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

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.moun

Imports & Setting Configuration

```
1 import re
2 import numpy as np
3 import pandas as pd
4 from textblob import TextBlob
```

```
1 import nltk
2 from nltk.corpus import stopwords
3 from nltk.stem import WordNetLemmatizer
4 nltk.download('stopwords')
```

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
True

```
1 from IPython.display import display, clear_output
2 pd.set_option('display.max_columns', 10000)
3 pd.set_option('display.max_rows', 10000)
```

```
1 lemmatizer = WordNetLemmatizer()
2 stop_words = set(stopwords.words('english'))
```

```
1 # found on stackoverflow @ https://stackoverflow.com/questions/33404752/removi
2 def deEmojify(text):
3     regrex_pattern = re.compile(pattern = "["
4         u"\U0001F600-\U0001F64F"  # emoticons
5         u"\U0001F300-\U0001F5FF"  # symbols & pictographs
6         u"\U0001F680-\U0001F6FF"  # transport & map symbols
7         u"\U0001F1E0-\U0001F1FF"  # flags (iOS)
8                            "]+", flags = re.UNICODE)
9     return regrex_pattern.sub(r'',text)
```

Import data & check it out

```
1 #importing our youtube comments
2 yt = pd.read_csv('/content/drive/MyDrive/data/3.26.youtube')
```

```
3 yt = yt.append(pd.read_csv('/content/drive/MyDrive/data/3.29.youtube'))
4 yt = yt.append(pd.read_csv('/content/drive/MyDrive/data/4.6.youtube'))
5 yt = yt.append(pd.read_csv('/content/drive/MyDrive/data/4.7.youtube'))
6 yt = yt.append(pd.read_csv('/content/drive/MyDrive/data/youtube_march20'))
7 yt = yt.append(pd.read_csv('/content/drive/MyDrive/data/youtube_march21'))
8 yt = yt.append(pd.read_csv('/content/drive/MyDrive/data/youtube_march28'))
```

```
1 yt.head()
```

	Unnamed: 0	Unnamed: 0.1	video_id	text	likes	date	channel_id
0	0	0.0	t59Ge4O70iM	Travel in 2020!!! 😂😂😂	0.0	2020-09-13 16:52:14	UCqUn0dTF2TW80kVyPzsESxg
1	1	1.0	t59Ge4O70iM	If we really went to the moon. They would have...	1.0	2019-12-31 16:24:03	UCEhvCzLOt9oWeWq5xarhUtw
2	2	2.0	t59Ge4O70iM	So will this finally shut the flat earthers up...	1.0	2019-12-31 14:42:40	UCnLxgJ9qvXkbT1iqAhmMB0G
3	3	3.0	AhF44UT2Alk	My school sent me here	0.0	2021-03-09 09:54:23	UC8rZCL_bU2lfjNL_XwdwkkA
4	4	4.0	AhF44UT2Alk	Whos here from mr bowles	0.0	2021-01-05 14:28:02	UC_48oH2diTx602x5T_dLOaw

```
1 yt.describe()
```

	Unnamed: 0	Unnamed: 0.1	likes	Unnamed: 0.1.1
count	482594.000000	226925.000000	482594.000000	146312.00000
mean	75431.468481	52394.775860	6.217614	73155.50000
std	59111.294792	44345.684815	132.582104	42236.78063
min	0.000000	0.000000	0.000000	0.00000
25%	20108.000000	14182.000000	0.000000	36577.75000

```
1 # getting rid of duplicate rows
2 yt = yt.drop_duplicates()
3 yt.reset_index(drop=True, inplace=True)
4 # starting a column to assign reposts to, so we have an equivalent
5 # to retweets
6 yt['repost_count'] = 0
7 # checking how many times each comment occurs
8 columns = {'index':'text', 'text':'count'}
9 txt_count = yt['text'].value_counts()
10 txt_count = txt_count.to_frame().reset_index().rename(columns=columns)
11 # editing our dataframe to add our repost column
12 for i in yt.index:
13   text = yt.at[i, 'text']
14   cnt = txt_count.loc[txt_count['text'] == text]
15   if len(cnt) > 0:
16     cnt = cnt.reset_index(drop=True)['count'][0]
17     yt.at[i, 'repost_count'] = cnt - 1
```

```
1 # importing our twitter data
2 twt = pd.read_csv('/content/drive/MyDrive/data/3.26.twitter', low_memory=False
3 twt = twt.append(pd.read_csv('/content/drive/MyDrive/data/tweets_matt.csv',low
```

```
1 twt.head()
```

```
1 twt.describe()
```

```
1 # renaming our columns so we can add our dataframes together
2 yt = yt.rename(columns={'comment_id':'post_id',
3                         'likes':'favorite_count',
4                         'channel_id':'user_id'})
5 twt = twt.rename(columns={'retweet_count':'repost_count'})
6 for i in [yt, twt]:
7   i = i[['text', 'favorite_count', 'user_id',
8          'mentions', 'repost_count', 'post_id']]
```

```
1 # starting a count and an empty list for gathering new hashtags
2 count = 0
3 empty list = []
```

```python
1 # look at the hashtags, see if there are any outliers to get rid of
2 for item in twt.hashtag.value_counts().to_frame().reset_index()['index']:
3     if count < 250:
4         hashtags = str(item).split()
5         for item in hashtags:
6             empty_list.append(item)
7             count = count + 1
8
9 # we don't want to count our original hashtags, so they're defined here
10 items = ['jaxa', 'esa', 'curiosityrover', 'areospace',
11         'internationalspacestation', 'JAXA', 'astronomy',
12         'oppertunityrover', 'virgingalactic', 'universe',
13         'sls', 'Starship', 'climate', 'starship', 'virginorbit',
14         'nasa', 'cosmos', 'mars', 'falconheavy', 'NASA',
15         'futurism', 'starliner', 'iss', 'spacex', 'falcon9',
16         'nasa_app', 'roscosmos', 'Roscosmos', 'blueorigin',
17         'ESA', 'spacetravel', 'artemis', 'marswebcam', 'starlink']
18
19 fresh_hashtags = []
20
21 for item in empty_list:
22     if item not in items:
23         fresh_hashtags.append(item)
24
25 print(set(fresh_hashtags))

    {'Mars2020', 'JUNMYEON', 'Venera3', 'StanWorld', 'FalconHeavy', 'archaeology', 'persever
```

```python
1 # since we don't have the ends of some tweets due to the api cutoff
2 # and hashtags usually falll at the end, to make sure we have them all
3 # we're going to check if they're in our sentences and if not add them
4 for i in range(len(twt)):
5     if type(twt.at[i, 'hashtag']) == float:
6         continue
7     else:
8       split_ = twt.at[i, 'hashtag']
9       for item in split_:
10        if type(twt.at[i, 'text']) == str:
11            twt.at[i, 'text'] + ' ' + item
```

```python
1 # merging our dataframes
2 df = twt.append(yt)
3 df = df.drop_duplicates(subset='post_id')
```

