Subreddit
Anxiety vs Diabetes
REPORT

NPL Analysis

SUBMITTED REDDIT TEXTS AND COMMENTS

THE NEED

This path to analyzing Anxiety and Diabetes subreddit text came initially looking at my own family, and seeing how these auto-immune diseases are connected to our genetics and developed over time.

We often think of human interactions and after Covid-19 it's become evident of how Reddit can become community support group, specially for these 2 specific subreddits. Where you can get support, help, advices, helpful tips and tools, sharing and venting your feelings, feeling connected to others in the same spectrum, it became I think (in my opinion) after taking a closer look at each posts, a very relevant support system.

According to the Journal of Psychosomatic Research, a review of the association of diabetes with anxiety, a systematic review and meta-analysis in which supports that diabetes is associated with an increased likelihood of having anxiety disorders and elevated anxiety symptoms.

MY PROBLEM STATEMENT ANSWERS A QUESTION OF:
IS THIS CORRELATION ALSO REFLECTED ON SUBREDDITS COMMENTS
AND SUBMISSIONS?

THE EVIDENCE

Comparing word Count distributions of anxiety and diabetes throughout subreddits since November 2020.

Evidence shows little significance in word count and characters. However anxiety subreddit has a higher average of 101 words and diabetes with an average of 76 words.

Character & Word Count Distributions
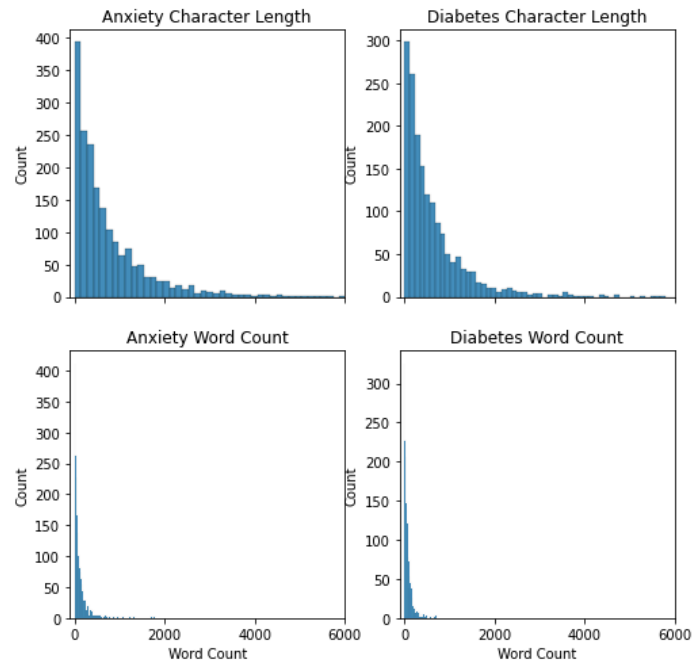
LOOKING AT TOP WORDS USED IN EACH SUBREDDIT

