

Social Media Data Analysis on Autism Spectrum Disorder

1. Introduction

With the new normal of living with Covid19, mental illness and suicide have increased dramatically (Wang, 2020). Out of many mental illnesses, one of the disorders which have a high suicidal tendency is Autism Spectrum Disorder. Computer science's own, Alan Turing committed suicide so as did Robin Williams and Vincent van Gogh. In academia, students learn many achievements from people with ASD or people suspected to have ASD: Hans Andersen, Emily Dickinson, Lewis Carroll (Author of Alice in Wonderland), Mark Twain and Issac Asimov (Author of I, Robot) in Literature, Wolfgang Mozart, Ludwig Beethoven, Bob Dylan in Music, Pablo Picasso, Vincent Van Gogh, Michelangelo and Andy Warhol in Fine Arts, Thomas Jefferson, Benjamin Franklin and Abraham Lincoln in political science, Steven Spielberg, Alfred Hitchcock, Stanley Kubrick and Tim Burton in Film, Bill Gates, Steve Jobs and Alan Turing in Computer Science, Issac Newton in Math, Albert Einstein, Thomas Edison, Charles Darwin and Nikola Tesla in Science (Blakey n.d.; Cerny n.d.; Joshi 2016; Fattig 2017). But in the modern era who comes to your mind when you say, Tesla? Tesla motors' founder Elon Musk recently announced that he has ASD.

So what exactly is ASD, Autism Spectrum Disorder/Aspergers? According to DSM-5, the newest edition of the Diagnostic and statistical manual of mental disorder by the American Psychiatric Association, ASD is characterized by "Persistent deficits in social communication and social interaction across multiple contexts" and "Restricted, repetitive patterns of behaviour, interests, or activities (Autism n.d). In 2018 Elon Musk has been sued for an offensive comment he made on Twitter regarding the Tham Luang cave rescue in Thai when 12 boys got trapped in the cave and the international rescue effort got attention around the world. Even though the mental illness does not justify his offence and crime, the price he paid for his communicational ability deficit marked by ASD is 190 million U.S. dollars (Reuters 2019; Lee 2019). As Elon Musk posted on Twitter, with the advent of Web 2.0, there is more need to understand its influence on mental illness. Social media could be beneficial for people with ASD since they have a chance to connect with other Autistic people and form their own culture with text communication rather than face-to-face verbal communication. Although in the online setting people can leave a comfortable situation with one click in the worst case, Autistic people tend to take everything literally and what happens in social media tends to affect their real life. They also have repetitive patterns of behaviour and interest that they can get obsessive about and cannot find the balance. The main focus of this project is to understand how people with ASD or suspected to have ASD are interacting with social media sites.

2. Background

According to some studies conducted on Children with ASD in regular class (in -person), Children with ASD tend to lose the connection with their classmates over time and receive more

non-preferred nominations by classmates as shown in Fig 1 in Appendix A (Anderson 2015). In Fig 1, the social network graph on the left shows the connection among students in term 1 and the left side shows the connection in terms. James has considerably less connection in term 2. When closely examined, the only connection that James managed to keep was one with Emily and the other one is a newly formed connection. Fig 2 shows that children with ASD have lower nomination by classmates when the survey was conducted to indicate who they like to play with and who they don't like to play with. Although the preferred nomination stayed evenly low throughout the school year, the non-preferred nomination rose over time steeper compared with children with other disorders as shown in Fig 3 (Locke 2013). Is this a problem? In 2011, Dr. Dawson from UBC developed Social Networks Adapting Pedagogical Practice (SNAPP) to identify the students who are isolated and help educators to develop an intervention if necessary (Bakharia 2011). How do you know when it is necessary?

One study shows that despite children with ASD experiencing lower involvement, companionship, reciprocity and acceptance, they did not report greater loneliness (Chamberlain 2007). The other study indicated that college students with higher ASD traits tend to use inappropriate manners for SNS and it was associated with loneliness. This study was conducted on their manner during LINE use (Suzuki 2021). The problem is that further study is needed to determine if and when you need to intervene and to leave them alone.

3. Data, Methodology and Result

The social media site used in this project is Reddit. Reddit is an American social news website where members can submit text posts, images, links and videos and members can vote up or down the posts which will be aggregated and organized by subject in user-created boards called subreddits. Reddit's covered a wide variety of topics and people can identify themselves with flair. It also has strict rules to prohibit abusive behaviour (Reddit n.d.). Reddit was the best one since it has multiple subreddits for ASD and Aspergers and it is a safer environment for people with mental disorders with built-in rules. One of the rules in the subreddit, Aspergers is to help prevent suicides and it is prohibited to promote or justify suicides. Below is the list of ASD-related subreddits

aspergers	109,504 members	created Mar 24, 2010
aspiememes	81,700 members	created Jan 9, 2018
aspergirls	47,867 members	created May 29, 2014
aspergers_dating	5,966 members	created Apr 7, 2013
asd	5,433 members	created Aug 6, 2009
aspiepositivity	3,300 members	created Jun 9, 2020
asperger	2,880 members	created Nov 14, 2010
Aspie	1,100 members	created Aug 13, 2009

According to the current version of DSM, Asperger's syndrome is now called Autism spectrum disorder, ASD. However as you can see on the list, most people who are related to ASD like to call it Asperger. Although in this project I will use ASD in order to be consistent with

DSM-5, this indicates that people with ASD have their own voice on Reddit and they don't care how experts label them.

I used Reddit API, Praw to collect data on a user's post including comments, the number of upvotes and downvotes, the time the post was created and aggregated score. Originally I tried evaluating the social network of friends compared to the social network graph from a regular classroom. Children with ASD have trouble sustaining the social connection and seeing if there are any differences in connection in social media. Also, I wanted to identify the most isolated user who does not actively participate in social media. Even if they do not actively participate and therefore they are technically not part of the culture, it might be more beneficial to be aware that other people with ASD share the same problem and see how they deal with it. However, this data was not available on Reddit. Therefore I need to change my project to use the data from users who actively participate in the subreddit in order to obtain data.

For this project, there are three methods used to examine user relationships and user patterns on Reddit which are social network analysis, sentiment analysis and regression analysis between user activity and the impact of reaction over time.

The first one is network analysis shown in Appendix B. Data is collected from a subreddit called aspergers and there are two categories, top and hot. The network graph is formed using a network with 20 nodes. Degree centrality is a measure of the average number of edges of a node in a graph. The average degree of centrality was 0.043. However when it is compared with popular subreddit "movies" and "worldNews", there were no significant differences. Betweenness and closeness were extremely close to 0. It successfully visualizes the relationship and cluster between users who made comments. The biggest finding was that the hot category is more likely to have higher connections than the new category.