```python
1 # unsure how a handlful of strings ended up in our favorites but dropping
2 df.reset_index(drop=True, inplace=True)
3 for i in range(len(df)):
```

```
4    if str(df.at[i, 'favorite_count']).isnumeric() != True:
5       df.drop(index=i, inplace=True)
```

```
1 # get rid of duplicate comments & retweets
2 df = df.sort_values(by='favorite_count', ascending=False)
3 df = df.drop_duplicates(subset='text', keep='first')
```

```
1 # much better
2 df.head(10)
```

```
1 df.describe()
```

```
1 # making a dictionary containing every contraction and acronym to replace
2 replacements = {
3                 "ain't": "have not",
4                 "aren't": "are not",
5                 "can't": "cannot",
6                 "can't've": "cannot have",
7                 "'cause": "because",
8                 "could've": "could have",
9                 "couldn't": "could not",
10                "couldn't've": "could not have",
11                "didn't": "did not",
12                "doesn't": "does not",
13                "don't": "do not",
14                "hadn't": "had not",
15                "hadn't've": "had not have",
16                "hasn't": "has not",
17                "haven't": "have not",
18                "he'd": "he had",
19                "he'd've": "he would have",
20                "he'll": "he shall / he will",
21                "he'll've": "he shall have",
22                "he's": "he is",
23                "how'd": "how did",
24                "how'd'y": "how do you",
25                "how'll": "how will",
26                "how's": "how is",
27                "I'd": "I had",
28                "I'd've": "I would have",
29                "I'll": "I will",
30                "I'll've": "I will have",
31                "I'm": "I am",
32                "I've": "I have",
33                "isn't": "is not",
34                "it'd": "it had",
35                "it'd've": "it would have",
36                "it'll": "it will",
37                "it'll've": " it will have",
```

```
38              "it's": "it is",
39              "let's": "let us",
40              "ma'am": "madam",
41              "mayn't": "may not",
42              "might've": "might have",
43              "mightn't": "might not",
44              "mightn't've": "might not have",
45              "must've": "must have",
46              "mustn't": "must not",
47              "mustn't've": "must not have",
48              "needn't": "need not",
49              "needn't've": "need not have",
50              "o'clock": "of the clock",
51              "oughtn't": "ought not",
52              "oughtn't've": "ought not have",
53              "shan't": "shall not",
54              "sha'n't": "shall not",
55              "shan't've": "shall not have",
56              "she'd": "she had ",
57              "she'd've": "she would have",
58              "she'll": "she will",
59              "she'll've": "she will have",
60              "she's": "she has ",
61              "should've": "should have",
62              "shouldn't": "should not",
63              "shouldn't've": "should not have",
64              "so've": "so have",
65              "so's": "so as ",
66              "that'd": "that would",
67              "that'd've": "that would have",
68              "that's": "that is",
69              "there'd": "there had",
70              "there'd've": "there would have",
71              "there's": "there has",
72              "they'd": "they had ",
73              "they'd've": "they would have",
74              "they'll": "they shall",
75              "they'll've": "they shall have ",
76              "they're": "they are",
77              "they've": "they have",
78              "to've": "to have",
79              "wasn't": "was not",
80              "we'd": "we had",
81              "we'd've": "we would have",
82              "we'll": "we will",
83              "we'll've": "we will have",
84              "we're": "we are",
85              "we've": "we have",
86              "weren't": "were not",
87              "what'll": "what shall",
88              "what'll've": "what shall have",
89              "what're": "what are",
```

```
90          "what's": "what has",
91          "what've": "what have",
92          "when's": "when has",
93          "when've": "when have",
94          "where'd": "where did",
95          "where's": "where has",
96          "where've": "where have",
97          "who'll": "who shall",
98          "who'll've": "who shall have",
99          "who's": "who has",
100         "who've": "who have",
101         "why's": "why has",
102         "why've": "why have",
103         "will've": "will have",
104         "won't": "will not",
105         "won't've": "will not have",
106         "would've": "would have",
107         "wouldn't": "would not",
108         "wouldn't've": "would not have",
109         "y'all": "you all",
110         "y'all'd": "you all would",
111         "y'all'd've": "you all would have",
112         "y'all're": "you all are",
113         "y'all've": "you all have",
114         "you'd": "you had",
115         "you'd've": "you would have",
116         "you'll": "you shall",
117         "you'll've": "you shall have",
118         "you're": "you are",
119         "you've": "you have",
120         " u " : " you ",
121         " w " : " with ",
122         " poc ": " person of color ",
123         " bpoc " : " black person of color ",
124         " bipoc ": " black indigenous people of color ",
125         " lgbt " : " lesbian gay bisexual transgender ",
126         " lgbtq ": " lesbian gay bisexual transgender queer ",
127         " lgbtqi ": " lesbian gay bisexual" +
128         " transgender queer intersex",
129         " lgbtqia ": " lesbian gay bisexual" +
130         " transgender queer intersex asexual ",
131         " omg": " oh my god ",
132         " lol ": " laughing out loud ",