Top 30 Words
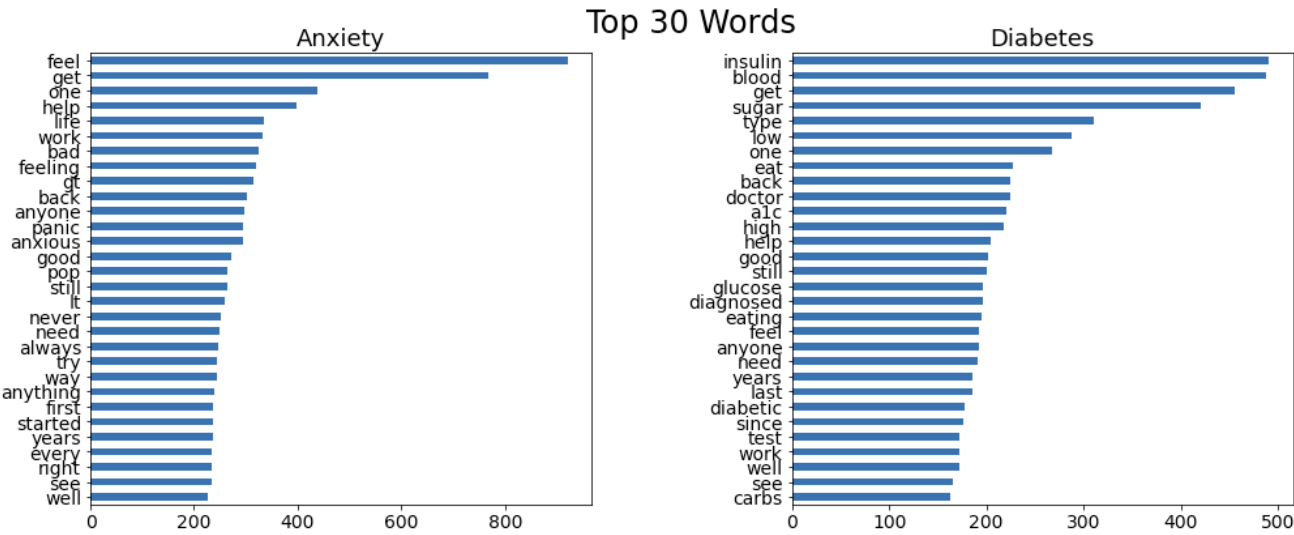
Words like feel, help, panic, anxious, bad, never, were prevalent and important to our analysis as individual words. Coupled words such as 'panic attack', 'side effects', 'feel guilty', 'whole life', 'heart palpitations', 'heart rate', 'intrusive thoughts' for coupled Anxiety words are concerning.

Diabetes has a confident feel towards language, where solution words are most often used. As, 'insulin', 'type', 'doctor', feel used on a lower basis, 'work', 'carbs', 'glucose', 'diet exercise', 'diagnosed', 'sugar levels', 'acting insulin', 'side effects' are all words indicative of different maturity towards their conditions, greater understanding of their symptoms, share of more positive words in terms of outcomes and coping. Where subreddit is used more to share their experiences, than crying for help.

LOOKING AT OVERLAPS

Overlapping coupled words such as 'anyone else', 'need help', 'side effects'.

Overlapping coupled time related words: 'last night', 'years ago' are significant to understand how it's a daily struggle with their conditions and long lasting.

On anxiety point of view they are more confused, looking for help and assistance. Feel, get, larger amounts of word count, used for expressing their feelings.

