721', 'Cancer 'vaccine' eliminates tumors in mice https://t.co/6SMS0M9UFs')
(0.1365, '977329371131232256', '@mavaho2 @Tagesspiegel Wenn die Republikaner als der Ukraine-konflikt begann an der Macht gewesen wären hätten
(0.1357, '963436863758401536', 'brain tumors in adults fuck sleeping in ass https://t.co/12qTcORBeW')
(0.1351, '989946388623577088', 'This is my cousin Gareth. He lost his leg to cancer 3 years ago and now has 15 lung tumours and is terminally
(0.1351, '98990633389545472', 'This is my cousin Gareth. He lost his leg to cancer 3 years ago and now has 15 lung tumours and is terminally
(0.1349, '1004385103060094977', '@NiciSa Tbh it sounded like something bad is about to happen , he seemed really nervous :/. So I think he h
(0.1347, '959867370427822080', 'Vaccines to be made from human cancer tumors - https://t.co/05mIjoUc15 https://t.co/aBEfD482Y2 via @HealthRan
(0.1343, '984809073223888897', '@Imp , könnte ich mal kurz was mit dier Privat besprechen, wegen #Hate und dat.')
(0.1343, '1009901369606668288', '@Fusionnist Not enough tumors and doesnt make me puke 0.5/5')
(0.1341, '975883452141768704', 'Spirituele verlichting, hoe voelt dat?: Spirituele verlichting, hoe voelt dat? Als je de... https://t.co/FMcP
(0.1337, '959192808509919234', 'A cancer 'vaccine' is completely eliminating tumors in mice - New York Daily News https://t.co/jsDdA79eTf')
(0.1337, '959175172967292928', 'A cancer 'vaccine' is completely eliminating tumors in mice - New York Daily News https://t.co/NZ2xXqeWYm')
(0.1337, '959175157054099456', 'A cancer 'vaccine' is completely eliminating tumors in mice - New York Daily News https://t.co/Cf2fALp6M1')
(0.1337, '959165358509903873', 'A cancer 'vaccine' is completely eliminating tumors in mice - New York Daily News https://t.co/1Yi4u2Cnbt')
(0.1333, '990179042686111745', 'Leg Length Inequality and Hip Osteoarthritis in the Multicenter Osteoarthritis Study and the Osteoarthritis I
(0.1308, '1000510963047690240', 'I'm: [NEWLINE][U+26AA] gay[NEWLINE][U+26AA] straight[NEWLINE][U+26AA] bisexual[NEWLINE][U+1F518] sorry if I
(0.1307, '975803939089240071', 'brain tumors in young adults zulu fucking https://t.co/GRdpzjR5H1')
(0.1307, '966593857382215680', 'brain tumors in young adults zulu fucking https://t.co/OPagjYGB8s')
(0.1292, '989292078143344641', '@AngelAlessandra sister pretty boy is a stutter er who stutters on the deep strike to straight sex , not on
(0.1291, '961737786607812609', '@GermanLetsPlay Alles Gute... uuund Gesundheit wünsche ich dir :3 (Gesundheit ist, denke ich mal, dat wichtig
(0.129, '960916397181292545', '@dixieldodd @YOURSPINEISMINE "anatomy" point of view: sacral nerve roots below s2 aren't responsible for legs.
(0.1278, '1009867312785252354', 'Using a Ketogenic Diet for Tumors & \xa0Cancer https://t.co/kse5E5XwnG https://t.co/Etg4ZyMoC6')
(0.1272, '964109856054669312', 'Green tumors.[NEWLINE][NEWLINE]Fotos: Gosia Janik[NEWLINE] [NEWLINE]#tallerdefeeas #feeas_shop #madetodislike
(0.1272, '963846540501639168', 'Green tumors.[NEWLINE][NEWLINE]Fotos: Gosia Janik [NEWLINE]#tallerdefeeas #feeas_shop #madetodislike #brokenp
(0.1272, '963391584996937729', 'Green tumors.[NEWLINE][NEWLINE]Fotos: Gosia Janik [NEWLINE]#tallerdefeeas #feeas_shop #madetodislike #brokenp
(0.1272, '963007064783060993', 'Green tumors.[NEWLINE][NEWLINE]Fotos: Gosia Janik [NEWLINE]#tallerdefeeas #feeas_shop #madetodislike
(0.1271, '965540385685680128', 'R.I.P. Salem[NEWLINE]2003 - 19.2.2018[NEWLINE]Aufgrund eines Tumors eingeschläfert https://t.co/egThScYtx7')
(0.1271, '1011376798238760960', '@the_necrosis @Erja_ @annalist Rischtisch... [NEWLINE][NEWLINE]Das Eine, muss dat Andere ja nich ausschließe
(0.1268, '1015663386284175361', 'Ichigo was straight. I knew that. It was pretty obvious from the beginning so yeah. I am not sad about that[
(0.1267, '985220670744399882', '#HSCP2018 Dr. Folpe on adipocytic tumors https://t.co/YzP8xb5vAI')
(0.1262, '1012082847069474817', '@mittelfinger @tantegerd1 Ein Austritt kommt der Entfernung eines Tumors gleich.)
(0.1261, '1013877406308945921', 'Reading this gave me several brain tumors at once https://t.co/EVrcNUNeL8')
(0.1259, '1019340567208251393', '@Reconisan those tumors only provide extra vitamins ..[U+1F440] thank you [U+1F49C]')
(0.1244, '974590688842190848', 'Cannabis Shrinks Tumors and the Government Knew in '74 https://t.co/Nfvzit5Rg7')
(0.1242, '1005596461218910209', '@David_Schraven @nytimes Naja, ganz so einfach ist es nicht. Aber die USA haben unter Obama die Pro-EU-Beweg
(0.1241, '9880600060807057409', '@zorua-su i know right, i can feel tumors growing in my eyeballs')
(0.1241, '984079129279594496', 'vagina tied naked pictures of anal tumors https://t.co/WwkybJ8moj')
(0.1241, '974398867876827136', 'vagina tied naked pictures of anal tumors https://t.co/5GXdYjNaOE')
(0.1241, '974201624515276800', 'vagina tied naked pictures of anal tumors https://t.co/Zz1Aeg3Rt')
(0.1239, '962084307102781441', '@lqjohnjae JSJDHD IT WOULD TAKE ME 10 MONTHS STRAIGHT TO GO BACK TO MY FIRST BIAS TO DO NSD WITH HIM AGAIN')
(0.1237, '985302598310924288', '@Maryam37747449 @littleboiphilip @kittischnitti @dyzzydyzzy Als ob ich jemanden gefunden habe der dat adam un

```
In [16]: for num, key in enumerate(data_index.keys()):
          if num == 85:
              print(num, data[key], '\n', key, data_index[key])
```