133         " lmao ": " laughing my ass off ",
134         " lmk ":" let me know ",
135         " icymi ": " in case you missed it ",
136         " til ": " today i learned ",
137         " omfg ": " oh my fucking god ",
138         " iirc ": " if i remember correctly ",
139         " ppl ": " people ",
140         " ppls ": " peoples "
```

```
141              " yts ": " white people "
142              " yt ppl": " white people"
143              " yt people": " white people"
144              " yt girl": " white girl"
145              " yt boy": " white boy"
146              " yt men": " white men"
147              " yt fragility ": " white fragility "
148              " yt mf": " white motherfucker"
149              " please ": " please "
150              " rn ": " right now "
151              " twt ": " tweet "
152              " pov ": " point of view "
153              " btw ": " by the way "
154              " srsly ": " seriously "
155              " fr ": " for real "
156              " tho ": " though "
157              " rlly ": " really "
158              " rly ": " really "
159              " ur ": " your "
160              " ai ": " artificial intelligence "
161              " ure ": " youre "
162              " jfc ": " jesus fucking christ "
163              " smh ": " shaking my head "
164              " smfh ": " shaking my fucking head "
165              " idiic ": " infinite diversity in infinite combinations "
166              " llap ": " live long and propser "
167              " idk ": " i dont know "
168              " idc ": " I dont care "
169              " ftw ": " for the win "
170              " tbt ": " throw back tuesday "
171              " n ": " and "
172              " em ": " them "
173              " oh em gee ": " oh my god "
174              " ngl ": " not gonna lie "
175              " b ": " be "
176              " hmu ": " hit me up "
177              " yt millennials ": " white millennials "
178              " yt character ": " white character "
179              " yt ": " youtube "
180              " tf ": " the fuck "
181              " af ": " as fuck "
182              " asf ": " as fuck "
183              " sm ": " so much "
184              " irl ": " in real life "
185              " aka ": " also known as "
186              " jsyk ": " just so you know "
187              " doe ": " though "
188              " dey ": " they "
189              " iwd": " international womens day "
190              " pos ": " piece of shit "
191              " mf ": " mother fucker "
192              " boi": " boy "
```

```
193                 " w e ": " whatever "
194                 " wtf" : " what the fuck "
195                 " andamp " : " and "
196                 " tfw ": " that feeling when "
197                 " tl ": " timeline "
198                 " l8r ": " later "
199                 " im ": " i am "
200                 " id ": " i had "
201                 " youre ": " you are "
202                 " theyre ": " they are "
203                 " weve ": " we have "
204                 " ikr ": " i know right "
205                 " yall ": " you all "
206                 " isnt ": " is not "
207                 " dont ": " do not "
208                 " fyi ": " for your information "
209                 " cant ": " can not "
210                 " wasnt ": " was not "
211                 " tbh ": " to be honest "
212                 " andamp ": " and "
213                 " & " : " and "
214                 " yt " : "youtube"
215                 }


  1 # after thouroughly examining the data, we found a number of oddities that
  2 # certainly dont belong, like slurs, pornographic content, dog photos
  3 # we have them all defined so we can get rid of them here
  4 list_to_drop = [' supaspicyy ', ' snyder', ' avenger ', 'persona',
  5         ' ghosting ', ' incel ', ' internetofblockchains ',
  6         ' ot8 ', 'jar jar', ' pineapple on pizza ', 'side chick',
  7         ' democrat ', ' prem ', ' remix ', ' fkn ', 'suppoawards',
  8         ' livestream ', 'hanamiiart', 'isolatiempo',
  9         ' my coffee this morning ', ' exo kai ', 'demoncrats',
 10         ' coivd 19 pandemic', 'spotify', 'sailormoon', 'chyna',
 11         'check it out later dylan ', 'stonkmarket', 'oprah',
 12         ' c 19', ' love your video', ' domestic au', ' sfw ',
 13         ' shinee ', ' reggae ', ' undertale ', 'transmental',
 14         'hogwarts', 'kakoadrawsstuff', 'ghhiphophubjoost',
 15         ' english only ', 'streamtwice', ' avengers ', 'miningbc',
 16         ' lady gaga ', ' dnda ', 'dodgecoin', ' hello please ',
 17         ' freestyle ', ' bollywood ', ' destiny tl ', ' earth sign',
 18         ' auschwitz ', ' an aries ', ' floof ', 'byunbaekyoong',
 19         'getbtc', ' gif ', ' pythonbot ', 'star wars',
 20         'downloaderbot', 'astrology', ' 1d ', 'bigdaddyxxi',
 21         ' st  patricks day ', ' nipple ', ' dadbod ', ' junichiixo ',
 22         'dallasparkrec', 'minsung', ' jefferson starship ',
 23         ' counterculture ', 'superjunior', 'page529',
 24         'guardiansofthegala', ' acnh ', 'my gf', ' learn code ',
 25         'realtorlife', ' mnet ', 'justice league', ' caucasian',
 26         'donsbot', ' prince charles ', 'fakeyogurtfacts',
 27         'medicalcannabis', ' djs ', 'slytherin', ' brexit ',
```

```
28      ' buy 1 get 1 free', 'miteyuk', 'meekonmingmong',
29      ' pogg', ' tpc ', ' amc ', ' chu ', 'the8films',
30      ' retard ', ' my cat ', ' ford motor company', ' au update',
31      ' i just remembered ', ' icc ', 'tenfictions',
32      'idrawcharts', ' snyd ', 'space tv', 'kaliuchis',
33      ' choco pie ', 'australianart', 'bytesofbooks', ' yeaji ',
34      ' hot wife ', ' cancel culture ', ' youtube show',
35      ' cbdhage ', 'paragliding', ' i love you', ' funsized ',