The second one is sentiment and content analysis shown in Appendix C. Sentiment analysis is conducted using Vader sentiment analysis from the Vader lexicon in nltk package and written out as a bar graph. Three different subreddits related to ASD are compared with three popular subreddits and there were no significant differences. Even though there are no significant differences, some people get negative impressions from asperger related subreddits. There is a subreddit called aspie positivity which implies that some users see the need to create a subreddit dedicated to positivity. In order to further examine the sentiment, content analysis is performed.

For the content analysis, all the comments from ASD-related subreddits are collected and text is preprocessed, tokenized, and stemmed using porter stemmer. For example, social, social and socialize are considered the same word as well as feel and feel. Also, the words less than 3 letters are eliminated so words such as "the", "I", "am" and "to" do not appear on the top of the list. Some of the words that were used most often were, people, think, understand, always, they, this, have, that, myself, some, what, because, social, never, attractive, conventionally, police, sex, communication and friend. Then I tried to apply a Tfidf vectorizer in order to make it relative to the length of the document. Even though further work is necessary to get the relative count to document length, the words used most often in comments had interpersonal and relationship nature.

The third one is regression analysis between a member's activity overtime and the positive or negative reaction they get from their posts shown in Appendix D. User comments, time and score are written out to excel as a data frame in CSV format and a scatter graph with

those three values are created in excel. Mainly there were three different patterns in which the user actively increased, decreased or unrelated to the feedback and reactions. I tried to build Bayesian networks in order to find the probability given the reaction, but I was not successful on this. The most interesting finding in regression analysis is that the average duration of user activity is about 1.63 years. Even among those people, some members do not participate constantly with long inactive time and their activity can be very scarce. So it might not be as simple as subtracting the newest post time from the first post date.

4. Implication from the findings and Discussion

There were some patterns in the change of activity which correlate to the reaction they got in prior posts. There are mainly three cases to consider. One possibility is that those who are not successful in an in-person social relationship but they find people who are similar to themselves and get good reactions on social media that their activity increases. These are the people that you can intervene and probably it doesn't hurt. The second type of people is those whose activity decreases for a while after they get high positive reactions. You need to further examine what kind of positive reaction they got in the ASD's point of view. The sentiment analysis of neurotypical people is not sufficient. It is especially important to be careful to intervene with people who never came back to actively participate after the reaction. However, most people have come back after a while and they got used to getting reactions. So we need to pay close attention when we intervene for the first time and how to intervene. The last group of people tend not to care if they get a reaction or not nor what kind of reaction such as positive or negative. These people might not be people with ASD and there should be no harm to intervene.

The other finding that was consistent across all types is that the average duration of active participation was about 1.63 years. If this is the case they only participate in active ASD discussion for a limited amount of time. According to content analysis users are interested in common topics and sentiment is not necessarily negative. So we need to further examine the nature of reaction and comment to see if there are some characteristics and patterns.

5. Evaluation and feedback

In order to evaluate and get feedback, I interviewed my peer students in an unstructured setting. One was an in-person interview with a UBC student (Martin Sung) who is in human kinetics and the other two people who have graduated from the UBC Computer Science department in an online setting (Caroline Zhang and Alex Zhang). The student from human kinetic background grew up in Malaysia and he said Reddit is not very popular in Malaysia. He found some of the findings very interesting but there is not enough evidence to make any concrete decision.

The other evaluator who has a computer science background was also not from Canada and she also mentioned that Reddit has a particular user and it is not common among people who are not fluent in English. She pointed out that other social media is way more popular in the

younger generation and most young people don't use Reddit. She suggested performing cluster analysis on the result of regression analysis in order to further support the hypothesis.

The last evaluator thought the analysis was not sufficient enough and tried to help me with the Bayesian network and cosine similarity with code, but we still could not do this in time. He also suggested that there might be something wrong in degree centrality calculation and aggregation.

6. Fundamental difficulty

The fundamental problem of study on people with ASD and social media is that it is extremely difficult to identify the people who have ASD. People with ASD tend not to choose to get diagnosed and they tend not to voluntarily label themselves with ASD or Asperger. You can only form social network graphs with people who comment on each other.

In order to solve this problem, you need to find people who are diagnosed with ASD and examine their social media use. However, this might be still biased since people with severe ASD tend not to participate in the study and these are the people who got identified when they were children or they chose to get diagnosed. It will exclude the type of ASD people who successfully mask their social and communication deficits. Most ASD people naturally avoid social interaction and they do not make an appointment or show up to go see a typical human to give them a diagnosis.

As suggested in the evaluation, I further looked into the age of Reddit users. Some of the subreddit especially "suicidewatch", have a large number of self reported younger users since they do not have other means to get help. Since you cannot get the age and location of a member in Reddit, it is impossible to prove it.

7. Conclusion and Future Work

The conclusion is that it seems like people who join the subreddit related to ASD are not sustaining the online interaction for the long term just like the in-person relationship. However, in Reddit, they are leaving voluntarily contrary to in-person relationships. So further work is necessary in order to even make any hypothesis. The reaction and feedback to the posts and comments seem to have some kind of impact on the user actively so as neurotypical users. Further work is needed to determine if it differs in degree. If you can successfully identify the pattern of social media use, you might be able to predict who has ASD and other members who joined the subreddit by interest. There was no significant difference in degree centrality or sentiment analysis. However, we need to examine if the atypical measure is applicable to people with ASD who have communication problems and distrust of atypical methods. Even though it gives start point and some idea, we need to be careful if this existing measure is applicable to the people doesn't belong to the mainstream.

There was not much difference from the mid-project milestone report. The main takeaway from this project was to get experience getting data using API, writing it out as a CSV file and performing data analysis. Although it is not at the level I want to be, I got the idea of how

data analysis research and social network analysis works. I still have problems handling large amounts of data, parsing necessary data and iterating successfully and need more experience with using appropriate libraries so I can do what I want to do with data.

If I could do this project again, I will structure it differently from the beginning. I did not know anything about Reddit as a user and most of the data I was planning to use was not available to obtain. Even though I learned so much as it progressed and I think Reddit is successful because it respects users' privacy, you can only obtain a very limited type of data. I also got a more negative impression from the posts in ASD-related subreddits. So I will look into the algorithm of sentiment analysis before I apply it. It was a great experience to get first-hand practice with data analysis and get an overview of how it works.

