Utilizing Reddit API!

```
!pip install praw
```

```
Requirement already satisfied: praw in ./anaconda3/lib/python3.10/site-packages (7.7.1)
Requirement already satisfied: prawcore<3,>=2.1 in ./anaconda3/lib/python3.10/site-packag
Requirement already satisfied: update-checker>=0.18 in ./anaconda3/lib/python3.10/site-pa
Requirement already satisfied: websocket-client>=0.54.0 in ./anaconda3/lib/python3.10/sit
Requirement already satisfied: requests<3.0,>=2.6.0 in ./anaconda3/lib/python3.10/site-pa
Requirement already satisfied: six in ./anaconda3/lib/python3.10/site-packages (from webs
Requirement already satisfied: charset-normalizer<4,>=2 in ./anaconda3/lib/python3.10/sit
Requirement already satisfied: idna<4,>=2.5 in ./anaconda3/lib/python3.10/site-packages (
Requirement already satisfied: urllib3<3,>=1.21.1 in ./anaconda3/lib/python3.10/site-pack
Requirement already satisfied: certifi>=2017.4.17 in ./anaconda3/lib/python3.10/site-pack
```

First collecting Schizophrenic patient data

```
import praw
import pandas as pd

reddit = praw.Reddit(
    [redacted]
    user_agent="Schiz/suicide correlation by /u/NecessaryClassic4135"
)

usernames = ["sunfloras", "No-Molasses-2247", 'Repulsive_Ring_2309', 'helsdog', 'Mental_

data = []
for username in usernames:
    # Get reddit user as object
    user = reddit.redditor(username)

    # get their top 20 posts (most recent)
    submissions = user.submissions.new(limit=20)

    # loop through and append the data
    for submission in submissions:
        data.append({
            "Username": username,
            "Post Text": submission.selftext
        })
```

```
# convert dictionaries into a dataframe
study = pd.DataFrame(data)
study['Diagnosis'] = "SCH"
study
```

	Username	Post Text	Diagnosis
0	sunfloras	when i meet new people i'm really scared that ...	SCH
1	sunfloras	my anxiety and sometimes psychosis acts up rea...	SCH
2	sunfloras	i was having an episode where i thought no one...	SCH
3	sunfloras	i've tried 6mg risperidone, 10mg abilify, and ...	SCH
4	sunfloras	are there any meds stronger than abilify, risp...	SCH
...	...	...	...
780	trashaccountturd	Ok, so I've been doing research on which flywh...	SCH
781	trashaccountturd	So, the best working theory I have is that an ...	SCH
782	trashaccountturd	A credit card was stuck, causing the cup holde...	SCH
783	trashaccountturd	This is the only spots like this on the tire, ...	SCH
784	trashaccountturd	Anyone else poppin' ollies and doing 3 flips?\...	SCH

785 rows × 3 columns

Second, collecting mental health users

```
from prawcore.exceptions import NotFound

reddit = praw.Reddit(
    [redacted]
    user_agent="Schiz/suicide correlation by /u/NecessaryClassic4135"
)

usernames = ["UtopianPariah", "ilovereditrabbit", "LisKoz1989", "joeym412", "SillyDot33

mhdata = []

for username in usernames:
    try:
        user = reddit.redditor(username)

        _ = user.name

        submissions = user.submissions.new(limit=20)

        for submission in submissions:
            mhdata.append({
                "Username": username,
                "Post Text": submission.selftext
            })

    except NotFound:
        print(f"User {username} not found, skipping.")
    except Exception as e:
        print(f"An error occurred for user {username}: {e}")
```

```
User DryInvestment1906 not found, skipping.
```

```
mentalhealth = pd.DataFrame(mhdata)
mentalhealth['Diagnosis'] = "MH"
mentalhealth
```

