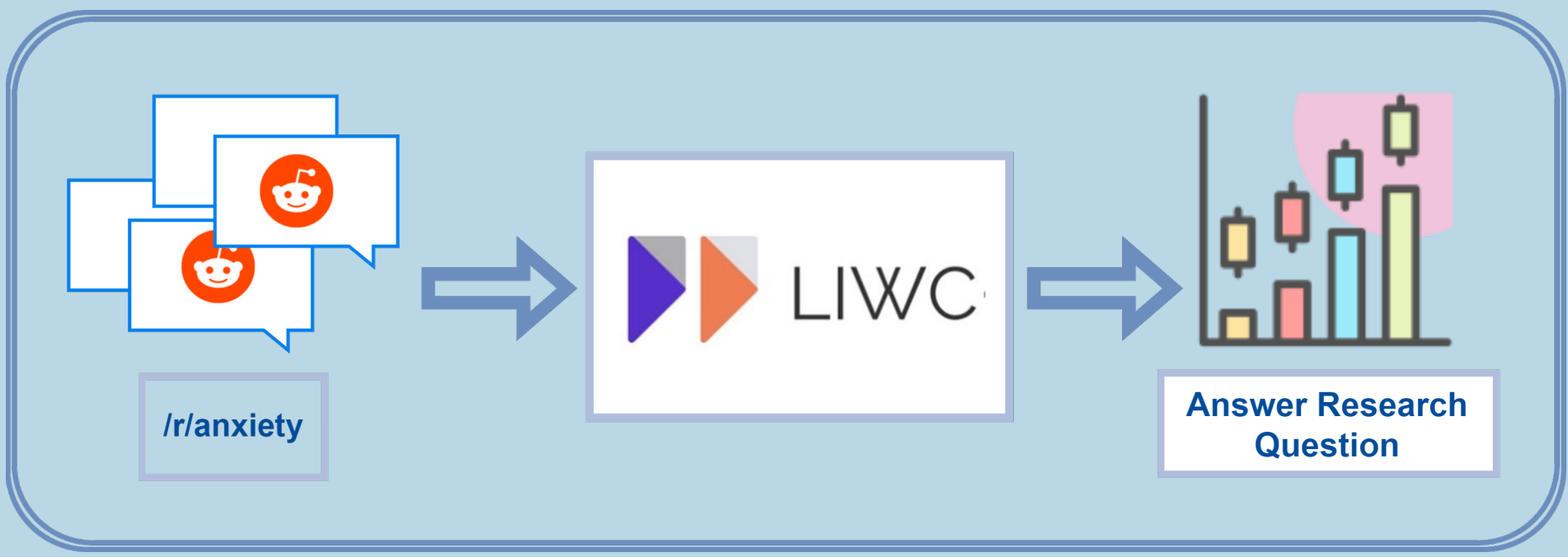


# Did COVID-19 Have a Significant Impact on the Mental Health Community 'Anxiety' on Reddit?



Group 33 - Thomas Aumüller, Muhammad Raisul Hassan, Golam Mostofa, Markus Perz  
thomas.aumueller@student.tugraz.at, muhammad.hassan@student.tugraz.at, g.mostof@student.tugraz.at, markus.perz@student.tugraz.at

## Concept



The COVID-19 pandemic left a strong impact on the lives of many worldwide. Due to lockdowns and restricted social interactions, the suspicion of increased development of mental health problems quickly rose. This work compares user-authored content posted on the **r/Anxiety** Reddit community shortly before and during the early phase of COVID-19 to identify user behaviour and evaluate state-of-the-art sentiment analysis tools.

## Related Work

Low et al. [1] analyzed mental health posts from 826,961 unique Reddit users, utilizing regression and supervised machine learning to identify trends and key features. Their study revealed early spikes in COVID-19-related posts highlighting shifts in themes such as economic stress, isolation, and suicidality, while topics like motion declined. In [2] similar Reddit communities were analyzed using topic modelling, Word2Vec embeddings, and various trend and thematic analysis techniques (e.g., time-to-event, regression, k-means clustering). Results suggest a 28-day post-event window for mental health concerns and economic stress as a primary suicide predictor.

## Models

**LIWC-22.** Linguistic Inquiry and Word Count is a dictionary-based text analysis framework. It classifies words into psychologically meaningful categories and quantitatively measures linguistic and psycho-social processes.