REFERENCES

- Autism speaks. Autism Diagnosis Criteria: DSM-5. Retrieved November 25, 2021, <https://www.autismspeaks.org/autism-diagnosis-criteria-dsm-5>
- Anderson A, Locke J, Kretzmann M, Kasari C. Social network analysis of children with autism spectrum disorder: Predictors of fragmentation and connectivity in elementary school classrooms. *Autism*. 2016;20(6):700-709. doi:[10.1177/1362361315603568](https://doi.org/10.1177/1362361315603568)
- Blakey, M. Famous people with autism. *Autism*. Retrieved December 1, 2021, from <https://www.aspergerstestsite.com/898/famous-people-autism/>
- Bakharia, A. Dawson, S. P. (2011). SNAPP: a bird's-eye view of temporal participant interaction. Proceeding of the 1st International Conference on Learning Analytics and Knowledge, LAK'11 (pp. 168-173). New York: ACM.
- Cerny, B. History's 30 Most Inspiring People on the Autism Spectrum. *Applied Behavior Analysis Programs Guide*. Retrieved November 28, 2021, from <https://www.appliedbehavioranalysisprograms.com/historys-30-most-inspiring-people-on-the-autism-spectrum/>
- Chamberlain, B., Kasari, C. & Rotheram-Fuller, E. Involvement or Isolation? The Social Networks of Children with Autism in Regular Classrooms. *J Autism Dev Disord* **37**, 230–242 (2007). <https://doi.org/10.1007/s10803-006-0164-4>
- Joshi, S. 17 Famous People Who Proved That Autism Is Much More Than A Condition. *Storypick 27th February (2016)*. Retrieved December 2, 2021, from <https://www.storypick.com/autism-could-be-a-boon/>
- Lee, D., Elon Musk wins defamation case over 'pedo guy' tweet about caver. *BBC 6 December (2019)* Retrieved December 2, 2021, from <https://www.bbc.com/news/world-us-canada-50695593>
- Locke, J., Kasari, C., Rotheram-Fuller, E., Kretzmann, M., & Jacobs, J. (2013). Social network changes over the school year among elementary school-aged children with and without an autism spectrum disorder. *School Mental Health*, 5(1), 38-47.
- Michelle Fattig. (2017, November 10). Famous People Who Have or Had Asperger's Syndrome. *Disabled World*. Retrieved December 1, 2021, from www.disabled-world.com/disability/awareness/famous/asp.php
- Reuters, T., Elon Musk testifies in defamation trial over his tweet about British cave diver. *CBC Posted: Dec 03, 2019, 7:18 AM ET | Last Updated: December 3, (2019)* Retrieved December 1, 2021, from <https://www.cbc.ca/news/world/elon-musk-defamation-trial-1.5382005>

Suzuki, K., Oi, Y. & Inagaki, M. The Relationships Among Autism Spectrum Disorder Traits, Loneliness, and Social Networking Service Use in College Students. *J Autism Dev Disord* 51, 2047–2056 (2021). <https://doi.org/10.1007/s10803-020-04701-2>

Wang X, Hegde S, Son C, Keller B, Smith A, Sasangohar F. Investigating Mental Health of US College Students During the COVID-19 Pandemic: Cross-Sectional Survey Study. *J Med Internet Res*. 2020 Sep 17;22(9):e22817. DOI: 10.2196/22817. PMID: 32897868; PMCID: PMC7505693.

Wikipedia contributors. (2021, November 28). Reddit. In *Wikipedia, The Free Encyclopedia*. Retrieved 05:52, December 2, 2021, from <https://en.wikipedia.org/w/index.php?title=Reddit&oldid=1057550082>

Appendix A: Background study

Fig 1

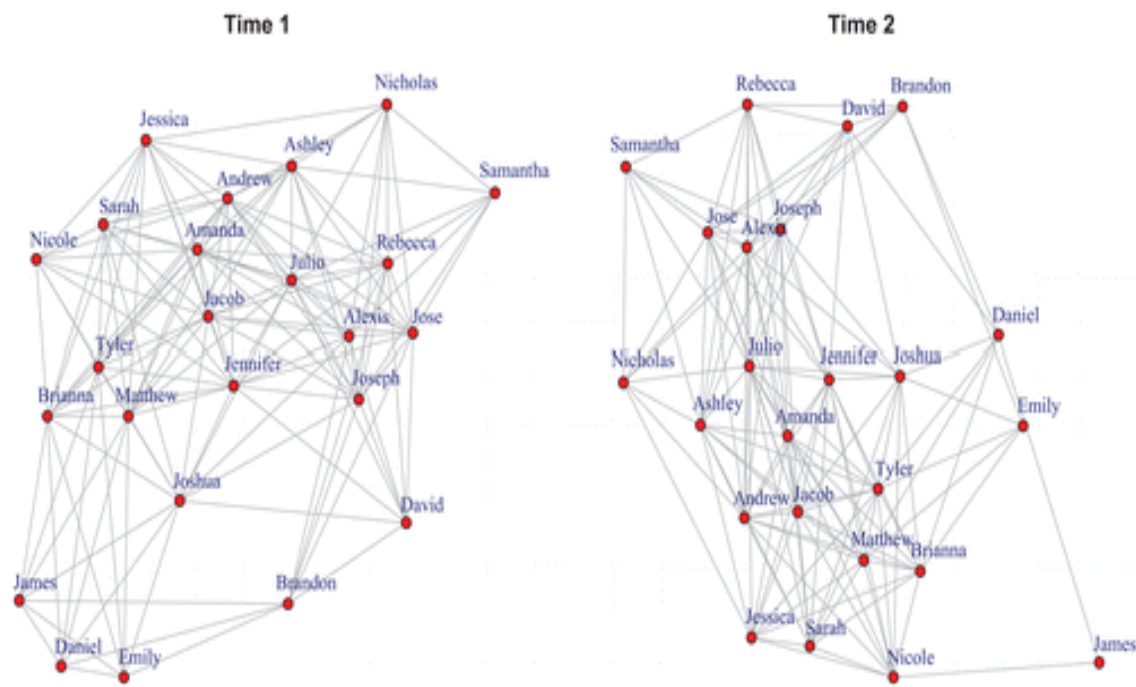


Fig 2 and Fig 3

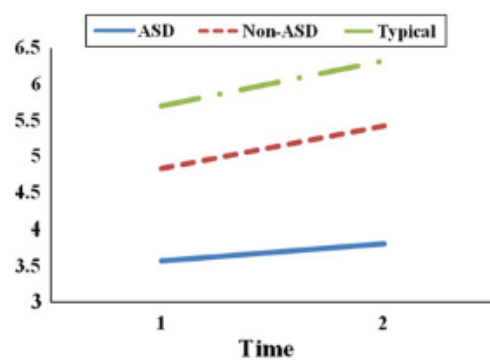


Fig. 2 Number of received friendship nominations (*indegrees*) by classmates between groups across one academic school year