36      'ladyboy', ' game', ' st patrick ', 'kirk', 'my mom',
37      ' doctorwho ', ' good friday ', 'got7', 'challengertask ',
38      ' shared on twitter ', ' out of character ',
39      ' fucked me up yo ', ' gaga ', 'artrabbit', ' norse ',
40      'definitelyfollow', ' the pandemic ', 'potterheadposts',
41      ' ellle em ', ' anarcho ', ' mcu ', 'universe is with me ',
42      'shouldhaveknownitwasascam', ' 8d ', ' basketball ',
43      ' on zoom ', 'illegal alien', ' top 10', ' piday ',
44      'ffvlyricsbot', 'serena williams', ' chauvin ', ' sk8 ',
45      ' leovil ', 'canon', 'lizard people', 'blender3d', ' selfie ',
46      ' twink ', ' prank ', 'gundamartemis', ' get popcorn ',
47      ' cassie and artemis ', ' net developer ', ' eddieizzard ',
48      'southern  border', 'babyboy', 'canadanewswire', ' alchemy ',
49      ' offseason ', ' au sommet ', ' game of thrones ', ' 9 11 ',
50      'lavigne', 'forsale', 'vanlife', ' blm ', ' webtoon ',
51      'hey fellas', ' good morning ', ' mac app ', 'tusculumsports',
52      'streetartsheff', 'italianostyle', 'suppoarea',
53      'dogagency', 'crazyforjoonie', ' handsome ', 'jaemjen',
54      ' new challenger appro', ' pound dumbbell ', ' ps4 ',
55      ' crossover ', 'hobbyconnector', ' klingon',
56      ' you dont know me ','africafactdaily',
57      ' international womens day ', 'euchre', ' au where ',
58      'we kissed', 'geminis ',
59      ' sk ', ' 911 ', ' discord servers ', 'brb', 'minimalisthyuk',
60      ' starship  au ', ' ftp ', 'cravitystarship', 'wheephoric',
61      ' new virtual fieldtrip ', ' antarctic survey ',
62      ' st first contact', 'plandemic', ' horse teams ',
63      ' seminar ', ' doggg ', ' tony hawk ', ' follows today ',
64      ' como ', ' tag 10 ', ' leo sun ', 'nintendoswitch',
65      'spacecoin', 'lawlietslolita', 'onlinecasino', ' patreon ',
66      ' senator ', ' baby momma ', 'bangtwicedaddyyungjoc',
67      'ariana grande', 'actualidadspacex', 'fandom', 'artpal',
68      'notamoviebuff', 'tony the tiger', ' pokemon', ' movie ',
69      ' sblv ', 'elysium', 'monchpublishing', 'witchinnreadin',
70      'trigger warning', ' otp ', 'happybirthday', 'hyeown',
71      'domainnameforsale', ' podcast ', ' $btc ', ' concert ',
72      ' moonbin ', 'no other grape', ' drink and a snack ',
73      ' meghan markle ', 'north korea', ' albums ', 'sailorfemme',
74      ' boba ', ' followed me ', ' toe suck', 'izone',
75      'planetfitness', ' sailor saturn ', ' waluigi ', 'vlogger',
76      'shaun vid', ' batman ', '99memes', 'walmartemployees',
77      ' drag race', ' codenewbies ', 'ps5s', ' kike ', 'sex appeal',
78      ' amazons bestselling ', 'ontariohealthc', 'yujin',
```

```
79          'funnyanimalbook', 'tsla', ' plot hole', ' bittorrent ',
80          ' giveway ', 'giveways', ' bucs ', ' stock ', 'gamingwithlance',
81          ' skywalker ', ' jedi ', ' endgame ', ' wandavision ',
82          ' girlgamer', ' celeste ', 'philzabot', 'includesartemismiss',
83          'taylor swift', ' my girlfriend ', ' gogh ', 'lyricbot',
84          ' stfu ', 'coinswitchkuber', ' nfts ', 'orbitalpodcast',
85          ' gypsy ', ' sherlock ', ' linkedin ', 'mythandfact',
86          'love the video', ' time travel au ', 'xiumin', ' boob ',
87          'suppostreamers', ' kid cudi ', 'mrbeast', 'yourdailytds',
88          ' papa ', 'realtorneworkng', ' pinche ', 'newsparodypk',
89          'mafia4rtemis', 'i have a crush', ' a taurus ', ' king kong ',
90          'submissivetraining', ' suga ', ' greek ', 'thelegendof',
91          ' yosemite ', 'new song', ' paddy', ' bestie ', ' vibe co',
92          'mutualaidrequest', 'haikyuusalbot', 'grumpyexpolice',
93          ' the winner of the competition ', ' bien ', 'join us tonight',
94          'moonchild', 'ogleyuk', 'starship ent', 'busty', ' k pop ',
95          ' 40k ', 'pockyoong', ' heckin ', 'nicki minaj',
96          ' chat tomorrow ', ' junhui ', ' danc', ' got me fucked up ',
97          ' fan union ', ' ebooks ', 'mp4', 'dreamchaewon',
98          'doe ever drive crazy fast night change', ' pi day ',
99          'crackhore2', ' overwatch ', 'space bun', 'yeongyu', '30stm',
100         ' italian mafia ', ' passed away yesterday ', 'twinsies',
101         'jongdae', 'seoho', 'pokmon', 'patrickstar', ' d r i n k  ',
102         'dprk', ' tron community ', 'sugacology', ' anime ',
103         ' ninetendo ', ' backup dance', ' otd ', ' bruno the goat',
104         ' drink and snack', ' rap ', 'club vibe', ' the titans ',
105         ' takki ', 'startrek lower decks', ' milf ', 'my song',
106         'exo member', 'world tour ', 'bakescookies4u', ' wonwoo ',
107         'portlandprotest', 'k followers', ' scandal ', ' nfl ',
108         ' uni au ', 'newhollandbrew', ' ipad ', ' soloist', 'libtard',
109         ' staan ', ' mars in gemini ', ' ggs ', ' woollim ',
110         'cutiesaturday', 'griffindor', ' feature on twitter ',
111         ' bidao ', ' blog pos', ' hyun ', ' ssbu ', ' my dog ',
112         ' girlsoftwitch ', ' pls rt', 'promoteamgamers', ' katalk',
113         ' tuning in tonight', ' childrens book', 'telepointo',
114         ' pinkmoontarot ', ' proofread my brothers ', ' eoy ',
115         ' graphic t ', ' his voice', ' mummy ', ' kanye ',
116         'lunarcrush', ' mew ', 'superiorartemis', ' trabajo',
117         'kokobop', ' service animal ', ' incest ', ' skinny jean',
118         'ootd ', 'anamnesisstudio', ' cardi ', 'sethmeyers',
119         'his pic', ' series e tournament ', ' monsta ', 'yedam',
120         'shemale', ' slashfic ', ' tienes ', ' joe manchin ',
121         ' ugly crying ', ' stans ', 'ebot ', ' gop challenger',
122         'unovan', ' breaking news', ' de ', ' watermark ',
123         ' exo vibe', ' painter ', 'spell check', ' maga ',
124         'tokentizedtweets', ' fanfic ', ' dodgers ',
125         'woman is killed', 'buffjjaeng', ' playlist ', 'jeonging',
