

```
In [1]: #IMPORT REQUIRED LIBRARIES
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb
```

```
In [2]: #IMPORT DATASET
df=pd.read_csv(r"C:\Users\mo hd arhab ahmad\OneDrive\Documents\Stress.csv")
```

```
In [3]: df
```

```
Out[3]:
```

	subreddit	post_id	sentence_range	text	label	confidence	social_timestamp
0	ptsd	8601tu	(15, 20)	He said he had not felt that way before, sugge...	1	0.800000	1521614353
1	assistance	8lbrx9	(0, 5)	Hey there r/assistance, Not sure if this is th...	0	1.000000	1527009817
2	ptsd	9ch1zh	(15, 20)	My mom then hit me with the newspaper and it s...	1	0.800000	1535935605
3	relationships	7rorpp	[5, 10]	until i met my new boyfriend, he is amazing, h...	1	0.600000	1516429555
4	survivorsofabuse	9p2gbc	[0, 5]	October is Domestic Violence Awareness Month a...	1	0.800000	1539809005
...	...	...	...	...	...	...	...
2833	relationships	7oe1t	[35, 40]	* Her, a week ago: Precious, how are you? (I i...	0	1.000000	1515187044
2834	ptsd	9p4ung	[20, 25]	I don't have the ability to cope with it anymo...	1	1.000000	1539827412
2835	anxiety	9nam6l	(5, 10)	In case this is the first time you're reading ...	0	1.000000	1539269312
2836	almosthomeless	5y53ya	[5, 10]	Do you find this normal? They have a good rela...	0	0.571429	1488938143
2837	ptsd	5y25cl	[0, 5]	I was talking to my mom this morning and she s...	1	0.571429	1488909516

2838 rows × 7 columns

```
In [4]: df.dtypes
```

```
Out[4]:
```

subreddit	object
post_id	object
sentence_range	object
text	object
label	int64
confidence	float64
social_timestamp	int64
dtype:	object

```
In [5]: print(df['text'].loc[df.index[0]])
```

He said he had not felt that way before, suggested I go rest and so ..TRIGGER AHEAD IF YOU I'RE A HYPOCONDRIAC LIKE ME: i decide to look up "feelings of doom" in hopes of maybe getting sucked into some rabbit hole of ludicrous conspiracy, a stupid "are you psychic" test or new age b.s., something I could even laugh at down the road. No, I ended up reading that this sense of doom can be indicative of various health ailments; one of which I am prone to.. So on top of my "doom" to my gloom..I am now f'n worried about my heart. I do happen to have a physical in 48 hours.

```
In [6]: df.describe()
```

	label	confidence	social_timestamp
count	2838.000000	2838.000000	2.838000e+03
mean	0.524313	0.808972	1.518107e+09
std	0.499497	0.177038	1.552209e+07
min	0.000000	0.428571	1.483274e+09
25%	0.000000	0.600000	1.509698e+09
50%	1.000000	0.800000	1.517066e+09
75%	1.000000	1.000000	1.530898e+09
max	1.000000	1.000000	1.542592e+09

```
In [7]: df.isna().any()
```

subreddit	False
post_id	False
sentence_range	False
text	False
label	False
confidence	False
social_timestamp	False
dtype: bool	

```
In [8]: df.isnull().any()
```

subreddit	False
post_id	False
sentence_range	False
text	False
label	False
confidence	False
social_timestamp	False
dtype: bool	

```
In [9]: df = df.drop_duplicates()
df
```

	subreddit	post_id	sentence_range	text	label	confidence	social_timestamp
0	ptsd	8601tu	(15, 20)	He said he had not felt that way before, sugge...	1	0.800000	1521614353
1	assistance	8lbrx9	(0, 5)	Hey there r/assistance, Not sure if this is th...	0	1.000000	1527009817
2	ptsd	9ch1zh	(15, 20)	My mom then hit me with the newspaper and it s...	1	0.800000	1535935605
3	relationships	7rorpp	[5, 10]	until i met my new boyfriend, he is amazing,	1	0.600000	1516429555

	subreddit	post_id	sentence_range	text	label	confidence	social_timestamp
				h...			
4	survivorsofabuse	9p2gbc	[0, 5]	October is Domestic Violence Awareness Month a...	1	0.800000	1539809005
...	...	...	...	...	...	...	...
2833	relationships	7oeel1t	[35, 40]	* Her, a week ago: Precious, how are you? (I i...	0	1.000000	1515187044
2834	ptsd	9p4ung	[20, 25]	I don't have the ability to cope with it anymo...	1	1.000000	1539827412
2835	anxiety	9nam6l	(5, 10)	In case this is the first time you're reading ...	0	1.000000	1539269312
2836	almosthomeless	5y53ya	[5, 10]	Do you find this normal? They have a good rela...	0	0.571429	1488938143
2837	ptsd	5y25cl	[0, 5]	I was talking to my mom this morning and she s...	1	0.571429	1488909516