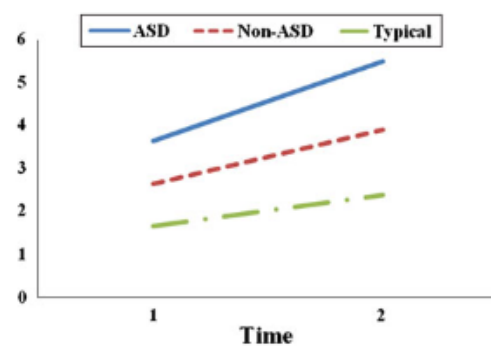


Fig. 3 Number of received non-preferred nominations by classmates between groups across one academic school year

Appendix B: Network analysis

Fig 4 and Fig 5 : Subreddit “aspergers” - a new category

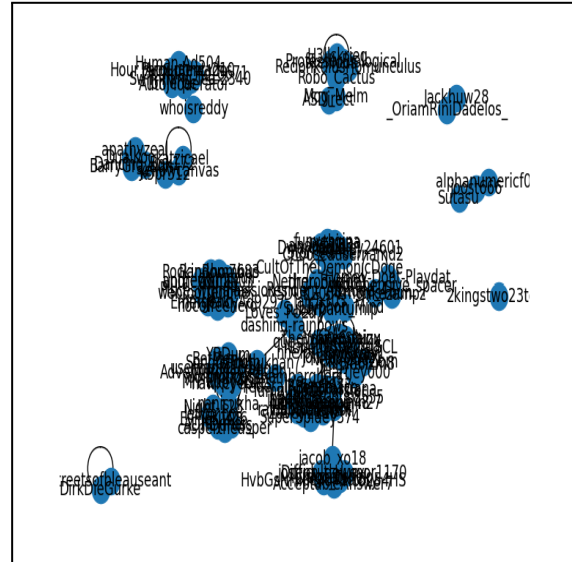
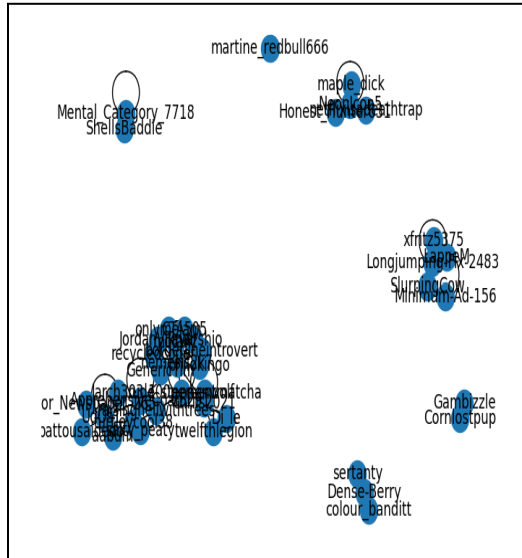
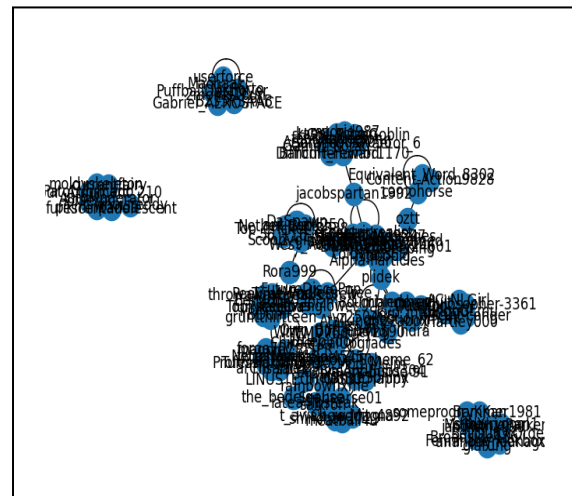
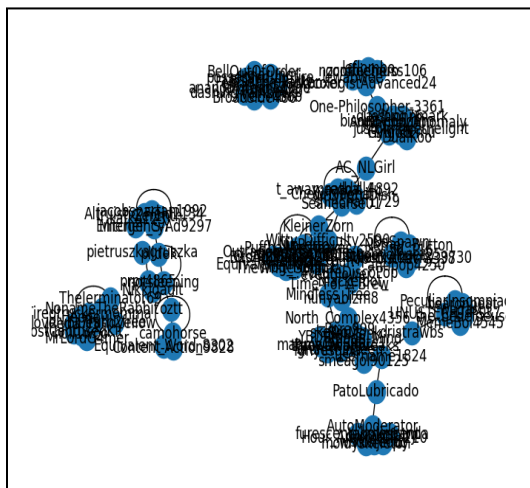
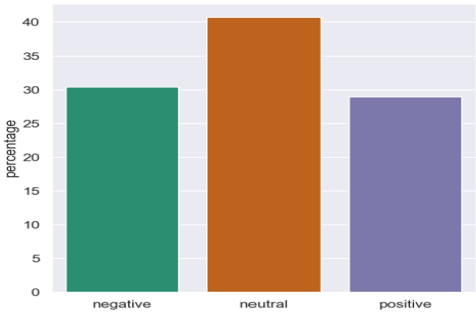