85 I'm currently looking for a psychologist bc I need help with my depression/ anxiety and I'd like to,,, talk to someone about it bc i feel
965731052525367296 i m currently looking for a psychologist bc i need help with my depression anxiety and i d like to talk to someone ab

```
In [24]: '''Tweet selection from dictionary'''
          print("Query:", data['965731052525367296'], '\n')
          top = top_100('965731052525367296')
          for i in top:
              print(i)
```

Query: I'm currently looking for a psychologist bc I need help with my depression/ anxiety and I'd like to,,, talk to someone about it bc i f

(0.315, '1005987366186684418', 'tomorrow i have my first appointment at the psychologist, i'm kinda scared but at the same time hopeful that
(0.2972, '1007392417920806912', 'I'm so unhappy with my life currently, I don't know why but I am")

(0.2918, '1020700466408296448', 'I'm so depressed I don't feel like food...')

(0.2779, '1000912376835837952', '01 so it's currently 3:30am in Hamburg Germany and this storm isn't that chill anymore like i just woke up b

(0.2774, '1007047898956271616', '@12Charlotte43 because I don't know anyone of the class I'm currently in and nobody was talking to me last t

(0.2682, '1010988561334751232', 'with each day the want and need to just jump on a train and visit @lipsilhouette is growing because I know t

(0.265, '998990804764028928', 'It makes me so incredibly sad that my parents don't support me when I talk about my future.')

(0.261, '990706608459182080', 'kinda want u back [NEWLINE]kinda fucked everything up [NEWLINE]kinda's not the right thing to do [NEWLINE]kind

(0.2602, '1004405763438206976', 'no matter what i do i'm so fucking unhappy at the moment like nothing is the way i want it to be and i feel

(0.2583, '997583650966396928', 'I'm really pissed and fucked up [U+1F62D]')

(0.2526, '1004965896148963328', 'I'm already nervous enough. I don't need this.')