Top coupled used words

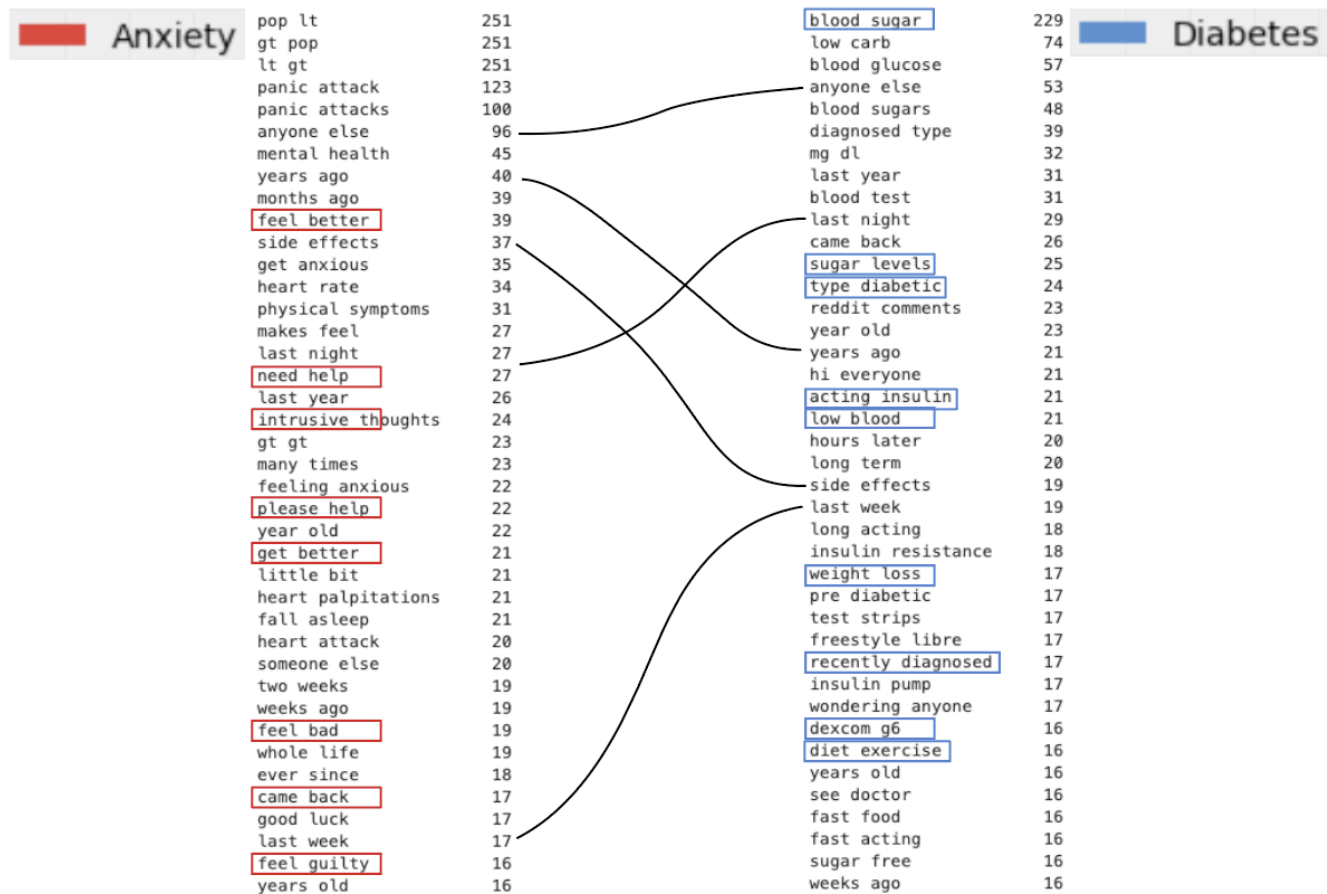|	Anxiety			Diabetes	
	pop lt	251	blood sugar	229	
	gt pop	251	low carb	74	
	lt gt	251	blood glucose	57	
	panic attack	123	anyone else	53	
	panic attacks	100	blood sugars	48	
	anyone else	96	diagnosed type	39	
	mental health	45	mg dl	32	
	years ago	40	last year	31	
	months ago	39	blood test	31	
	feel better	39	last night	29	
	side effects	37	came back	26	
	get anxious	35	sugar levels	25	
	heart rate	34	type diabetic	24	
	physical symptoms	31	reddit comments	23	
	makes feel	27	year old	23	
	last night	27	years ago	21	
	need help	27	hi everyone	21	
	last year	26	acting insulin	21	
	intrusive thoughts	24	low blood	21	
	gt gt	23	hours later	20	
	many times	23	long term	20	
	feeling anxious	22	side effects	19	
	please help	22	last week	19	
	year old	22	long acting	18	
	get better	21	insulin resistance	18	
	little bit	21	weight loss	17	
	heart palpitations	21	pre diabetic	17	
	fall asleep	21	test strips	17	
	heart attack	20	freestyle libre	17	
	someone else	20	recently diagnosed	17	
	two weeks	19	insulin pump	17	
	weeks ago	19	wondering anyone	17	
	feel bad	19	dexcom g6	16	
	whole life	19	diet exercise	16	
	ever since	18	years old	16	
	came back	17	see doctor	16	
	good luck	17	fast food	16	
	last week	17	fast acting	16	
	feel guilty	16	sugar free	16	
	years old	16	weeks ago	16	

THEY ARE IN DIFFERENT TIME STAGES:

One has already initialized their process of recovery in a way. Was already pre-diagnosed, has clear understanding of their conditions, already initialized their process of recovery in a way. On anxiety point of view they are more confused, using time terms that are less specific and a bit subjective, looking for help and assistance.