Fig. 6 and 7: Subreddit aspergers - hot category

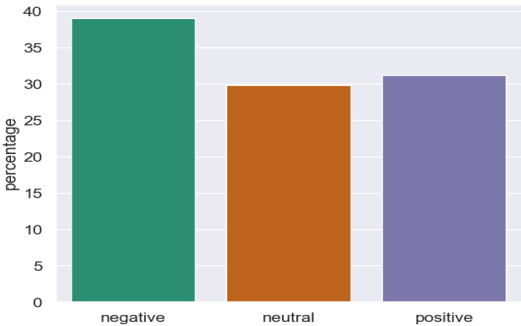


Appendix C: Sentiment Analysis

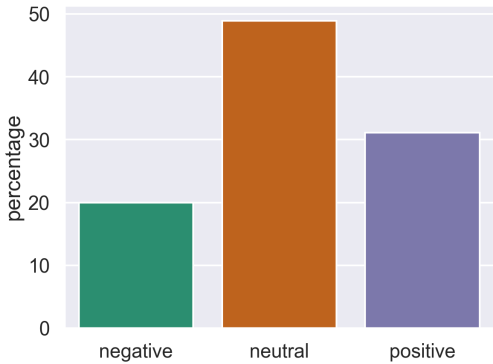
Aspergers Hot



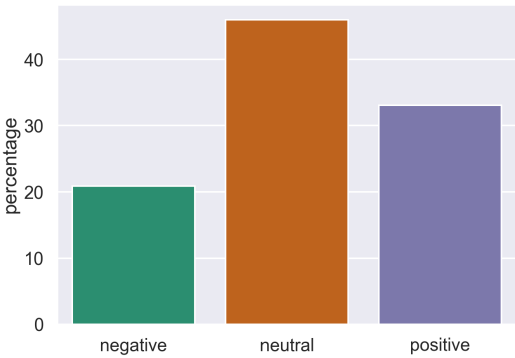
Asperger's Top



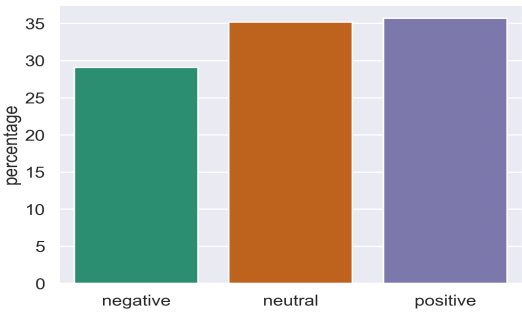
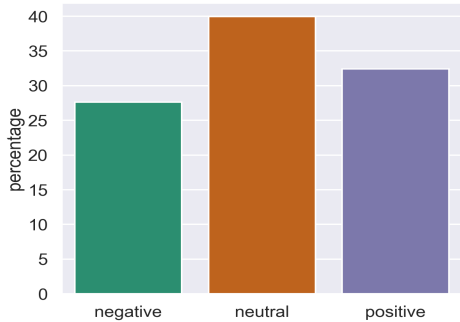
Aspiememes Hot



Top

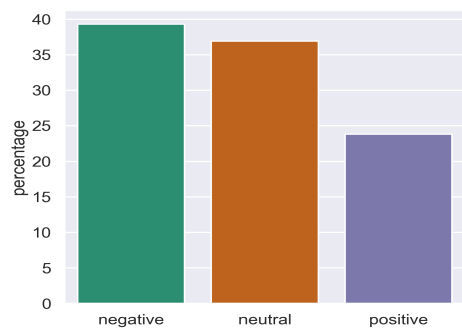


Aspergirls hot

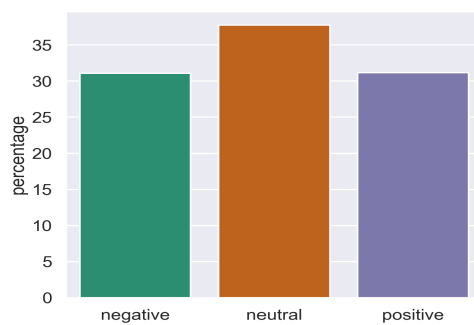


Canada hot

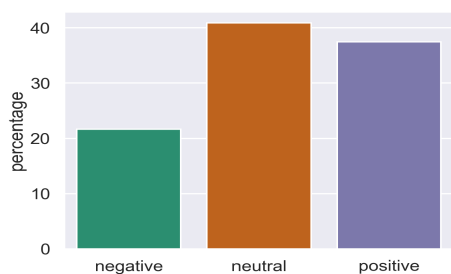
Canada top



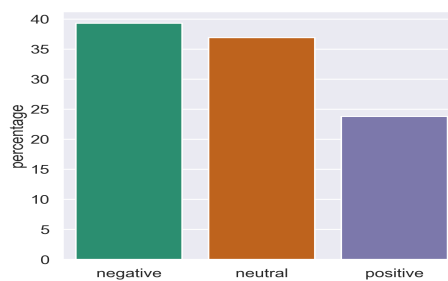
Movie hot



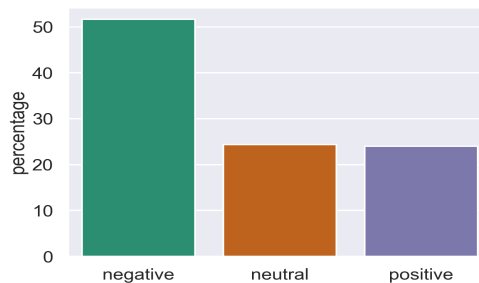
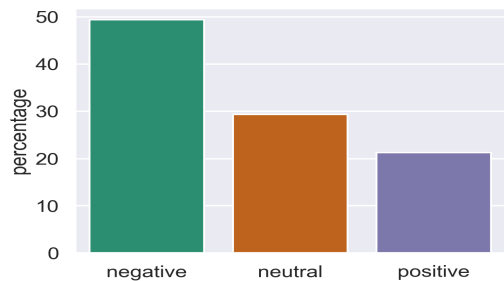
top