126         ' uber ', ' a pices ', 'joonsroom', ' hyde ', ' whore ',
127         ' senpai ', ' dj ', ' my office ', 'eeveelution', ' shran ',
128         ' youre so gorgeous ', ' baal', 'ujunghaeyo', ' baby girl ',
129         'rantarxamami', ' mommy ', ' data leak ', 'moondustyoon',
130         ' cr4vity ', 'hg2films', ' pfp ', ' vuelve ', 'for exo',
```

131	'huehue', ' snl ', 'lovato', ' channeling these vibes ',
132	'askabureson', ' nightvale', '911milfs', ' bois ', ' an esa ',
133	' chanyeol ', ' exo perform ', 'ksjnluvbot', 'takeonthegame',
134	' me gusta ', ' my bf ', ' video game ', ' magic spell ',
135	' spacex teq', ' sonic the hedge', ' terf ', 'suppodepartment',
136	'pokeballs', 'bearschoice2020', ' tits ', ' shadow the hedge',
137	' air sign', 'pajeet26nazinazi', ' musician ', ' video recap ',
138	'youngboycharts', 'k iss', ' baby cakes ', ' qtum ',
139	' plot holes ', 'fanart', 'leagueoflegends', ' vtube',
140	'webanime', ' romluan warbird ', ' cecil hotel ', ' pissbab',
141	'desiraiart', ' t4t ', ' au we all need', ' mfs ',
142	' influencer ', ' instagram live ', ' a new article in no ',
143	' ranbo', 'personal space', 'marchofrobots', ' you won ',
144	' furry ', ' water sign', ' arachno futurism ', ' covid',
145	' dank ', ' ethererum ', 'seuljhi', ' desis vibe ',
146	' horror movie', ' precious mom ', 'famedcelebrity',
147	'incantationsoverwater', 'join us at', 'fuckboy', ' jongin ',
148	' the god ', ' seokjin', ' follow train ', ' challenger pack ',
149	' scam call ', 'highschool au', 'rick and morty',
150	'asminternatnl','justinbieb',' p l a y ',
151	'dashboardconfessional', ' lads ', 'ramtrucks', ' stanning ',
152	'bad attitude ruin ', 'ig live', ' oc ', ' ugg ', 'dabak',
153	' artemis twt ', ' smashbrosultimate ', ' gaming ',
154	' jinhoon', ' mercedes benz ', ' champion show ',
155	' yung ', 'whatsupgurgaon', ' misogynist ', ' rape ',
156	'minsungbingo', ' superhero ', 'download the ap', 'stockpro',
157	' hey babe ', ' internationalwomensday ', ' spacexcountdown ',
158	'jaechan', ' popcorn spell', 'donaldtrump', ' no context',
159	'6ix9inebtcpromo', 'bruno mars', ' mama ', ' fingering',
160	'immigrantsnews', ' burger king ', 'bangers only', ' bd ',
161	' sun moon rising ', ' fancam ', ' fan page', ' my toes ',
162	' facist ', ' saint ', 'seungcheols', 'toryincompetence',
163	'exol dks', ' titty ', 'y00nggas', 'stoner', 'ilor mercury ',
164	'happyeasygo', ' commie ', ' audition ', ' vader ',
165	'hufflepuff', 'smash ultimate', ' gme ', ' sextile ',
166	'pelosi', ' fifa ', ' starlink ', 'dateless', 'love geraniums',
167	' au fic ', 'evertonschool', ' edgy ', 'leaked photo',
168	'jewish space lazer', ' itunes ', 'my commissions',
169	' football ', 'faggot', ' pandora ',' cishet ',
170	'jewish space laser', 'fanartfriday', 'bbhlyuv', 's day ',
171	' sailor moon ', ' 42069 ', 'ohmygirl', ' hot 100 ',
172	' delivery fee ', ' blockchain', ' follow back', ' bitch ',
173	' teleheath ', ' if there were two guys on ', 'butthole',
174	'yvesjeong', ' donghyuk', ' women are killed ', 'disneyplus',
175	'icone', 'rebranded', 'mawandegraphics', 'happening now ',
176	' zaddy ', 'herecomesshawty', ' police corrupt', ' stimmy',
177	'minecraft', 'toploadersports', ' followers please', ' rin ',
178	' followforfollow ', ' nationalpublicdefenderday ',' zionist ',
179	'transgenders', ' esta ', ' exokai', ' march is ', 'accent',
180	' ya boi ', ' leto ', ' pos coin', ' to code ', 'kylieart3',
181	' yoda ', 'breesunshinee', 'headcannon', ' whatsapp ',
182	' spell weaver ', ' page au ', 'i made out with ', 'ease rt ',

```
182    spell weaver ', ' page au ', ' i made out with ', ' ease it ',
183    ' lin manuel ', 'onlyfans', 'weyoun', ' registered esa ',
184    'vmineverytime', ' thumbnail ', ' stoner', 'itsshowtimena',
185    ' fck', ' my holdings', ' epiosde ', ' instagram ',
186    'humanesociety', 'song', ' zhongchi ', ' blockchain ',
187    ' btob ', ' bet wa ', ' cinnamon tea ', ' vibrator ', ' bnb ',
188    ' my dad', ' abt abuse ', 'sumcoin', ' hades ', 'lupefiasco',
189    ' cum ', ' dm me ', ' rapist ', ' waifu ', ' cravity ',
190    'chrysler', ' fancon ', 'hyoshin', 'releaseofliquidators',
191    ' iwd', 'chipsmutcorner', ' itsyourmoment ', 'mp3',
192    ' alexandria ocas', ' mirrorverse ', 'stanislav', ' jk ',
193    'cygames', ' rookie group ', ' henna ', 'btsarmy', ' yolo ',
194    ' stan ', ' history month ', ' 35mm film ', ' in exo',
195    ' super junior ', ' praise god', 'magnesium for anx',
196    ' hi ', 'choose character', 'mailsport','tedcruz',
197    'sn11 almost ready to fly ', ' dc uni', ' of the week',
198    ' dnd ', ' woke mob ', ' grader', ' stonline',
199    ' an aquarius ', ' roblox ', ' set your alarm',
200    ' cameo ', 'transgendered', 'sayhername',
201    ' reality design shows ', 'genre', ' boy group ',
202    'armyrotc', ' xunqi', ' tip4da4', ' shino ', 'exhibitabrewing',
203    'furnituregurus', ' koreab00 ', ' dls ', ' binance ',
204    'doggo', 'orcaowlart', 'jaldi', '  did lockdown ',
205    ' top track ', ' lola and maya ', ' top 5 ', ' f4f ',
206    ' music video ', 'bobsburgersfox', 'criminaltiktoks',
207    ' kai exo ', ' rts ', 'paramoremusicom', ' shit post ',
208    ' yedam ', ' realistic arm', ' side part', ' friendsofkrispy ',
209    ' prime timeline ', 'cobra starship', ' std ', ' cat food ',
210    'philwademusic', ' an yujin ', ' disney ', ' sl dorr ',
211    ' footbal ', ' dm me for info ', ' au contraire', ' comedy ',
212    'btfoshizzle', ' by exo ', 'jumpoon', ' st artemis ',
213    'comedybangbang', ' cardassia ', ' spoilers ', ' babymomma',
214    ' when exo ', ' cj enm ', 'theestallion', ' what is grammar ',
215    ' besties ', ' your own vibe ', 'justice4tamilgenocide',
216    'mindsofanime', ' sountrack ', ' we are hiring ', 'lola bunny',
217    ' bb0un ', ' click link', ' sucking d', ' wonho ', 'bananocoin',