Anxiety top time words are :

9th: years ago

10th: month ago
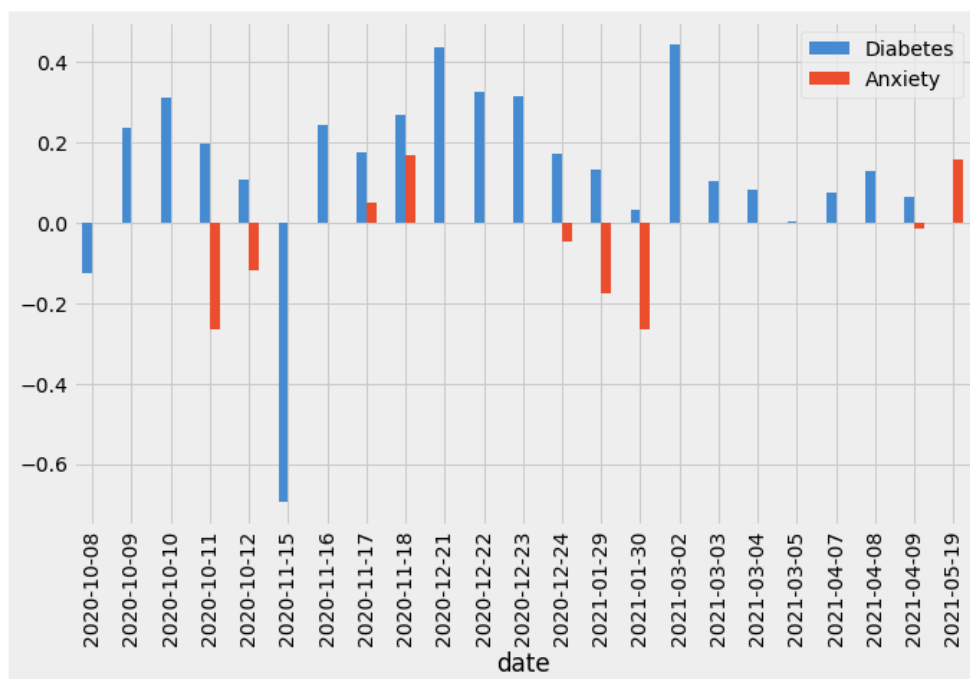
Diabetes top time words are :

7th: last year

9th: last night

15th: years ago

19th: hours later

Also indicative of maturity towards their conditions, reporting of the facts as a coping mechanism.
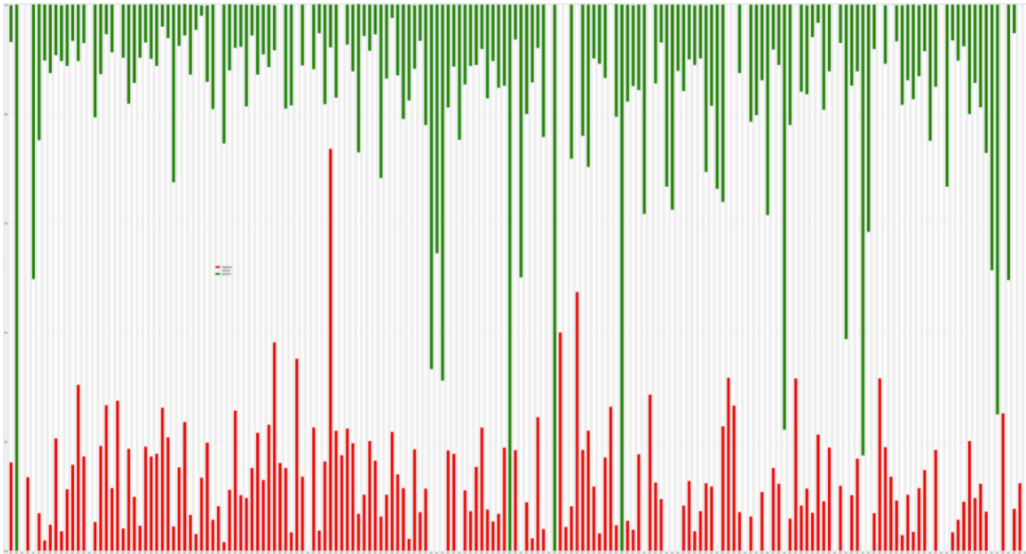
SENTIMENT ANALYSIS