(0.2524, '984932219629326336', 'I really want to tell my dad tmrw that I don't feel well so we can cancel our plans cause I don't know what i

(0.2512, '1009833171754643456', 'I KINDA WANT TO SHAVE MY HEAD TO LIKE 8 OR 9 MILIMETERS BUT NOW I'M KINDA SCARED HOW PEOPLE WOULD REACT...L

(0.2509, '1002263846919208960', 'I'm getting more nervous help')

(0.2498, '997517935584391168', 'hello i just wanted to say the future is scaring me a lot and i'm very afraid of fucking up my whole life bc

(0.2478, '996400932706451456', 'i'm so fucking nervous and i'm shaking and i feel like i'll throw up and i want to cry')

(0.2468, '966045320491696128', 'I can't get you out of my head[NEWLINE]like a tumor...')

(0.2459, '993577154033090561', 'I am really scared of going to school tomorrow and I don't even know why. I wish I had someone to talk to abou

(0.2459, '1006646063619563520', 'im so nervous for tomorrow :(but i don't need to because i know i can do it')

(0.2449, '1018591122405429248', 'I'm actually kinda scared for my parents to see me at a concert bc idk they don't usually see me as happy as

(0.2436, '1016067247792586752', 'Okhaleesizara I don't know!! It just came like I have SUCH a bad headache and I feel like throwing up the wh

(0.2434, '997966117443395584', 'I'm fucking scared.. I feel like I'm getting into my 2012/2013 phase again')

(0.2433, '998879542579290112', 'I'm just kinda afraid bc I don't wanna smoke alone and also definitely not just with him but I also don't wan

(0.2432, '979858165255888896', 'anxiety is back, i don't want this again')

(0.2428, '100658483664627969', 'I'm so frustrated and I want to talk to someone but I know talking won't solve the "problem" so urghhh')

(0.2427, '984576556038262785', '@ZacharyLevi I'm battling with depression & anxiety + burnout & I'm so sick of it! I want to live & enjoy my

(0.2425, '977687133749235714', 'at days like today where i overthink so much that it makes my anxiety worse and depresses me to the bones, i

(0.2418, '1023664158720643074', 'i'm kinda sad bc i don't know if he has a gf')

(0.2408, '992080043630321665', 'So my Psychologist and i talked about 13RW and He also thinks that it's a harmfull Show and He Said that He w

(0.2407, '1022631754639843328', 'Ostarrynightgray @createthan_ You can always talk to me if you ever feel like it. [U+1F60A] I think it alway

(0.2399, '1004411730351489025', 'I hate it to speak but I also really love it. I love to tell stories. I also really like it to tell someone

(0.2398, '1013524574170906624', '@IovabeI I'm so pissed help')

(0.2397, '995422010133090312', 'Help I'm so nervous. [U+1F62D] #Eurovision #ESC2018')

(0.2397, '993955794893799425', 'I'm so nervous help[NEWLINE]#ESC2018 #Eurovision #AllAboard')

(0.2388, '991465496762601472', 'i hate depression and anxiety can someone kill me')

(0.2381, '1003065805200941057', '@WillTWDG I literally want to die. [NEWLINE]I just don't do it because I feel like I should give myself a lit

(0.238, '1012020796833828871', 'I'm only sad that germany lost bc I had a bet with my maths teacher today and now I'm not getting a better gr

(0.2375, '989529850749882368', 'I finally got my ear pierced today. It was a big step for me bc I was so scared and my parents didn't want me

(0.2371, '1014205255280680962', 'the problem is, that i *cant* talk to them about it, bc theyll be very depressed&start feeling bad bc of me

(0.2371, '1010226360881700864', 'my parents know how fucking pissed I an at them and they know I have a good reason so i'd rather not talk to

(0.2367, '993586230951587840', 'I feel really depressed lately.. I don't know how long i can handle this shit anymore . Just want to lock mys

(0.2362, '1006971728420130817', 'I'm feeling kinda depressed right now and I need someone to talk to, but I can't think of anyone who is in m

(0.2355, '967910869840683008', 'i guess i'm just more sad that i'm not part of that world bc i really would like to i don't even know why')