**BERT.** Bidirectional Encoder Representations from Transformers is a deep neural model that processes text in both directions commonly used for natural language tasks.

**VADER.** Valence Aware Dictionary and Entiment Reasoner is a lexicon and rule-based tool tailored for social media content calculating sentiment polarity.

## References

[1] Low D. et al., *Natural Language Processing Reveals Vulnerable Mental Health Support Groups and Heightened Health Anxiety on Reddit During COVID-19: Observational Study*, Journal of Medical Internet Research, Vol 22, 12.10.2020

[2] Zhu J. et al., *Investigating COVID-19's Impact on Mental Health: Trend and Thematic Analysis of Reddit Users' Discourse*, Journal of Medical Internet Research, Vol 25, 12.07.2023



## Datasets

The datasets used consist of **Submissions and Comments** posted in the community of **r/Anxiety** on the Social News Platform Reddit. This data was part of a manually retrieved dataset as part of a seminar work, crawled via the official Reddit API. Included in the dataset are **216,128 unique users**, **183,093 Submissions** and **1,023,554 Comments** ranging from January 2018 to June 2021.

**January 2018 - December 2019:** Representing a time frame before the COVID-19 pandemic  
**January 2020 - June 2021:** Reference data for posts after the COVID-19 pandemic

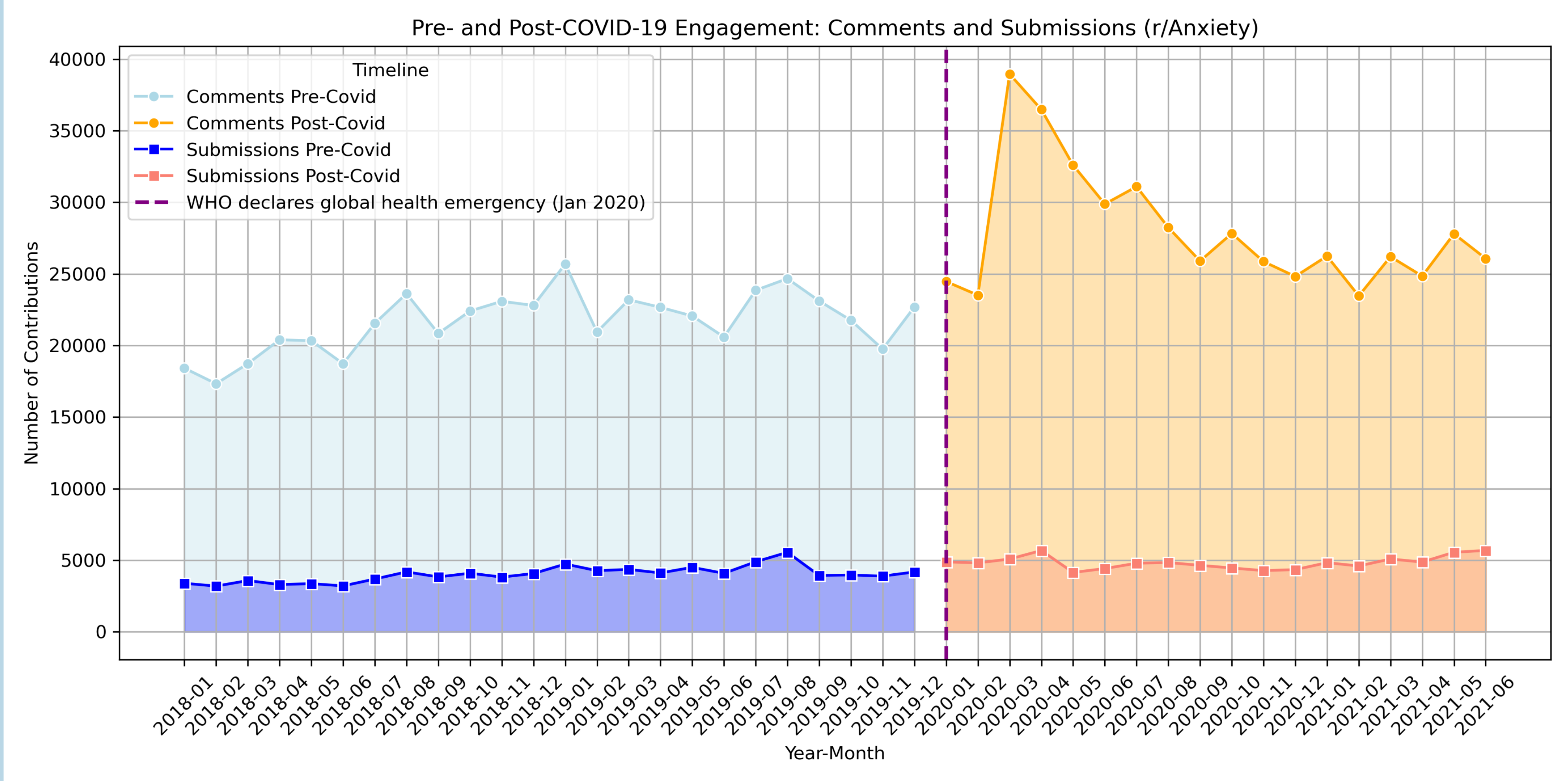


Figure 1: Engagement in r/Anxiety before and after COVID-19.

## Results

The results of the sentiment scores before and after the COVID-19 outbreak were calculated with **LIWC-22**, **BERT** and **VADER**. Visual observations hinted at some interesting trends such as a noticeable increase in the number of comments, thus statistical tests were conducted to further evaluate the impact of the pandemic on sentiment scores in this specific mental health community.