	Username	Post Text	Diagnosis
0	UtopianPariah	I'm 23, just came to New Jersey 3 months ago, ...	MH
1	UtopianPariah	Anyone willing to be friends can text me, I'm ...	MH
2	UtopianPariah	after finishing elden ring, I bought sekiro in...	MH
3	UtopianPariah	so, all my friends have either drifted apart o...	MH
4	UtopianPariah	Please DM if interested to have a conversation.	MH
...	...	...	...
703	Guilty_Critic	Hello, I have a true phobia of fish. I believe...	MH
704	Guilty_Critic	So I have a new neighbor and their dog barks a...	MH
705	Guilty_Critic	My apartment has a door that cannot be opened ...	MH
706	Guilty_Critic	I'm not sure if it's something i need to get l...	MH
707	Guilty_Critic	Is it possible to add remote start to a 2015 m...	MH

708 rows × 3 columns

Lastly, with neither diagnosis

```
from prawcore.exceptions import NotFound

reddit = praw.Reddit(
    [redacted]
    user_agent="Schiz/suicide correlation by /u/NecessaryClassic4135"
)

usernames = ["youre-welcome5557777", "maybesaydie", "Current_Variety_9577", "MileHighMon

nonedata = []

for username in usernames:
    try:
        user = reddit.redditor(username)

        _ = user.name

        submissions = user.submissions.new(limit=20)

        for submission in submissions:
            nonedata.append({
                "Username": username,
                "Post Text": submission.selftext
            })

    except NotFound:
        print(f"User {username} not found, skipping.")
    except Exception as e:
        print(f"An error occurred for user {username}: {e}")
```

```
User Wide-Refuse-7815 not found, skipping.
```

```
none = pd.DataFrame(nonedata)
none['Diagnosis'] = "Neither"
none
```

	Username	Post Text	Diagnosis
0	youre-welcome5557777	I've heard that the O's fanbase used to stretc...	Neither
1	youre-welcome5557777		Neither
2	youre-welcome5557777		Neither
3	youre-welcome5557777		Neither
4	youre-welcome5557777		Neither
...	...	...	...
817	Natural-Brilliant-95	Hey everyone just a couple of questions 1st if...	Neither
818	Natural-Brilliant-95	Hi everyone I have a player who wants to use m...	Neither
819	Natural-Brilliant-95	Hi I am wondering if its possible to randomize...	Neither
820	Natural-Brilliant-95	Hi all I am making my second ever cleric and I...	Neither
821	Natural-Brilliant-95	So I just lost the randomized run i was doing ...	Neither

822 rows × 3 columns

Combining all of my data

```
import pandas as pd
all_posts = pd.concat([study, mentalhealth, none])
all_posts
```

	Username	Post Text	Diagnosis
0	sunfloras	when i meet new people i'm really scared that ...	SCH
1	sunfloras	my anxiety and sometimes psychosis acts up rea...	SCH
2	sunfloras	i was having an episode where i thought no one...	SCH
3	sunfloras	i've tried 6mg risperidone, 10mg abilify, and ...	SCH
4	sunfloras	are there any meds stronger than abilify, risp...	SCH
...	...	...	...
817	Natural-Brilliant-95	Hey everyone just a couple of questions 1st if...	Neither
818	Natural-Brilliant-95	Hi everyone I have a player who wants to use m...	Neither
819	Natural-Brilliant-95	Hi I am wondering if its possible to randomize...	Neither
820	Natural-Brilliant-95	Hi all I am making my second ever cleric and I...	Neither
821	Natural-Brilliant-95	So I just lost the randomized run i was doing ...	Neither