Having a detailed understanding of the sentiment of each texts, diabetes has a positive connotation on their posts while anxiety has a more negatively charged posts as well as positives too.
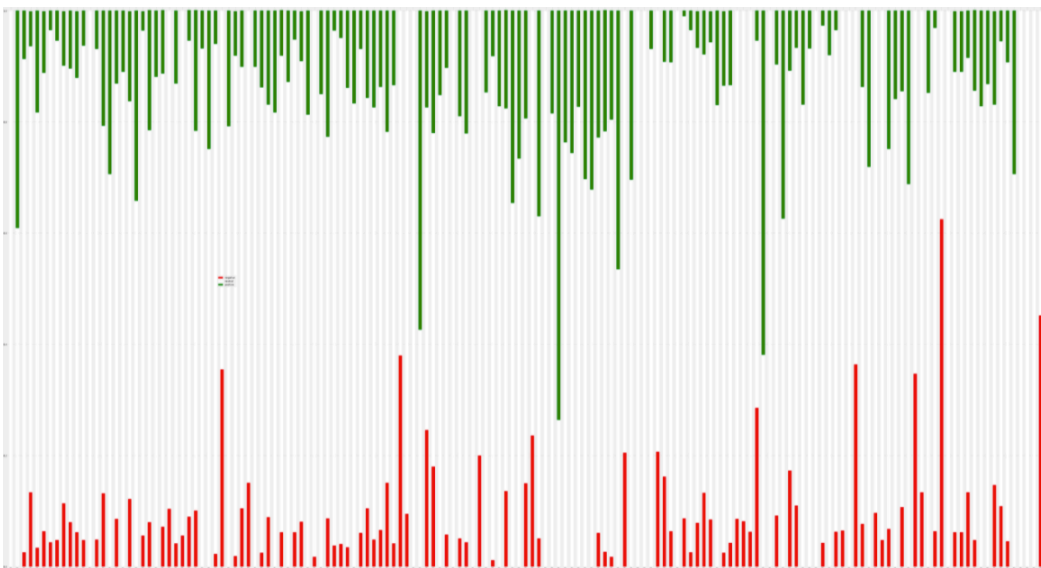
OVERALL LOOK AT SENTIMENT OF POSTS