218    ' exo history', ' nike ', ' yuehua ', 'kroll show',
219    ' artemis fowl ', 'gayshit', 'antiasian', ' i woke up ',
220    'proudofjungkook', ' rodimus ', ' rpg ', 'massbuildergame',
221    ' sa ', ' cop burial ', ' hot fire test', ' isp ', ' yer ',
222    ' pic of me ', ' thanos ', 'fleshlight', 'inthemoonlight',
223    ' white cat ', ' firstcontactday ', ' startrekonline ',
224    'vintage4moms', 'dblyrajsifonvhumgvo68ecys39dlcfc5e', ' dox ',
225    ' wap ', 'vampire', ' mythology ', ' purplekiss ', 'vocalist',
226    ' heck ', ' tiktok ', ' catering company ', 'in2deeprecords',
227    ' starship enterprise ', 'roddenberry', 'wonyoung', '9doodles',
228    ' yikes ', ' high end pc ', 'fleshnft', ' student au',
229    'bohemiandlive', 'hyunjujae', 'harry potter', ' yaoi ',
230    ' myanmar military ', ' aint kissed ', ' meme ', ' que ',
231    ' hiphop ', ' existential crisis ', ' stanned ', ' eth ',
232    ' ig ', ' ufc ', 'insomniacgames', ' my ex', '420 dabs',
233    ' infrastructure bill', ' senate ', ' nft ', 'ustedes',
```

```
234        'bossladybtarot', ' abortion ', 'theweeknd', 'geepoetry',
235        ' owo ', 'unicefso', ' lana del ', 'larrytruther',
236        ' championship ', 'riker', ' lyricists ', 'writeconnectllc',
237        ' reylo ', ' gap teeth ', ' the bulls ', ' vpn ', 'otelcoinc',
238        ' i fucked ', ' cyberpunk2077 ', 'weareoneexo', 'betikake',
239        'grammy', 'kawaiivibes4', 'ateez', ' exo breathes ', ' volta ',
240        'digitalarts', '11x16', ' easter ', ' wizard ', ' qb ',
241        'thestuffofmemes', 'casualxtopher', ' apj ', 'smoke weed ',
242        ' trustwaller', ' emmy', 'jarirbookstore', ' snowflake ',
243        ' alien franchise ', 'douglas adams', ' a scorpio ', ' stoned ',
244        ' cunt', ' tabloid', 'esports ', ' steelers ', 'proudboy',
245        'artemisstanding', ' fisheye lens ', ' yashiro ', 'party vibe',
246        ' btc ', ' prime minister ', ' shitcoin ', ' sjw ',
247        'startrek', ' acab ', ' monster ', 'boysw', ' sia ',
248        ' syllabus ','sayhisname',
249        'selenaasslut', ' bts ', 'hypersexualiz', ' subs ', '4biden ',
250        ' worldhealthday ', 'squishyunoh', ' coinapp ', '1bitcoinash',
251        'universe is god ', ' guided meditation ', 'ship troopers',
252        'picture of the day', 'tr76000670100000076380895',
253        ' fire sign', ' pixar movie ', 'chartdata', ' contact day ',
254        'cryptocurren', ' my youtube', ' sexc ', 'sunnycheungky',
255        'wildriftvideos', '4chan', 'fanboys', 'jungkookers',
256        ' hitler ', ' bebe ', ' baby boy ', 'saytheirname',
257        ' you exo ', ' brady ', ' star crossed ', 'exo gsd',
258        'confomovie', ' stabeing ', 'yayabuzzgame', ' psth ',
259        ' burgerking ', 'startrekrp', ' video gam', ' safe space ',
260        ' muggle', 'au writer', '666666', 'photocontests', '80x25',
261        ' fineart ', 'senqyoun', ' kelvin timeline ', ' slayed ',
262        ' gmb ', ' karma ', 'tyler the creator', 'bibliography',
263        'frick ', ' asmr', ' ad ', ' exo leaving sm ', ' jets fans ',
264        ' shitpost', 'cinematic universe', ' 8th house ', 'ur gay',
265        ' your handle', 'harley quinn', ' playoff', 'homeschoolfun',
266        ' bik3r boi ', ' st pa', ' april fool', 'jesus christ',
267        ' morning vibes ', 'seattlesurge', ' super bowl ', ' sith ',
268        ' industry night ', ' doink ', 'jongin','jaxdotcom',
269        ' nobody  absolutely nobody ', ' a saggitarius ', ' s n a z ',
270        ' super mario ', ' brittania ', ' miley cyrus ', ' bandom ',
271        'erotica', ' wtnv ', ' dabbing dad ', ' y i k e s ',
272        ' code jam', ' first comment ', 'mansion', ' knock knock ',
273        ' i just got ', ' werewolf ', 'letsnamjoon', ' cosmetology ',
274        ' sim ', ' men aint shit ', ' merch ', ' wjsn ', 'nydevito',
275        ' best spaghetti ', ' fics ', 'ust uploaded ', ' ratchet ',
276        'didyouknownfact', ' i just wanna ', 'seokgrls', 'fotografia',
277        ' kong ', '10th grade', ' 9gag ', 'witchesof', 'livetweet',
278        'bajenge', 'arrested', 'traytaymakeup', ' learn to spell',
279        ' detects deepfakes ', 'club  price', ' barking ',
280        'dongmincheeks', ' ppl be fucked', 'billboards',
281        ' centre for seismology ', 'face mask', 'workingfromhome',
282        'rasicsmisnotcomedy', 'superframemedia', ' halsey ',
283        ' artemis au ', ' this character ', ' new lego ',
284        ' women killed', ' lucky to have each', 'artminx',
285        ' orville ', 'postmates', ' lyricist ', ' wts ',
```

```
286        'loomingartemis', ' barista au', 'bluemoonstudios',
287        'irishnewsbiz', 'on twitch ', 'dickpic', ' bitcoin ',
288        'teslaowners', ' birthday ', ' human au ', ' singer ',
289        ' go live ', 'horny', ' chicken ad ', ' nutted ',
290        ' au portrayal ', ' solo art', 'newlisting', ' natal ',
291        ' a virgo ', 'cherrykissjoon', ' corn syrup ', ' a leo ',
292        'fanfic', 'yerenewenjoygi', ' b0mb ', ' kink ', ' gg ',
293        'sun moon rising', ' kehlani ', ' babcock ', 'shatner',
294        ' au sketch', ' mingi ', 'hatecrimebill', ' idol ',
295        ' your download ', 'wonrobot', ' 30 seconds to mars ',
296        ' tesla stock ', ' midget ', ' april is', 'nvslive',
297        ' the devil ', ' a libra ', 'dhwani', 'jared leto',
298        ' findom ', ' comcast ', ' riverdale ', ' woman day ',
299        ' notstakedtokens ', 'familyhistoryandgenealogyseries',
300        ' like tweet ', ' vagina ', 'twitchnetwork',
301        ' dodge challenger ', ' yg ', ' midweekliving', ' ass ',
302        'newyorkerunion', 'treasuremembers', ' kitkat ', ' bitter ex',
303        'four seasons la', 'middaycontest', 'guinness world',
304        '  breaking  ', ' rapper ', ' sl character ', ' dawg ',
305        'jam magno ', 'latenightmovie', 'couchbrewery',