2315 rows × 3 columns

```
all_posts.to_csv("all_posts.csv", index=False)
```

Creating the training data set

```
reddit = praw.Reddit(
    [redacted]
    user_agent="Schiz/suicide correlation by /u/NecessaryClassic4135"
)

subreddit = reddit.subreddit("SuicideWatch")

posts = []
for submission in subreddit.new(limit=None):
    title = submission.title
    text = submission.selftext
    combined_text = f"{title}. {text}"
    posts.append({
        "combined_text": combined_text,
        "is_suicidal": True  # Manually label as suicidal
    })
```

```
suicide = pd.DataFrame(posts)
suicide.to_csv("suicide_data.csv", index=False)
```

```
import random
reddit = praw.Reddit(
    [redacted]
    user_agent="Schiz/suicide correlation by /u/NecessaryClassic4135"
)

subreddit_names = ["Coachella", "assassinscreed", "audiobooks", "aviation", "budgetfood"

post_texts = []
```

```
for subreddit_name in subreddit_names:
    # get 4 random posts
    subreddit = reddit.subreddit(subreddit_name)
    random_submissions = random.sample(list(subreddit.new(limit=5)), 5)

    for submission in random_submissions:
        post_texts.append(submission.selftext)
```

```
nonsuicide = pd.DataFrame(post_texts)
nonsuicide.to_csv("nonsuicide_data.csv", index=False)
```

I went through and took the top responses. I then self-identified whether they were suicidal text or not.

```
training_data = pd.read_csv('data_test.csv')
training_data
```

	compiled_text	is_suicidal
0	I'll be there Friday, leave Tuesday and I will...	False
1	Wishlist item. Again.	False
2	I know most people would rather see them make ...	False
3	Just got the game and put in about four hours ...	False
4	Honestly I'm hoping for a Japanese dub. Some g...	False
5	I completed every AC games and stopped after U...	False
6	Been playing ac 4 again recently and am really...	False
7	Why are American based books read by people wi...	False
8	So it's the Arabian Nights based on an 1811 tr...	False
9	Hello! Thank you for clicking in!\n\nI am a wr...	False
10	I need an audio book for the book players by J...	False
11	People flying thru Orlando are not well season...	False
12	Which is the minimum sink speed for the Cessna...	False
13	A long time ago I hard the term "Red Flag ops,...	False
14	14.5 oz can of diced tomatoes \n\n4 oz can of ...	False
15	For the next 2 weeks I have very little to spe...	False
16	Due to kidney stones I'm limited in a lot of c...	False
17	unsure of how much each sandwich costs exactly...	False
18	I dont want to make it to my birthday. Its in ...	True
19	I think I'm ready. I don't know where to write...	True
20	This will never end and I can't make it end ei...	True
21	Life is annoying. I'm 18(M) \n\nDon't you hate...	True
22	I plan on killing myself in 35 days and all I ...	True
23	I am suicidal but scared . I think the only th...	True
24	Already took my pills. Throwaway causel don't ...	True
25	Want to end it, but want to stay.. Look... I'm...	True
26	I'm out of fuel . Just dragging myself through...	True
27	nobody loves me. im not going to make it to 20...	True
28	I realized I'll never really be happy so why b...	True
29	There's really no point to life. There is no p...	True

34	Planned. I have it planned for Wednesday yet I...	True
35	I can do it.. 15/M\n\nI can't believe it. Thes...	True
36	"But I don't wanna die". Ugh. Ykwim? Just ugh....	True
37	Probably going to kill myself over AI soon. I'...	True

Training + Testing Model!

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report

X = training_data['compiled_text']
y = training_data['is_suicidal']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42

vectorizer = TfidfVectorizer(max_features=1000)
X_train_vec = vectorizer.fit_transform(X_train)
X_test_vec = vectorizer.transform(X_test)

model = LogisticRegression()
model.fit(X_train_vec, y_train)

y_pred = model.predict(X_test_vec)
print(classification_report(y_test, y_pred))

all_posts_vec = vectorizer.transform(all_posts['Post Text'])
all_posts['predicted_is_suicidal'] = model.predict(all_posts_vec)

print(all_posts)
```

```
              precision    recall  f1-score   support

       False       0.75      1.00      0.86         3
        True       1.00      0.80      0.89         5

    accuracy                           0.88         8
   macro avg       0.88      0.90      0.87         8
weighted avg       0.91      0.88      0.88         8

                Username                                          Post Text  \
0              sunfloras  when i meet new people i'm really scared that ...
1              sunfloras  my anxiety and sometimes psychosis acts up rea...
2              sunfloras  i was having an episode where i thought no one...
3              sunfloras  i've tried 6mg risperidone, 10mg abilify, and ...
4              sunfloras  are there any meds stronger than abilify, risp...
..                   ...                                                ...
817  Natural-Brilliant-95  Hey everyone just a couple of questions 1st if...
818  Natural-Brilliant-95  Hi everyone I have a player who wants to use m...
819  Natural-Brilliant-95  Hi I am wondering if its possible to randomize...
820  Natural-Brilliant-95  Hi all I am making my second ever cleric and I...
821  Natural-Brilliant-95  So I just lost the randomized run i was doing ...