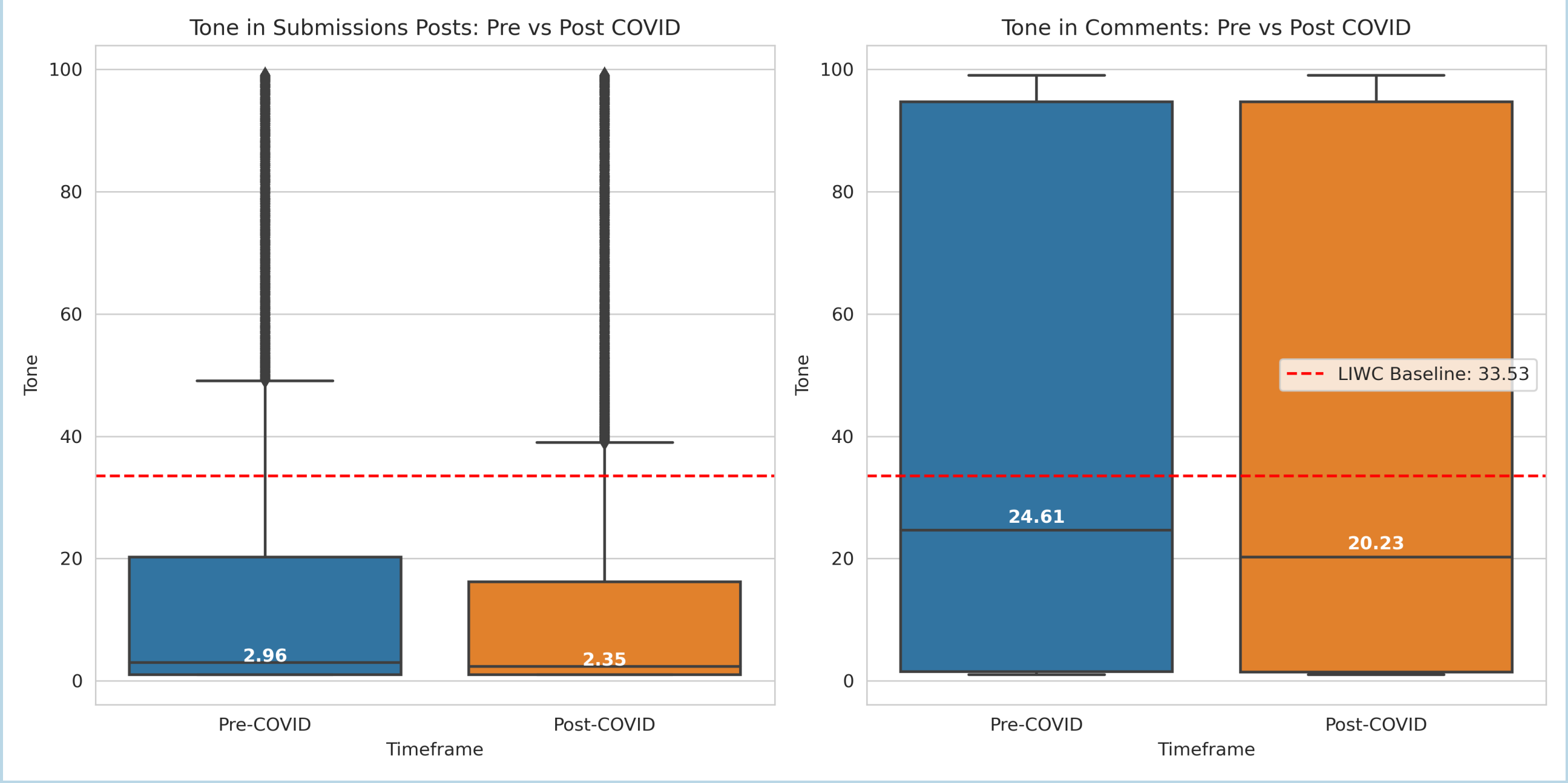


Figure 2: Comparison of LIWC Tone (Sentiment) in Submissions and Comments before and after COVID-19.

Changes in sentiment score were observed to be statistically significant after comparing sentiments before and after covid with the **Mann-Witney U Test**. Furthermore, a conducted **T-Test** for all sentiment permutations confirmed these significant changes in submissions for BERT score, VADER compound, and LIWC Tone. The comment sections only showed significant differences in their BERT scores. When calculating the magnitude of change in the sentiment scores with the **Cohen's d Test**, the effect sizes proved to be generally quite small across all sentiment scores. The largest effect size was observed for Tone in Submissions ( $d = 0.045$ ), which still indicates minimal practical significance. Ultimately, while statistically significant differences were detected, the magnitude of sentiment changes was minimal, as Cohen's d effect sizes indicated. These findings highlighted nuanced shifts in sentiment and shed some light on the increased engagement in the comment sections of posts in the community of r/Anxiety after COVID-19.