Worldnews hot

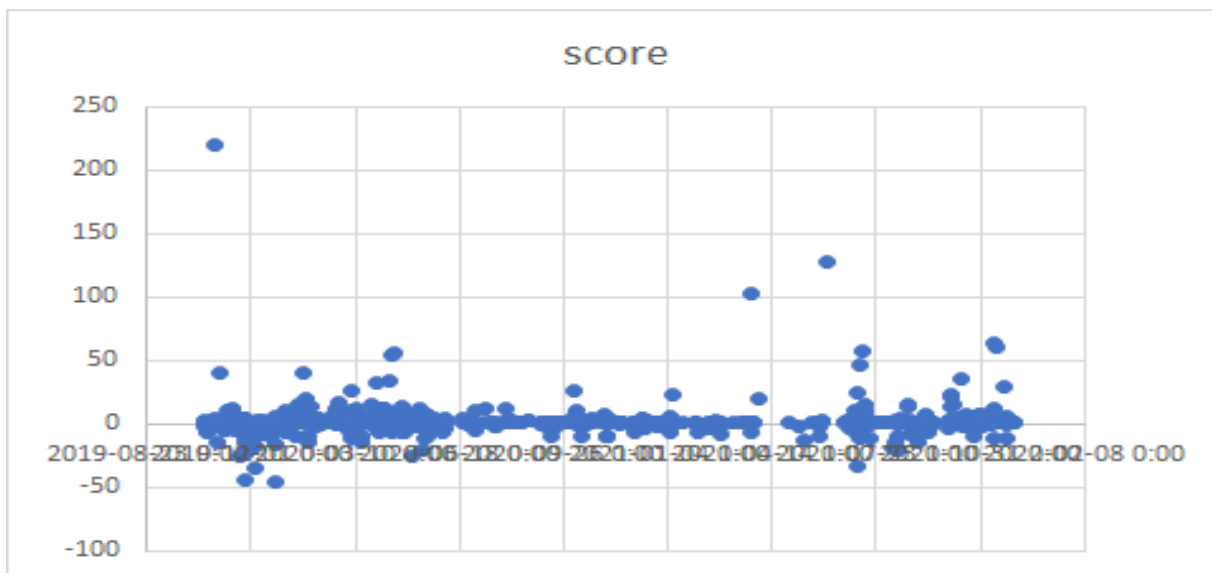
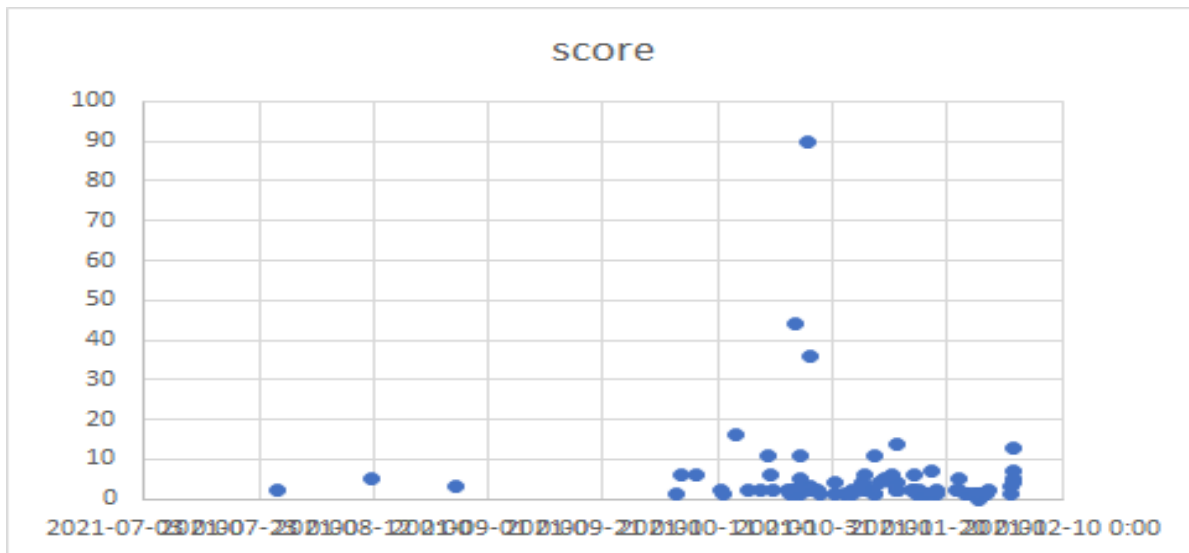