    Diagnosis  predicted_is_suicidal
0         SCH                 False
1         SCH                  True
2         SCH                  True
3         SCH                 False
4         SCH                  True
..        ...                   ...
817   Neither                 False
818   Neither                 False
819   Neither                  True
820   Neither                  True
821   Neither                 False

[2315 rows x 4 columns]
```

```
all_posts.head(2)
```

	Username	Post Text	Diagnosis	predicted_is_suicidal
0	sunfloras	when i meet new people i'm really scared that they can tell there's something wrong with me. i mask really well so it appears nothing is wrong but masking all the time can be so tiring.	SCH	False
1	sunfloras	my anxiety and sometimes psychosis acts up really bad during life changes. i'm currently going through a life change- moving back to my home state but a bit away from my family. i'll have a room with roommates but it's my first time paying rent somewhere. i'm really scared something could go wrong and i'll end up homeless again. i can barely sleep or eat. how do you guys deal with life changes??	SCH	True

Analysis!

```
pd.set_option('display.max_colwidth', None)
all_posts.iloc[708]
```

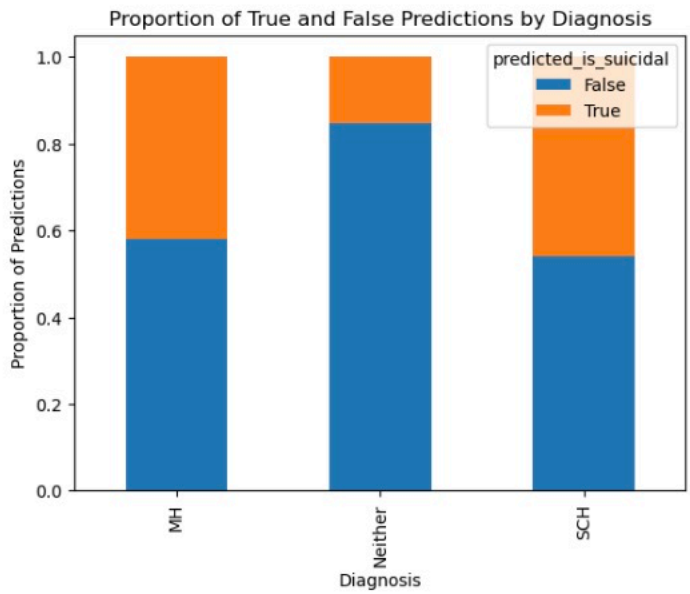
```
Username
Post Text                I don't really get what the average person seems to get from alc
Diagnosis
predicted_is_suicidal
Name: 708, dtype: object
```

```
import matplotlib.pyplot as plt

counts_df = all_posts.groupby(['Diagnosis', 'predicted_is_suicidal']).size().unstack(fil
proportions_df = counts_df.div(counts_df.sum(axis=1), axis=0)
proportions_df.plot(kind='bar', stacked=True)

plt.xlabel('Diagnosis')
plt.ylabel('Proportion of Predictions')
plt.title('Proportion of True and False Predictions by Diagnosis')

plt.show()
```



```
proportions_df
```

predicted_is_suicidal	False	True
Diagnosis		
MH	0.581921	0.418079
Neither	0.846715	0.153285
SCH	0.541401	0.458599