(0.2351, '98979982993723397', '@wingedwhalien i'm looking forward to one entry in particular, lol. [NEWLINE][NEWLINE]tbh i'm so surprised wit

(0.2345, '1011709678781435904', 'i really want this character drawn' vs 'i'm really scared that it won't look like it looks in my head bc

(0.2328, '101163782700249088', 'Urgh I'm quite nervous bc of Jenadu and I don't like it')

(0.2326, '991039664776720389', '@emWuuu @PTXofficial @scotthoying @mitchgrassi @kirstin @mattsalleemusic @KOlusola Same I'm super nervous bc

(0.2326, '1001914317204226048', 'my life is so boring omg i'm getting so depressed about it')

(0.2324, '975516709120630784', '@OfAnnoyance "First my hip and now my head??? You really want to kill me. Don't you?")

(0.2318, '1008097731229966336', '@sevimxb This might be a little weird, bc I don't really know you, but I'm getting so angry about this ð.ð.

(0.2314, '994306347872194562', 'i'm really afraid to say it because you all know i hate myself and my body but: i feel good about my icon and

(0.2308, '990222601292263426', '@itanimeirl I am kinda happy about this, so I don't need to worry about getting spoiled')

(0.2308, '1001485519002402816', 'I'm so fucking pissed rn don't talk to me')

(0.2306, '962083269633363968', '@sontzuuu @svebara @SerinaTirek @albinokid Don't be embarrassed - if you need something to help you feel bett

(0.2303, '1002200233198215169', '@BTSEuropeARMY @BTS_twt I feel sick and anxious tbh...especially bc I feel that everything is so unorganized

(0.2298, '1014655293328973824', 'Every night after my boyfriend read me to sleep, i wake up again and just lie there, thinking about how this

(0.2277, '980881892202475520', 'i should really start looking into getting an add diagnosis bc i need to know what the fuck is up with me[NEW

(0.2276, '1005184558583730177', 'hello so i'm kinda drunk and also kinda sad bc i kinda miss my crush and i'm upset bc i want him to like me

(0.2271, '99501976645853058', 'I'm so unhappy. cheer me up guys. [NEWLINE]I want to cry but I don't know why [U+1F622]')

(0.2271, '1001205441915912192', '@LoyakkVega I like it')

(0.2267, '981251035825082370', '@Candy__Spice @Pilnut @Nerso54 @VisionaryArt_ @3mpireTimes @Ericof_ @LickMyMexican @Tallboy_100 @alex_ane90

(0.2265, '1009904064782651392', 'Currently I'm really depressed and cry a lot at night... I don't know what's wrong with me but I hope it's n

(0.2264, '1018496654675251200', 'Putting together the corset was surprisingly easy... [U+1F914] [NEWLINE]I'm just afraid I messed up the patter

(0.2261, '995242345560989696', 'I'm actually quite happy that I didn't pay for this shit. I'm already throwing up because of it at home, don't

(0.2253, '1001233780894523392', 'i want it to be june 1st already i want to know if i got the tickets or not i'm so nervous already and it's

(0.2252, '1014545820144848898', '@Herthes_ Same lmao. Like I'm honestly also a lil sad how radical society is with Pedophiles which is why i

(0.2251, '985211247229579266', '#Fuck #War I feel a #shame for my Country! <https://t.co/p5iVGizGay>)

(0.2225, '1014645979415568385', 'I don't know how to describe it. I feel so scared of everything. I feel scared of..leaving my room. I feel s

(0.2225, '1013683136432046080', 'I'm kinda happy to see my parents today [U+1F495]')

(0.2224, '963558209293254666', 'my anxiety has been really getting worse the last days :/')

(0.2224, '1013557956288368643', '@KirschKid Hey, you want to talk about it? I'm also totally depressed right now, maybe we could help eachoth

(0.2223, '980224485436678144', 'I feel like I'm in the worst emotional crisis ever at the moment and idk how to deal with it, it's like depre

(0.2222, '1013204179727331328', 'I'm so scared tbh <https://t.co/cks2k6WDnq>)

(0.2218, '1007358485250695174', 'I'm really nervous about my talk in history class tomorrow [NEWLINE]I feel unprepared and I can't talk in fr