2838 rows × 7 columns

```
In [10]: df.groupby("subreddit").count().sort_values(by=["confidence"])
```

```
Out[10]:
```

	subreddit	post_id	sentence_range	text	label	confidence	social_timestamp
	food_pantry	37	37	37	37	37	37
	stress	64	64	64	64	64	64
	almosthomeless	80	80	80	80	80	80
	homeless	168	168	168	168	168	168
	survivorsofabuse	245	245	245	245	245	245
	assistance	289	289	289	289	289	289
	domesticviolence	316	316	316	316	316	316
	anxiety	503	503	503	503	503	503
	relationships	552	552	552	552	552	552
	ptsd	584	584	584	584	584	584

```
In [11]: df.confidence.max()
```

```
Out[11]: 1.0
```

```
In [12]: df.confidence.min()
```

```
Out[12]: 0.428571429
```

This dataset is labelled as 0 and 1, where 0 indicates no stress and 1 indicates stress.

Subreddit indicates that what are the primary reasons for stress.

```
In [13]: df[(df.subreddit=="anxiety")&(df.label==1)].shape
```

```
Out[13]: (331, 7)
```

```
In [14]: df[(df.subreddit=="anxiety")&(df.label==0)].shape
```

```
Out[14]: (172, 7)
```

```
In [15]: 331/503
```

```
Out[15]: 0.658051689860835
```

```
In [16]: df[(df.subreddit=="relationships")&(df.label==1)].count()
```

```
Out[16]: subreddit      247
post_id      247
sentence_range  247
text      247
label      247
confidence  247
social_timestamp  247
dtype: int64
```

```
In [17]: df[(df.subreddit=="relationships")&(df.label==0)].count()
```

```
Out[17]: subreddit      305
post_id      305
sentence_range  305
text      305
label      305
confidence  305
social_timestamp  305
dtype: int64
```

```
In [18]: 247/552
```

```
Out[18]: 0.447463768115942
```

```
In [19]: df[(df.subreddit=="ptsd")&(df.label==1)].count()
```

```
Out[19]: subreddit      332
post_id      332
sentence_range  332
text      332
label      332
confidence  332
social_timestamp  332
dtype: int64
```

```
In [20]: df[(df.subreddit=="ptsd")&(df.label==0)].count()
```

```
Out[20]: subreddit      252
post_id      252
sentence_range  252
text      252
```

```
label          252
confidence      252
social_timestamp 252
dtype: int64
```

```
In [21]: 332/584
```

```
Out[21]: 0.5684931506849316
```

```
In [22]: df[(df.subreddit=="ptsd")&(df.confidence>0.8)].count()
```

```
Out[22]: subreddit          243
post_id                    243
sentence_range             243
text                       243
label                      243
confidence                  243
social_timestamp           243
dtype: int64
```

```
In [23]: df[(df.subreddit=="anxiety")&(df.confidence>0.8)].count()
```

```
Out[23]: subreddit          257
post_id                    257
sentence_range             257
text                       257
label                      257
confidence                  257
social_timestamp           257
dtype: int64
```

```
In [24]: df[(df.subreddit=="relationships")&(df.confidence>0.8)].count()
```

```
Out[24]: subreddit          211
post_id                    211
sentence_range             211
text                       211
label                      211
confidence                  211
social_timestamp           211
dtype: int64
```

```
In [25]: #sum of null values for each title of the given dataframe.
df.isnull().sum()
```

```
Out[25]: subreddit          0
post_id                    0
sentence_range             0
text                       0
label                      0
confidence                  0
social_timestamp           0
dtype: int64
```