306        ' car clogged la ', ' idgaf ', 'stray kids', 'k the pol',
307        ' spot the difference', 'cheemsburbglar',
308        ' exos special appearance ', ' fandom ', ' soundcloud ',
309        ' deffo', ' esa dog ', ' yoshi ', ' hi hi ', ' domme ',
310        'bodybuilding', ' kawaii ', ' tucker carlson ',
311        'villagers in russia', 'bitches', ' sims ', ' cuck ',
312        'magicschoolbus', ' iphone ', 'pixielatedesign', ' cleopatra ',
313        ' warhammer', 'ex0atmospheric', 'killerklowns', ' bruno mf ',
314        ' twittersupport ', ' a gemini ', 'selfpromosaturday',
315        ' sk 8', ' puta ', 'direct message me', 'cb4yr', ' iost ',
316        ' q iss ', 'enhypen', 'realestateagent', ' godzilla ',
317        ' nobody absolutely nobody ', ' my gf ', ' mlp ', ' lyrics ',
318        'qosicart', ' gay space communism ', ' ootd ',
319        ' hardest jobs in the universe ', ' ive fucked ', '5sos',
320        ' hoe ', ' this footage ', ' fag ', 'pizza delivery guy',
321        ' my mother', ' insta ', 'jazzersize', 'ravenclaw',
322        ' dio ', ' tw ', 'mostrologyy', 'nodegree2dev',
323        'this drawing', ' pixel 5 ', '13b603b66', 'kpop',
324        ' clowning ', ' puns ', 'my aesthetic vibe',
325        ' ufotwitter ', 'sookramsbrewing', ' cloud space ',
326        ' exos ', ' starshipent ', ' fanmeeting ',
327        ' animal crossing ', ' primary school ', 'sexysaturday',
328        ' xiuweet ', ' album ', ' being sad is not low ',
329        'yall tl v yall twitter', 'breadfootage', ' tokyo drift ',
330        ' cis actors ', 'exols', ' beyonce ', 'germanyiniraq',
331        'sportscenter', ' kp0p', 'sungjin', ' woollim', '4teaching',
332        ' innit ', 'perv', 'cybertruck', ' ambisverse ',
333        'mobicommunitycorner', ' lego idea ', 'parkjihoon',
334        'abroadinjapan', 'mysticism', 'talkradio', ' recap',
335        ' space needle ', ' trailer ', ' add me ', ' mlhc ',
336        ' github daily development activity ', ' exo got ',
```

```
337          ' religion ', ' puto ', ' mmj ', 'btsieverse', ' ama ',
338          'ventuskycom', ' dm ', 'toyassgirl', 'minhyu', ' pepe ',
339          'tarotcards', ' tengo ', ' chad ', ' e3 ', 'scamfake',
340          'ratedfgamingllc', 'thatmlpgamer', ' l e n g t h y ',
341          ' clowned ', 'yooyeonseokph', ' gamestop ', ' name for a cat',
342          ' anris ', 'the jews', 'australian open', ' kitten content ',
343          ' garak ', ' george bush ', 'crypto', 'chaesuart', 'chonky',
344          'steven universe', 'shibas', 'spacegabpodcast', 'tranny',
345          ' sailor venus ', ' song ', ' cpn ', 'baehoyuong',
346          ' avrillavigne ', 'meghanandharry', ' astronaut au',
347          ' uwu ', ' daddy ', ' grade student ', ' space au ',
348          ' mr fantastic ', ' exo concert', ' marvel universe ',
349          ' want exo to ', 'furrson', ' a capricorn ', ' yo lo ',
350          'picard', ' antifa ', 'happybdaytome11', 'onlineexhibition',
351          'put into the universe',' variety show ',
352          'ubereats', ' el catpo ', 'attackont', 'chuueoy', ' piss on ',
353          ' yo no ', ' spelling errors ', 'fortnite', 'smallbeerpress',
354          'thetoychronicle', ' pendejo', ' justice league', ' romulus',
355          ' wifey ', 'achrisart', 'wwwteslagiveorg', ' freebie ',
356          'drunkscribe', 'sunwoosunn', ' superintendant ',
357          ' a gang ', 'tgirl', 'ozonies', 'imbofication', ' dnd5e ',
358          'pridepublishing', ' 2017 republicans ', ' sponsor update ',
359          ' murder mystery ', ' yoongi ', 'gyulmeetsworld', ' pm modi ',
360          'tarot', 'british museum', 'miumiuofficial', ' stonk ',
361          'selenagomez', ' feburary is ', ' in my lap ',
362          'buyastrangerabook', 'during lockdown', 'sabesdoesart',
363          ' xbox ', 'musicmonday', ' british tabloid ', ' dios ',
364          'pseudoscience', 'ncontextmukbang', 'ringabelcaaffordable',
365          'outofcontext', ' mandalorian ', ' stonks ', 'super smash',
366          'japancore', 'gimped', ' abuelita ', 'live interview', 'simp',
367          'thankyoucirce', ' spock ', 'voucher code', 'ilysm', ' nigg',
368          'piersmorgan', 'doge', 'ghanasport', 'brunomars', 'spongebob',
369          'souljaboy', 'classicstarwars', 'artisqueer', 'coinmarket',
370          ' dionysus ', ' ziggy ', 'boycottutah', 'nbatopshotthis',
371          ' liminal', ' hogayi ', ' james charles ', 'taekoobots',
372          ' reddit ', ' roleplay ', 'artemisentm', ' junior school ',
373          'followyourbuyer', ' this link ', ' misinformation ',
374          ' altseason ', ' our music ','emgauthor', 'bootahflytattoo',
375          'jobincore', ' green run', ' f*ck', 'poofbuttmutt',
376          'whatartshealth', 'vantetaetaekook', 'indiegame', ' hey guys ',
377          ' aoc', ' sailor mars ', 'superstarwonwoo', 'freeddawg',
378          'like comment', 'killerstockcall', ' therapy dog',' god bless ',
379          'complete our survey', ' gucci ', 'lostartlimited',
380          ' want an easy way to get a', ' road closure ', ' anus ',
381          'nothingbutsex', ' ufc', ' dilf ', ' my esa', ' mk ',
382          ' gamergirl', ' my pup ', ' rona ', ' twizz', ' an exo ',
383          ' spacex drink', ' facebook ', '30 seconds to mars',
384          'army exo', ' por ', 'gravy train', 'low low price',
385          ' erotic ', 'petenajarian', 'bert thompson',
386          ' good space mentaility', 'throw back tuesday',
387          'hotdogstand', 'test math brutal', 'exo cock', 'th house',
388          'st house', 'rd house', 'beautiful without makeup',
```

```
389        'modern day pagan', 'worklifebalance', ' piano melody',
390        ' kin luna ', ' hbd ', 'burfday', 'wellnesswednesday',
391        'artemis troof', ' dominate cricket',