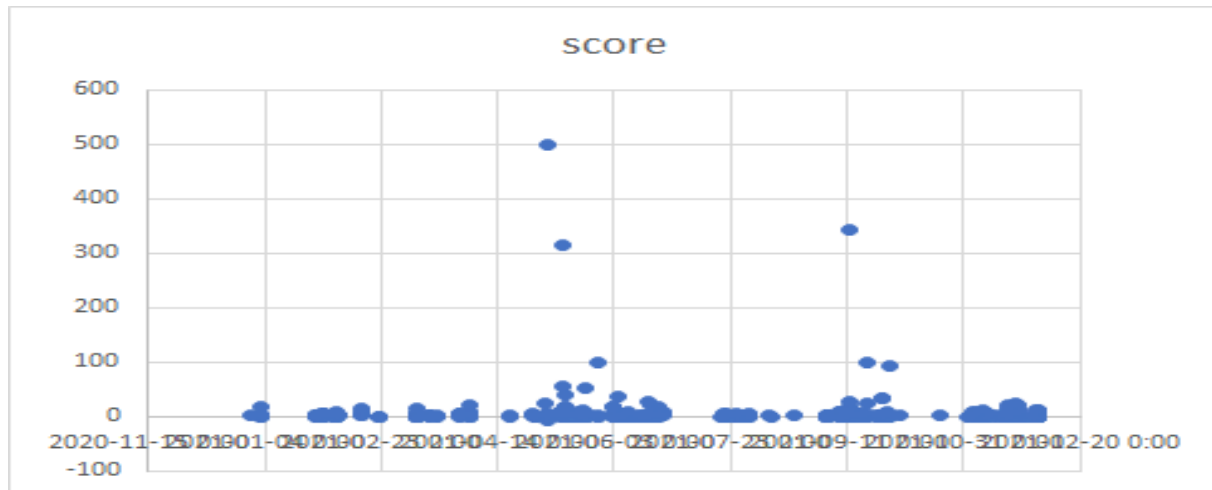
top



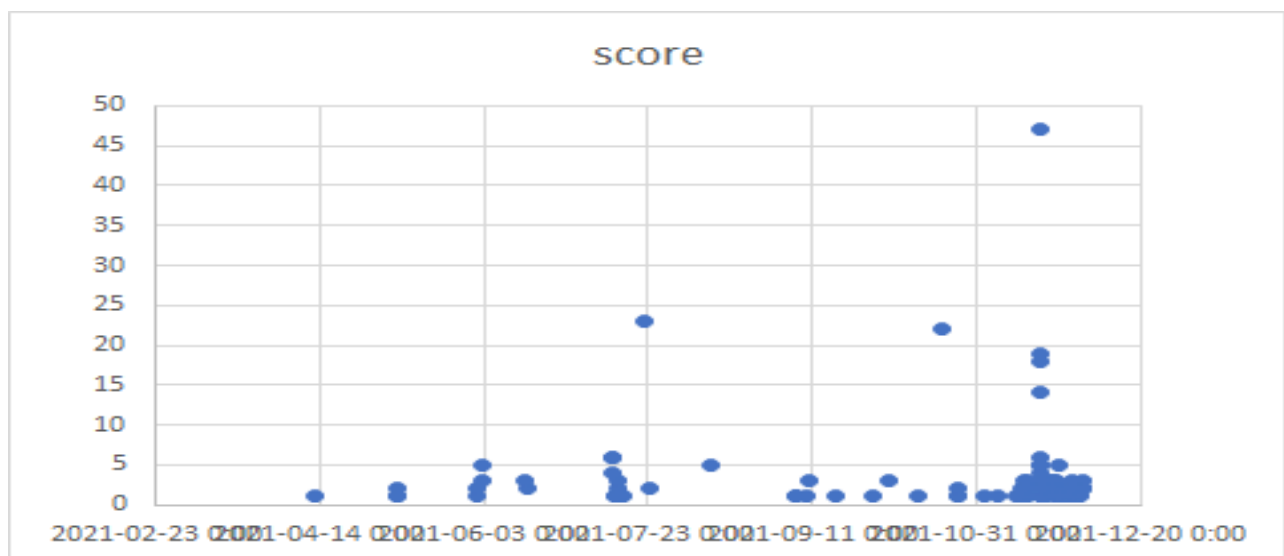
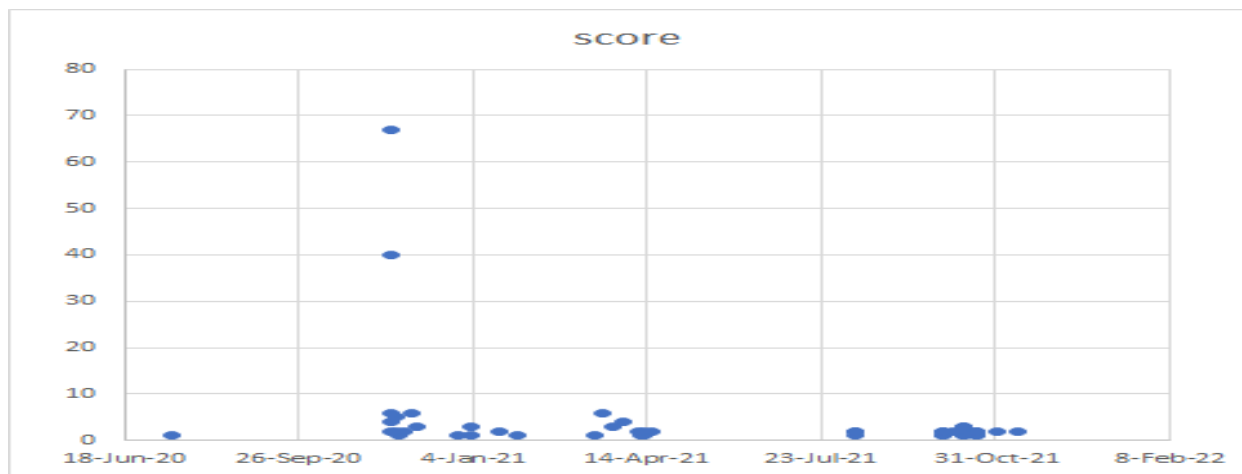
Appendix D: Regression Analysis

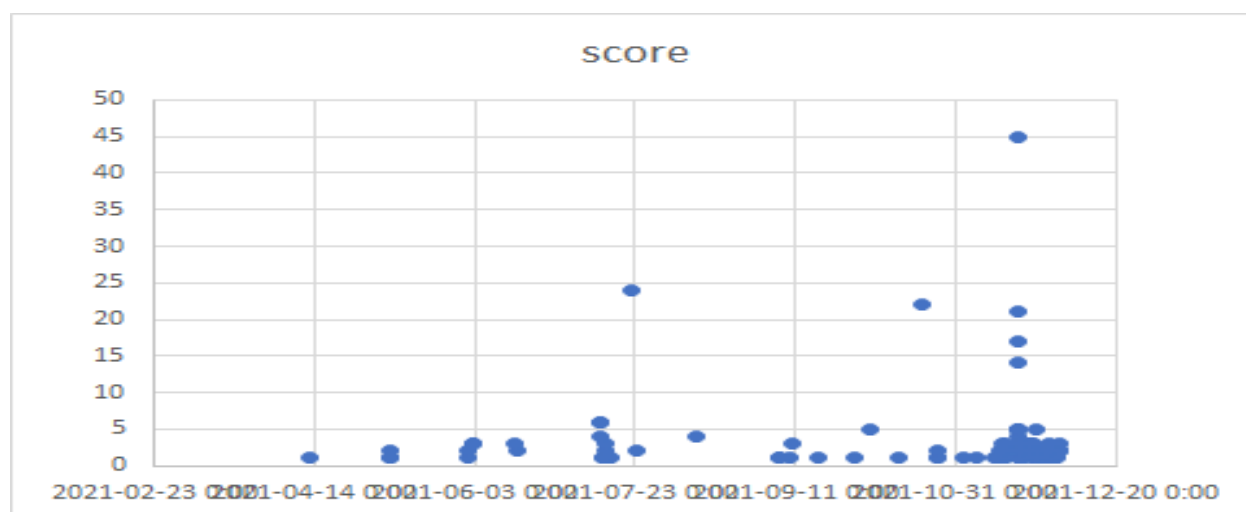
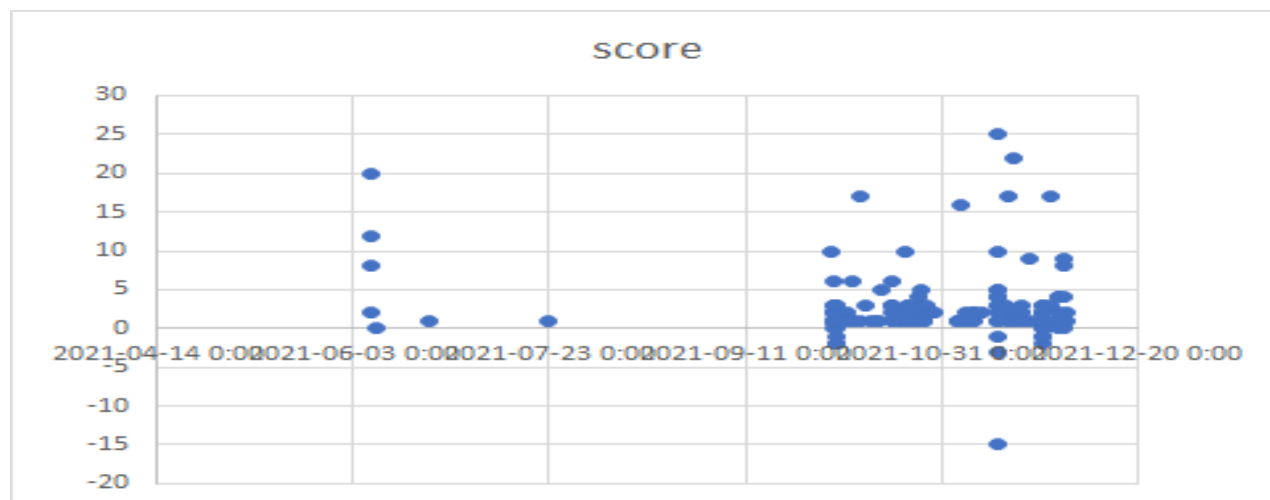
Increase



