

The dataset

The following dataset was used for the analysis:

Stress Analysis in social media.

https://www.kaggle.com/ruchi798/stress-analysis-in-social-media?select=dreaddit-test.csv

The reddit community dataset found several stress-related categories, where contains expressions related to the stress perceived by users of this platform, where it is found in different topics such as Interpersonal conflict that is related to abuse and social environment. Also, mental illnesses such as anxiety and post-traumatic stress disorder (PTSD) and financial problems.

There are three domains principal which are abuse, anxiety and financial. These categories have a diversity of subreddits such as domestic violence, stress, almost homeless, among others.

Research Question

Are there similar causes to the stress of people in these communities and what are those causes?



As mentioned before, this dataset includes people's experiences with stress. In each register in the dataset people are expressing how and why they feel one way.

Everyone experiences stress in one way or another, and it's caused by our ways to live. Comprehending some of the main reasons people have stress may helps us understand them a little bit better.

By understanding some of the causes of stress, we may be able to not only comprehend these communities, it also helps us to see what are some major causes of stress in our own lifes.

Understanding the data

Assistance	Anxiety	Domestic Anxiety Violence		Stress	Food Pantry
Act of helping or assisting someone or the help supplied.	Feeling of unease, such as worry or fear, that can be mild or severe.	Violence committed by someone in the victim's domestic circle.		To feel worried and nervous	Individual site that d distributes bags or boxes of food directly to those in need who reside in a specified area
Relationships	Survivor of abuse			eless/ homeless	PTSD
The way in which two or more people or things are connected, or the state of being connected.	Individual who has been a victim of financial, emotional, psychological, sexual or physical abuse, among others.		have a usually k	vho do not a home, pecause of Il reasons.	Psychiatric disorder that may occur on people who have experienced or witnessed a traumatic event.

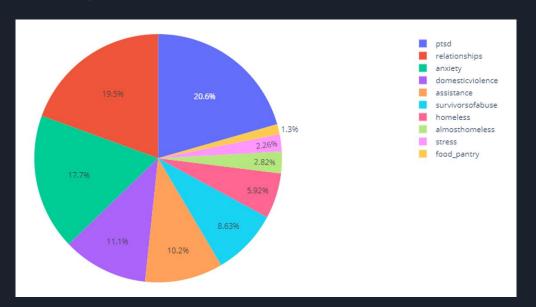
Distribution of Subreddits

We know that in the dataset we have 10 different communities or subreddits.

PTSD, Assistance, Relationships, Survivors of abuse, Domestic violence, Anxiety, Homeless, Stress, Almost homeless, and Food pantry

The first thing to do is to separate the dataset in these different groups before starting to detect similarities. This is very simple, because the dataset already contains a column with the name of the subreddit each post is part of which subreddit.

Distribution of Subreddits in the dataset



The graph shows the categories that have the highest incidence in the users, which are PTSD, relationships and anxiety. This suggests that these are recurring problems in the users of this platform who seek to communicate their realities and find possible answers to such problems.

Word Clouds

Word clouds, also known as tag clouds, are visual representations of text, which is usually used to identify keywords from large portions of text. It allows us to visualize the keywords and their usage according to the category.

Thus, it is helpful to visualize the most common words in a category in a visual, simple, and comprehensible way. Allowing us to see some similarities between the subreddits

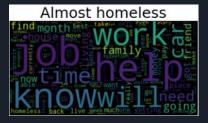
Word Clouds by subreddit

As seen in these images, we are able to understand the existence of some possible correlation between categories using the problems that users have in each subreddit, and the similarity in the way they communicate their problem.

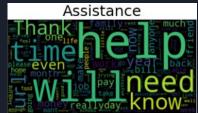


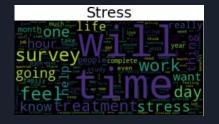




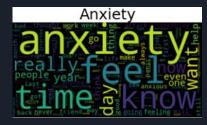














Conclusion

According to the results obtained, it is possible to observe a concordance between categories according to the keywords that are used in each one of them, giving importance to keywords such as time, know, feel, among others. Giving to understand that there are certain similarities in the problems of the users of this platform, since they present common phenomena among them.

References

https://www.kaggle.com/ruchi798/stress-analysis-in-social-media?select=dreaddi t-test.csv

https://www.datacamp.com/courses/sentiment-analysis-in-python