In the search for trends, which the beginning of October and November 2020 where covid-19 lock downs were more severe. You can visually spot more negatively charged than after January 2021 negatively charged posts are more spread out.
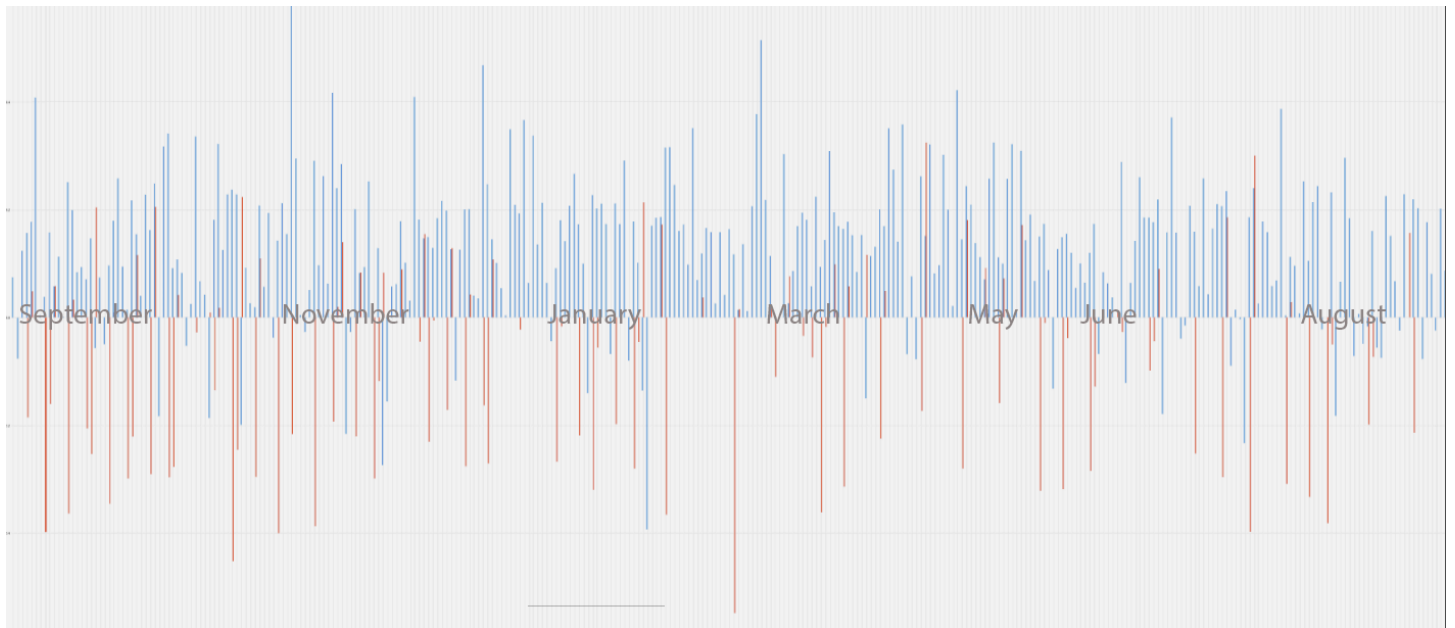
NOVEMBER ANXIETY POSTS SENTIMENT

NOVEMBER DIABETES POSTS SENTIMENT

This visual is indicative of how SentimentIntensityAnalyser is working defining compound sentiment value. Compound value takes the values of negative, positive and neutral values and sets for the higher value as the overall sentiment. This visualization gives a more accurate sentiment to each posts.

PREDICTION RESULTS

Predicting from which subreddit (Anxiety or Diabetes) each post was posted from.

MultinomialNB with TfidfVectorizer:

{'tvec__max_features': 5000,
'tvec__ngram_range': (1, 1),
'tvec__stop_words': 'english'}

Test score: 0.892

MultinomialNB with CountVectorizer:

{'cvec__max_df': 0.9,
 'cvec__max_features': 5000,
 'cvec__min_df': 3,
 'cvec__ngram_range': (1, 1)}

Test score: 0.896

KNeighbors:

{'knn__n_neighbors': 9,
'tvec__max_features': 5000,
'tvec__ngram_range': (1, 1),
'tvec__stop_words': None}

Test Score: 0.708

Random Forest: (small dataset)

{'rf__max_depth': None, 'rf__max_features':
'sqrt', 'rf__n_estimators': 100,
'tvec__max_features': 2000,
'tvec__ngram_range': (1, 2),
'tvec__stop_words': 'english'}

Test score: 0.835

Random Forest: (larger dataset)

{'rf__max_depth': None,
 'rf__max_features': 'sqrt',
 'rf__n_estimators': 80,
 'tvec__max_features': 2000,
 'tvec__ngram_range': (1, 3),
 'tvec__stop_words': 'english'}

Test score: 0.878

CONCLUSIONS

The 40% of patients with diabetes have heightened anxiety symptoms reported from The Journal of Psychosomatic Research, is visible throughout our text analysis, existence of coupled words overlap are both evidence of correlation mentioned in the Journal of Psychosomatic Research, which there is an overlap of words, but influences little on classification predictions. Our predictions are at range of 89% testing scores, which considering the correlations and causalities it's an overall 'good enough' prediction score.

THE CALL-TO-ACTION

Promoting mental health: Offering a variety of services and programs, moderated online discussion forum. Possibly an official support group for anxiety, such as which medications might be helpful and how to dial with specific symptoms.

Screening tools: Offer screening tools that can help you identify if you might have a mental illness, such as anxiety or PTSD. Giving a few extra questions to answer about your symptoms. Having actionable advice to talk to your physician or a mental health professional.