392        'rookie exo', ' en esa ', ' qui ' , 'bookmark link',
393        'taylorswift', 'hit music', 'white supremacist terror',
394        'plastic bunny hold', 'pre order', 'cmsphupdates',
395        ' usagi ', 'artemis thing app', 'luna artemis',
396        'historymonth', 'codepink', 'wonder woman',
397        'fordnation', 'ice cream', 'license photo',
398        'akira', 'bar support', 'torontosun', 'artemis mgmt',
399        'takatifu', 'florida insurance', 'primetime tv',
400        'freesenegal', 'deeplearn', '4th winner', 'art museum',
401        'birth chart', 'junyoung', 'ultimo', 'gizbot',
402        'via youtube', 'hb 2013', 'internationalspacestation passing',
403        'torontoobserver', 'radio edit', ' rompe ',
404        'starship take group', ' adorable video', 'video edits ',
405        'datang', 'izmminone', 'premierleague',
406        'kingkong', 'etsy', 'wontonin', 'annyeongz', 'pleaseretweet',
407        'fast food twitter', 'attack titan', 'wake bake',
408        'web drama', 'ted cruz', 'fuck child', '100daysof', 'banger',
409        'tv special', '5x3', 'skyrim', 'jungkook', 'lengwrites',
410        'ejaculation', 'lyric video', 'cardodging', 'brujeria',
411        ' woth magic ', 'youtube son', 'paste link', 'woosan',
412        'artemis richochet', ' whipped chai ', ' swaddle art',
413        'mar entering gemini', 'javascript', 'reupload',
414        ' acoustic cover ', 'andrew elementary', 'spring grove int',
415        'dunseith public', 'student comment', 'weird pronunciation',
416        'christian school', 'gaither high', 'stranship',
417        'geographyteacher', 'bandtan', 'ssg', 'icky thong',
418        'sex class', 'player coach staff', 'click photo', ' almanac',
419        'new menu', 'face post', 'astrological sign', 'trudeau',
420        'degree 2017', 'new blog', ' kid suspended ', 'marco rubio'
421        ' elmer ', ' fudd ', ' impediment ', ' mawrs ', ' retard ',
422        ' voiceover ', ' lisp ', ' studder ', ' accent ',
423        ' antichrist ' , ' sjw ', ' dwopping ', ' arthursday ',
424        ' junior high ', ' shout out', 'shoutout', 'acadamy',
425        ' middle school ', ' high school ', ' secondary school ',
426        ' academy ', ' ibc ', 'esport', 'examcancelhoga',
427        ' dcpublicschool', ' brewsters ', ' head coach ',
428        'findyourthing', 'gangsters', 'rbandme', 'music',
429        'homework',' ovni ','wandavision','dogecoin','bcu_ry',
430        'termpaper', 'newmusic','nowplaying', 'fridaylivestream',
431        'btspace', 'cravity',  'bitcoin', 'stanworld', 'izone',
432        'colchester', 'nowplaying', 'onlineclass', 'worldbookday',
433        'satyamevajayate', 'snydercut', 'mamamoo',  'eva71',
434        'paralleluniversecontinues', 'dogecoin', 'showtimebet',
435        'didyouknow', 'harrystyles',  'grammy', 'etsy',
436        'myanmarmilitaryterrorists', 'poshmark', 'stpatricksday',
437        'mindbreeze', ' ad ', 'artoftheblue', 'essaydue', 'bigdata',
438        'aylesbury',  'piday', 'harpenden',  'soundcloud',
439        ' doge ', 'zacksnyders', ' nft ',  'ifttt', 'shopee',
```

```
440          bjpmodel , chanyeol , baekhyun , xiuweet ,  xiumin ,
441          'oprahmeghanharry', 'auspixius', ' suho ', 'dollwith',
442          'icanimagine', 'thewildsspace',' dyk ', 'artemis_twt',
443          'esa_celebnews', 'superstraight', 'seekthetruth', 'izone',
444          'tarotbybronx', 'cryptoart', 'meme king', 'minecraft',
445          'artemis and luna', 'brasileiro', 'myanmar coup', 'baekhyun',
446          'doyoung', 'kpop', 'cuddles', 'bruno mars',
447          'space jam', 'bible says', 'templetonrel']
```

```python
1  regex = r"\*s(http[s]?:\S)|(\S.((com)|(org)|(gov)|(co)|(uk))\S*)\*s"
2  #  to update our display to see how long it's taking
3  end point = len(df)
4  reset(df)
5  # go through the dataframe and really clean up the text
6  for i in range(len(df)):
7      num = len(df)
8      to_lem = df.at[i, 'text']
9      # first we remove emojis
10     to_lem = deEmojify(to_lem)
11     # then urls
12     to_lem = re.sub(regex, "", to_lem, flags=re.I)
13     # then non-ascii characters
14     to_lem = to_lem.encode("ascii", "ignore").decode()
15     # turn into a list to replace contractions
16     to_lem = to_lem.split()
17     for j in range(len(to_lem)):
18       if to_lem[j] in replacements:
19         to_lem[j] = replacements[to_lem[j]]
20     lemmed = []
21     for item in to_lem:
22         # remove punctuation
23         item = re.sub(r'[^\w\s]', '', item)
24         # split it up again, since the contractions would be one item,
25         # regardless of length
26         item = item.strip()
27         if item != ' ':
28           lemmed.append(lemmatizer.lemmatize(item))
29     # remove stop words
30     filt_sen = [w for w in lemmed if not w in stop_words and len(w) >=2]
31     # addition I'd like to make for the future
32     # filt_sen = [w for w in filt_sen if w.startswith('@') == False]
33     # see if there is any innapropriate content
34     check =  any(item in filt_sen for item in list_to_drop)
35     # filter to minimum sentence length
36     if len(filt_sen) >= 6 and check == False:
37       lemstr = ' '.join(filt_sen)
38       # check and correct spelling to the best we can
39       textBlb = TextBlob(lemstr)
40       df.at[i, 'text'] =  textBlb.correct()
41     # if it isn't long enough, we can't draw meaningful conclusions
42     else:
```

```python
43        df.drop(index=1, inplace=True)
44    # this takes hours so good to print it so you know it's really
45    # working and not stuck
46    display(f"Checking text { i + 1 } of {end_point}:   ",
47            f"{(i + 1) * 10000 // (len(df)) / 100}% Processed",
48            f"{num)} Remaining ")
```

```python
1 # check that it worked
2 df.head(10)
```

```python
1 df.describe()
```

```python
1 # export the data
2 df = df.to_csv('cleaned.csv')
```