

```
# INSTALL REQUIRED PACKAGES
# pandas is a Python library, to install via the pip package manager, for working with structured data (tables,
CSVs, etc)
# quiet makes the output cleaner

!pip install --quiet pandas
import pandas as pd
from google.colab import files

# SET SOURCE CSV TO USE FOR SEARCHING (UPLOADED TO GOOGLE COLAB)
filename = '/content/all_youtube_comments-5.csv'
df = pd.read_csv(filename)
print(df.head())

# DEFINE CATEGORIES AND KEYWORDS AND PHRASES WITHIN THEM
keyword_categories = {
    'Structural denial': [
        'not a real woman', 'you are not real', 'only two genders exist',
        'your identity is made up', 'trans people don't exist', 'it's just a trend',
        'you're mentally ill', 'you're delusional', 'you mutilated yourself',
        'he's a man', 'still a man in a dress', 'you'll never be one of us',
        'born a man = always a man', 'you're sick', 'men in dresses',
        'man in a dress', 'man in dress', 'men wearing dresses', 'mutilated',
        'mutilation', 'mutil8ed', 'mutillated', 'ull never be a woman',
        'ull never be a man', 'youre not a woman', 'youre not a man',
        'you are not a woman', 'you are not a man', 'only two genders',
        'only2genders', 'biological male', 'biological female', 'bio male',
        'bio female', 'born a man', 'born a woman', 'still a woman',
        "he's a man", 'she's a man', 'hes a man', 'shes a man', 'delusional',
        'dillusional', 'mentally ill', 'mentaly ill', 'mental illness',
        'confused', 'confuzed', 'unnatural', 'abomination', 'abominayshun',
        'ab0mination', 'disgusting', 'disgosting', 'brainwashed', 'brainwash',
        'convert them back', 'revert them', 'reverse the surgery', 'detrans them',
```

```
'destroy their hormones', 'destroy hormones', 'no hormones', 'hormone junkie',
'surgery addict', 'gender delusion', 'genderdelusion', 'just a fetish',
'fetishizing womanhood', 'fetishizes women', 'womanface', 'manface',
'self-mutilation', 'self mutilation', 'butchered body', 'butchered',
'you're sick', 'youre sick', 'ur sick'
],
'Dog whistles': [
'gender ideology', 'woke cult', 'LGBTQ lobby', 'protect the children',
'radical trans agenda', 'biological truth', 'natural order',
'agenda-driven', 'save the kids', 'agendadriven', 'saveourkids',
'LGBTQlobby', 'trans mafia', 'transmafia', 'indoctrinating kids',
'indoctrinatekids', 'child abuser', 'childabuser'
],
'Slurs': [
'trans agenda', 'transagenda', 'men in dresses', 'groomer', 'tranny',
'pedophile', 'pedo', 'shemale', 'freak', 'mutant', 'pervert', 'failed man',
'failed woman', 'degenerate', 'demon', 'satanic', 'trannny', 'Tranni',
'trnny', 'gromer', 'groomr', 'pedofile', 'pedogroomer', 'trans cult',
'transcult', 'crossdresser', 'cross dresser', 'freak', 'fr33k', 'perv',
'degen', 's4tanic', 's8nic', 'trans freak', 'transfreak', 'trans garbage',
'transgarbage'
],
'Neutral': [
'trans woman', 'trans man', 'nonbinary', 'trans rights', 'trans', 'FTM',
'MTF', 'male to female', 'female to male', 'trans joy', 'trans pride',
'trans community', 'enby', 'gender diverse', 'genderqueer', 'agender',
'genderfluid', 'transmasc', 'transfem', 't4t', 'they/them', 'she/they',
'he/they', 'xe/xem', 'ze/hir', 'trans visibility', 'trans awareness',
'transition journey', 'top surgery', 'bottom surgery', 'gender euphoria',
'queer'
],
'Threats': [
'burn in hell', 'burninhell', 'burn u in hell', 'kill yourself',
```

```

        'killyourself', 'kys', 'kys pls', 'shoot them', 'hang them', 'execute them',
        'corrective rape', 'correctiverape', 'rape them straight'
    ]
}

# BUILD A DICTIONARY CONTAINING THESE KEY WORDS AND PHRASES MAPPED TO CATEGORIES
keyword_to_category = {}
for category, words in keyword_categories.items():
    for kw in words:
        keyword_to_category[kw.lower()] = category

# SELECT COLUMNS TO SEARCH (CHANGE IF NEEDED)
columns_to_search = df.columns.tolist() # or ['comment', 'description']

# DEFINE FUNCTION TO FIND KEYWORDS/PHRASES AND CATEGORIES IN EACH ROW
def find_keywords_and_categories(row):
    found_keywords = set()
    found_categories = set()

    for col in columns_to_search:
        text = str(row[col]).lower()
        for kw in keyword_to_category:
            if kw in text:
                found_keywords.add(kw)
                found_categories.add(keyword_to_category[kw])

    return pd.Series({
        'keyword': ';'.join(sorted(found_keywords)),
        'category': ';'.join(sorted(found_categories))
    })

# APPLY THIS FUNCTION TO THE DATA, ROW BY ROW
# Identified keywords/phrases and their categories are outputted in new columns

```

```
df[['keyword', 'category']] = df.apply(find_keywords_and_categories, axis=1)

# SAVE TO CSV
output_filename = "csv_with_keywords_and_categories.csv"
df.to_csv(output_filename, index=False)
files.download(output_filename)
```