```
grouped_df = all_posts.groupby(['Username', 'Diagnosis'])
at_least_one_suicidal = grouped_df['predicted_is_suicidal'].any().groupby('Diagnosis').s

total_users = grouped_df['Username'].nunique().groupby('Diagnosis').sum()

proportion_suicidal_users = at_least_one_suicidal / total_users

print(proportion_suicidal_users)
```

```
Diagnosis
MH         0.957447
Neither    0.530612
SCH        0.895833
dtype: float64
```

```
def count_suicide_mentions(posts):
    return sum(('suicide' in post) or ('suicidal' in post) for post in posts)

grouped = all_posts.groupby(['Username', 'Diagnosis'])['Post Text'].apply(count_suicide_

users_with_suicide_mentions = grouped[grouped['suicide_mentions'] > 0]

users_with_suicide_mentions
```

	Username	Diagnosis	suicide_mentions
3	ASleepyB0i	MH	1
4	AcceptableSmoke9129	MH	1
10	Basil-the-bagel	SCH	1
12	BitersAnon	SCH	1
16	Careless-Scratch-658	MH	1
26	Cute-Avali	SCH	1
37	Guilty_Critic	MH	1
43	InterestingKiwi5004	SCH	1
59	MistWeaver80	Neither	1
64	No-Molasses-2247	SCH	1
109	drowsyneon	SCH	1
120	justarandomer_	MH	1
142	x37h4n	SCH	1

```
# Did my model think this text from MistWeaver80 was suicidal?

try:
    all_posts[(all_posts.predicted_is_suicidal == True) and (all_posts.Username == "Mist
except Exception as e:
    print("No predictions of suicidal text from this user")
```

```
No predictions of suicidal text from this user
```

```
# It seems like this was not necessarily suicidal text...

all_posts[all_posts.Username == "MistWeaver80"].head(1)
```

	Username	Post Text	Diagnosis	predicted_is_suicidal
183	MistWeaver80	Link to the article: https://www.telegraph.co.uk/global-health/women-and-girls/kurdistan-iraq-suicide-self-immolation-domestic-violence/	Neither	False

```
from scipy.stats import chi2_contingency

contingency_table = pd.crosstab(all_posts['Diagnosis'] == 'SCH', all_posts['predicted_is
print("Contingency Table:")
print(contingency_table)

chi2, p, dof, expected = chi2_contingency(contingency_table)

print(f"\nChi2 Statistic: {chi2}")
print(f"P-value: {p}")

if p < 0.05:
    print("There is a significant association between schizophrenia diagnosis and suicid
else:
    print("There is no significant association between schizophrenia diagnosis and suici
```

```
Contingency Table:
predicted_is_suicidal  False  True
Diagnosis
False                   1108    422
True                     425    360

Chi2 Statistic: 76.67201765213719
P-value: 2.018431476549952e-18
There is a significant association between schizophrenia diagnosis and suicidal text (p <
```

```
import numpy as np
from scipy.stats import chi2_contingency
from sklearn.linear_model import LinearRegression

new_posts = all_posts.copy()

new_posts['predicted_is_suicidal'] = new_posts['predicted_is_suicidal'].astype(int)

expected_categories = ['MH', 'SCH', 'Neither']

new_posts = pd.get_dummies(new_posts, columns=['Diagnosis'], prefix='Diagnosis')
for category in expected_categories:
    column_name = f'Diagnosis_{category}'
    if column_name not in new_posts.columns:
        new_posts[column_name] = 0


contingency_table = pd.crosstab(new_posts['Diagnosis_SCH'], new_posts['predicted_is_suic

chi2, p, dof, expected = chi2_contingency(contingency_table)

n = contingency_table.sum().sum()
cramers_v = np.sqrt(chi2 / (n * (min(contingency_table.shape) - 1)))
print(f"Cramér's V: {cramers_v}")

X = new_posts[['Diagnosis_MH', 'Diagnosis_Neither', 'Diagnosis_SCH']]
y = new_posts['predicted_is_suicidal']

model = LinearRegression()
model.fit(X, y)
r_squared = model.score(X, y)
print(f'R-squared: {r_squared}')
```

```
Cramér's V: 0.18198808217231985
R-squared: 0.08461516541704761
```