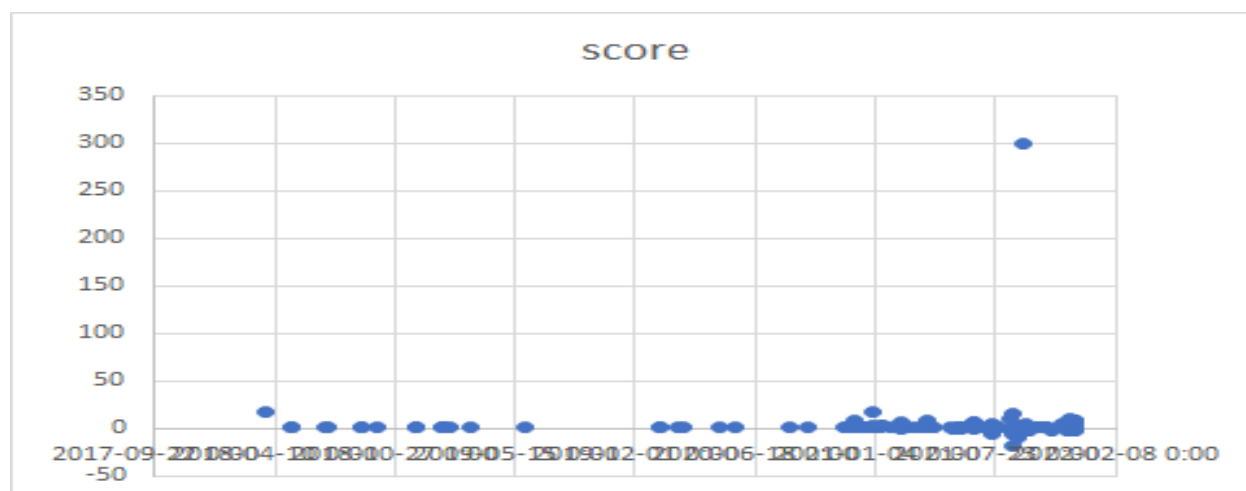


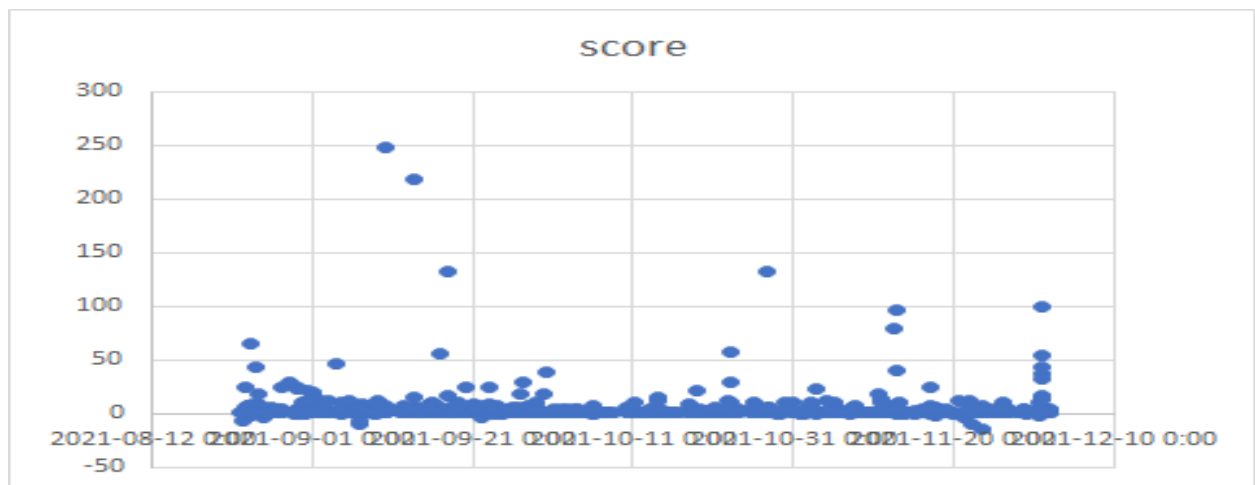
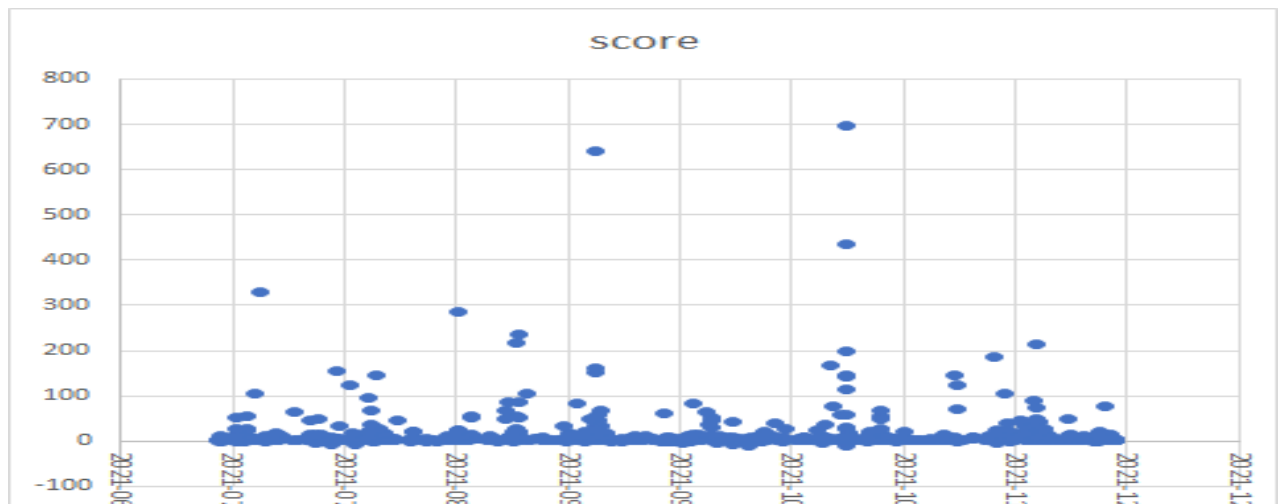
Decrease or decrease initially





No correlation with reaction





Scarce or discontinued activity