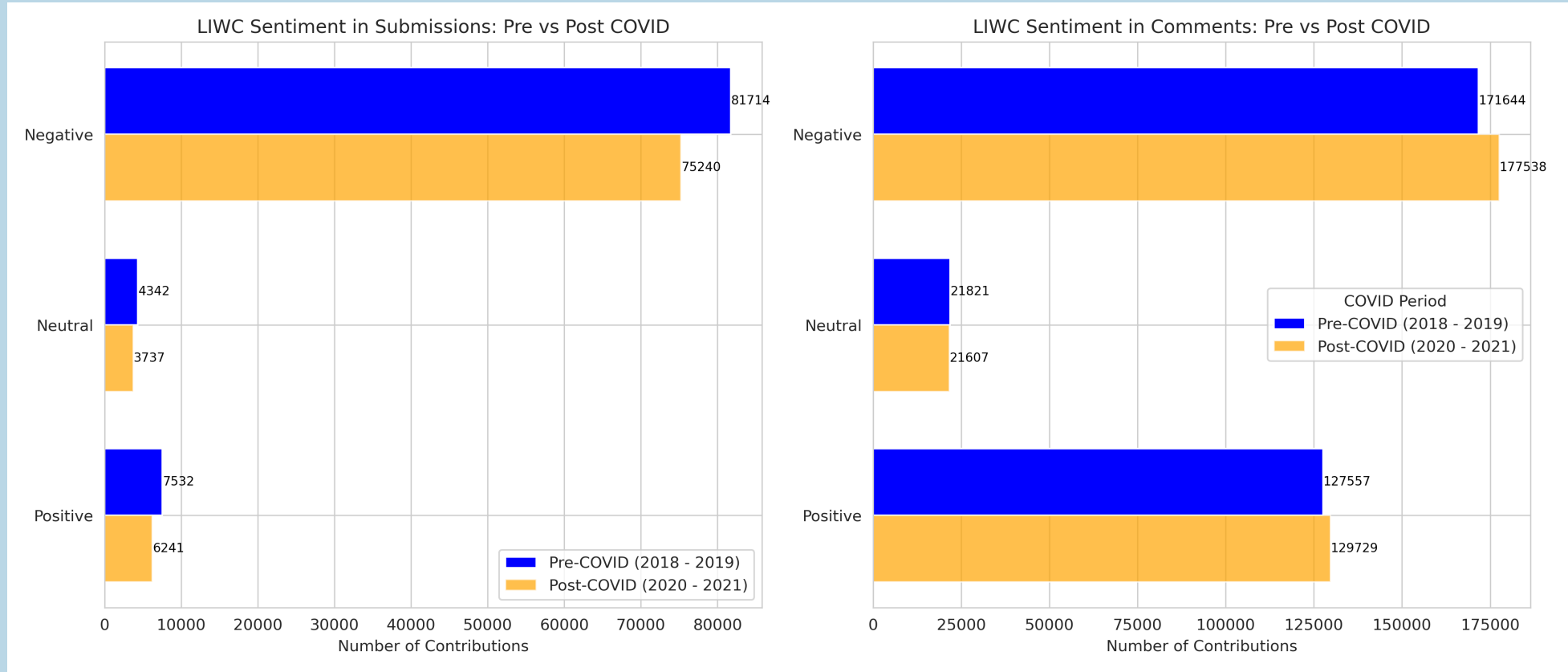


Figure 3: LIWC Tone of Submissions and Comments.