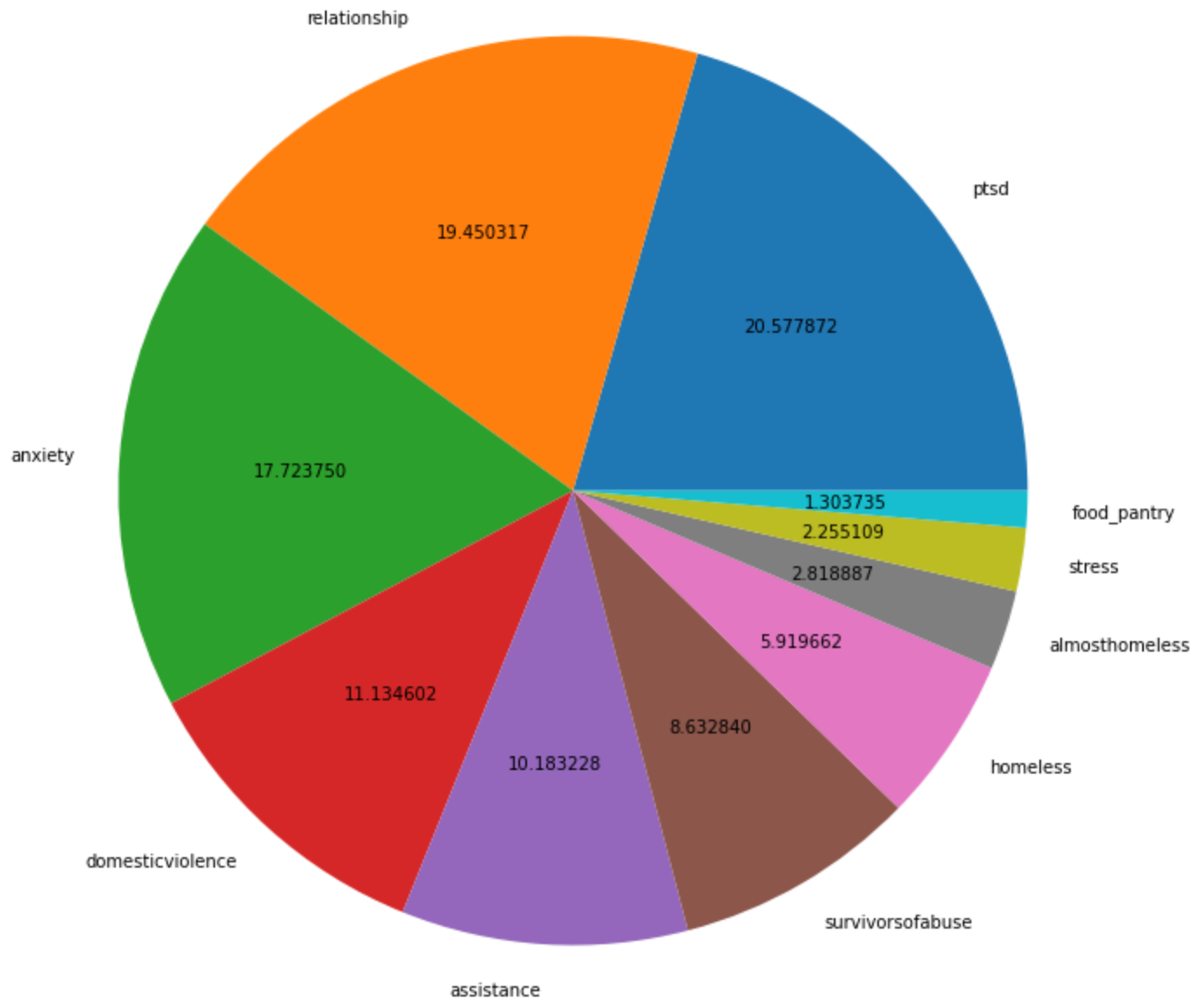
```
In [26]: df.subreddit.value_counts()
```

```
Out[26]: ptsd                584
relationships                552
anxiety                      503
```

```
assistance      289
survivorsofabuse 245
homeless        168
almosthomeless  80
stress          64
food_pantry     37
Name: subreddit, dtype: int64
```

In [27]:

```
#This piechart illustrates the proportion of a particular subreddit(reason of stress).
plt.figure(figsize=(14,12))
plt.pie(x=[584,552,503,316,289,245,168,80,64,37],labels=["ptsd","relationship","anxiety","domesticviolence","assistance","survivorsofabuse","homeless","almosthomeless","stress","food_pantry"],
plt.show()
```



In [28]:

```
df.label.value_counts()
```

Out[28]:

```
1    1488
0    1350
Name: label, dtype: int64
```

In [29]:

```
df.head(8)
```

Out [29]:

	subreddit	post_id	sentence_range	text	label	confidence	social_timestamp
0	ptsd	8601tu	(15, 20)	He said he had not felt that way before, sugge...	1	0.8	1521614353
1	assistance	8lbrx9	(0, 5)	Hey there r/assistance, Not sure if this is th...	0	1.0	1527009817
2	ptsd	9ch1zh	(15, 20)	My mom then hit me with the newspaper and it s...	1	0.8	1535935605
3	relationships	7rorpp	[5, 10]	until i met my new boyfriend, he is amazing, h...	1	0.6	1516429555
4	survivorsofabuse	9p2gbc	[0, 5]	October is Domestic Violence Awareness Month a...	1	0.8	1539809005
5	relationships	7tx7et	(30, 35)	I think he doesn't want to put in the effort f...	1	1.0	1517274027
6	domesticviolence	7iphly	[25, 30]	It was a big company so luckily I didn't have ...	0	0.8	1512854409
7	anxiety	5m3k80	(5, 10)	It cleared up and I was okay but. On Monday ...	1	0.8	1483582174

In [30]:

```
#This chart shows the label of stress and no stress for a particular subreddit.
list=["label"]
plt.figure(figsize=(28,12))
sb.countplot(data=df,x='subreddit',hue='label')
plt.yticks(range(0,400,25))
plt.grid()
plt.savefig(r'C:\Users\mohd arhab ahmad\OneDrive\Pictures.png')
plt.show()
```