(0.2215, '992433066135126016', '@redheadexgf Yeah maybe but getting there is just eh no[NEWLINE]I get so fucking angry when someone takes my

(0.2213, '960250072809209856', '@ilikeferns No but I feel like it will never happen. I'm so unlucky yk.. plus; with the way I look now I don't

(0.2213, '1021553752472317952', 'his parents are kinda pissed since I'm there so often even tho I livw close [U+1F605] so I sleep in my apart

(0.2213, '1009858963951833089', 'My dad just made me slightly nervous because he started a serious talk about what i want to do after the bac

(0.221, '1007224909834215424', 'I'm so excited for my pic with Camila, but at the same time I'm completely terrified bc I already know that I
(0.2208, '997251513637658625', 'I'm literally crying right now. I feel like such an asshole, but I was so shocked and surprised. It came out
(0.2205, '1014281876171841536', 'I want to see more happy people. If you have the power to make someone happy, do it. Help those who need you
(0.2205, '1014281829657075712', 'I want to see more happy people. If you have the power to make someone happy, do it. Help those who need you
(0.2203, '993239508857606149', 'What makes me really mad is this. Stop lying for attention. I don't think any of my mutuals got a follow bc o
(0.2201, '999077298769944576', '@saara121029 Actually I'm on that journey at the moment. If you want/need someone to talk to, my DM/Line are
(0.22, '1004367376748511232', '@Snibella Seriously, when I was younger I could barely talk in front of people. [NEWLINE]Today I'm always very
(0.2198, '997953232428306432', 'this is fucked up i'm really mad rn")
(0.2198, '1002825583993204736', 'Yesterday my grandpa died and I still don't know how to feel about that.[NEWLINE][NEWLINE]But I'm kinda happ
(0.2192, '996868856579809281', '@vioxixconfused Hahahaha yea we don't talk to strangers really [U+1F602] he asked if it would bother me if I
(0.2189, '1007401378250215426', 'How can a person hate someone so much who literally did nothing to them? I'm so disgusted about that articl
(0.2185, '995350123424165888', 'For me also, to read about it was so hard. I didn't know him but it touched me that someone who is so loved d
(0.2183, '1002399187730911232', 'I don't feel well at all..[NEWLINE]My body hurts so much..[NEWLINE]My head, my stomach, even my heart beats
(0.2179, '1015787333470351361', '@CaptainNwalps @Auron530 You know depression comes with lack of motivation to do ANYTHING at all like it's
(0.2179, '1005633254412750848', 'Does someone want to give me 200 Euros so I can pay the general anaesthetic for my wisdom teeth surgery? I'm
(0.2179, '1002333475620442113', 'So I'm really happy here and I enjoy life, but it's been 3 weeks and I feel like everybody back home is forg
(0.2178, '995410550652899330', '@flightradar24 please don't use red for planes one is currently watching anymore! I was so shocked and worrie
(0.2175, '993646251529981952', 'I'm pretty angry. Not anyone special.[NEWLINE]Rather at me. [NEWLINE]Because i feel like i have lost somethin
(0.2175, '1006629240949886976', 'Btw I'm kinda sad I can't watch the whole PlayStation stream live bc much of it is in the middle of the nigh
(0.2173, '997574887005925376', 'I'm so depressed and I just can't hide it. I'm about to lose control and I think I like it [U+1F60B]')
(0.2172, '1022586412351320068', 'when the past knocks on the door...I'm kinda happy and upset at the same time now...I don't know how to talk
(0.2169, '999388720653656065', 'This is me applying to stuff and suddenly getting super scared [NEWLINE]I'm currently in the cinema and I'm s

Overall we are quite pleasantly suprised by the quality of the results. In most cases there is a strong thematic link between the query text and the most highly ranked tweets.

In the first two queries it is easy to distinguish the importance of a few relatively rare keywords in the results. The terms cervical, cancer and HPV appear to play a very important role in the ranking of results for query one. Japan did as well, but the first occurrence of this word in the ranking is much lower. The second tweet was very short, and each word seems to have played an important role in ranking. Tumor,leg, mccain and straight are all extremely common in the results. Also noteworthy is the fact that we have many short tweets in the results, showing that shorter documents receive higher ranks.The final query was relatively longer and we can see a greater diversity of important terms in the results. Keywords include depression, anxiety, psychologist, currently, and I'm.