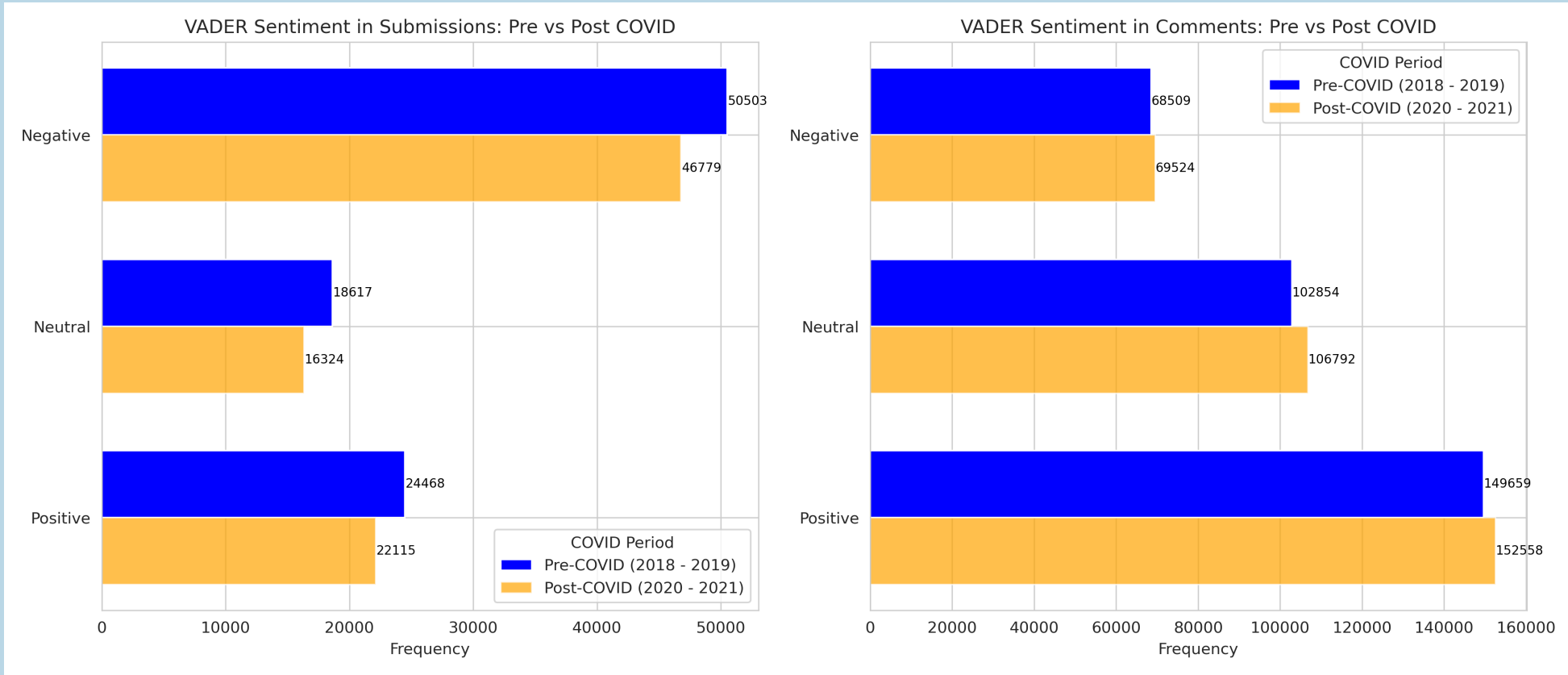


Figure 4: VADER Sentiment of Submissions and Comments.