
Depression Analysis

December 22, 2022

1. Introduction

Depression is a mental health disorder that changes the mood to feeling of sadness, hopelessness and losing interest in activities. It can further lead to other types of emotional, physical problems and distract the daily lives of the patients, such as difficulty sleeping, changes of appetite, difficulty focusing, fatigue, and thoughts of suicide [1]. This could impact patients' lives significantly, interfering with their ability to study, work, or even to maintain their social relationships.

Mental health disorders such as depression is a common condition that affects many people around the globe. According to WHO, approximately 3.8% of the world's population is diagnosed with depression. It can occur at any age, 5% of the adults, and it can be caused by various factors including genetics, life events, or medical conditions. It is a leading cause of disability and one of the main contributors to the global burden of disease [2].

There are effective medications and treatments that can improve mental health. The examples of the treatments are cognitive behavioral therapy and interpersonal psychotherapy, behavioral activation, and the medications include antidepressant medication such as selective serotonin reuptake inhibitors (SSRIs) and tricyclic antidepressants (TCAs). Although, antidepressant medications are not supposed to be prescribed as the first line of treatment, especially for mild depression symptoms.

The objective and motivation of the project is to gain valuable information from the data and get results whether the data is enough to analyze depression. The analysis will include exploration of plots with different methods such as comparison between parameters of data and time series analysis, in addition, some articles and knowledge about depression are included in this project to aid the analysis.

In the end of the studies, the results suggest that activity levels and MADRS scores can be the indicators of depression symptoms.

2. Problem Formulation

Depression could lead to suicide if left untreated. Therefore, it is essential to seek help from mental health professionals when we experience symptoms of depression. This project is motivated to analyze potential symptoms that indicate or might lead to depression.

In general, there are several symptoms that can lead to depression. The symptoms include feelings of blues(sadness, emptiness, or hopelessness), anxiety, sleep disturbances, and lack of passion in normal activities [1].

MADRS or Montgomery Åsberg Depression Rating Scale is a method to scale symptoms of depression based on 10 parameters diagnosing core symptoms of depression. The parameters include mostly patient reports of the mood symptoms such as 1. Apparent sadness, 2. Reported sadness, 3. Inner tension, 4. Reduced sleep, 5. Reduced appetite, 6. Concentration difficulties, 7. Lassitude, 8. Inability to feel, 9. Pessimistic thoughts, and 10. Suicidal thoughts[8, 9].

The scores of MADRS are then divided as:

- 0-6 = normal
- 7-19 = Mild symptoms of depression
- 20 - 34 = Moderate symptoms of depression
- >34 = Severe symptoms of depression

This research includes the amount of activities of depressed and non-depressed individuals and MADRS scores to evaluate the symptoms of depression. These factors are then applied for the plots and to determine whether they are applicable enough to indicate depression.

3. Dataset Description

The dataset utilized in this project is from kaggle, The Depression Dataset [3]. There are two separate data folders, one contains the Controls files and the other contains the Condition group files, and a scores file, that contains mostly scores of the Condition group. The Control group includes individuals without depression, whereas the Condition group is individuals with depression. All the data is in csv files that are collected over time from patients.

Condition and Control files data include the activity data (actigraph) observations of each individual during the studies, approximately 5-20 days. Each patient has their own csv file that contains a timestamp with one minute intervals, date and activity.

In more details, Scores.csv file includes background information of the participants, that is categorical and numerical. Categorical data includes the gender (1:female, 2:male), age, education, marital status, employment status, type as afftype(1:bipolar II, 2:unipolar depressive, 3:bipolar I), melanch(1:melancholia, 2: no melancholia), inpatient(1: inpatient, 2:outpatient). Whereas numerical data includes days, madsr1, which is MADRS score when the measurement started, madsr2, MADRS score when the measurement stopped, and additional data is included, that is the difference between madsr2 and madsr1 that is called DeltaMADRS (figure 2). All the data included in scores is only for the Condition group. [5]

First the data is preprocessed for the scores data. The condition and control groups are separated. Condition group is also then separated as categorical and numerical data. In addition, for the Condition group, there are no missing values for most of the attributes, except the melanch parameter. There are in particular three missing values for condition_7, condition_8, and condition_9 individuals. Thus, in the figure 1 of categorical data plots below, in plot of melanch data, these three missing values are not included.

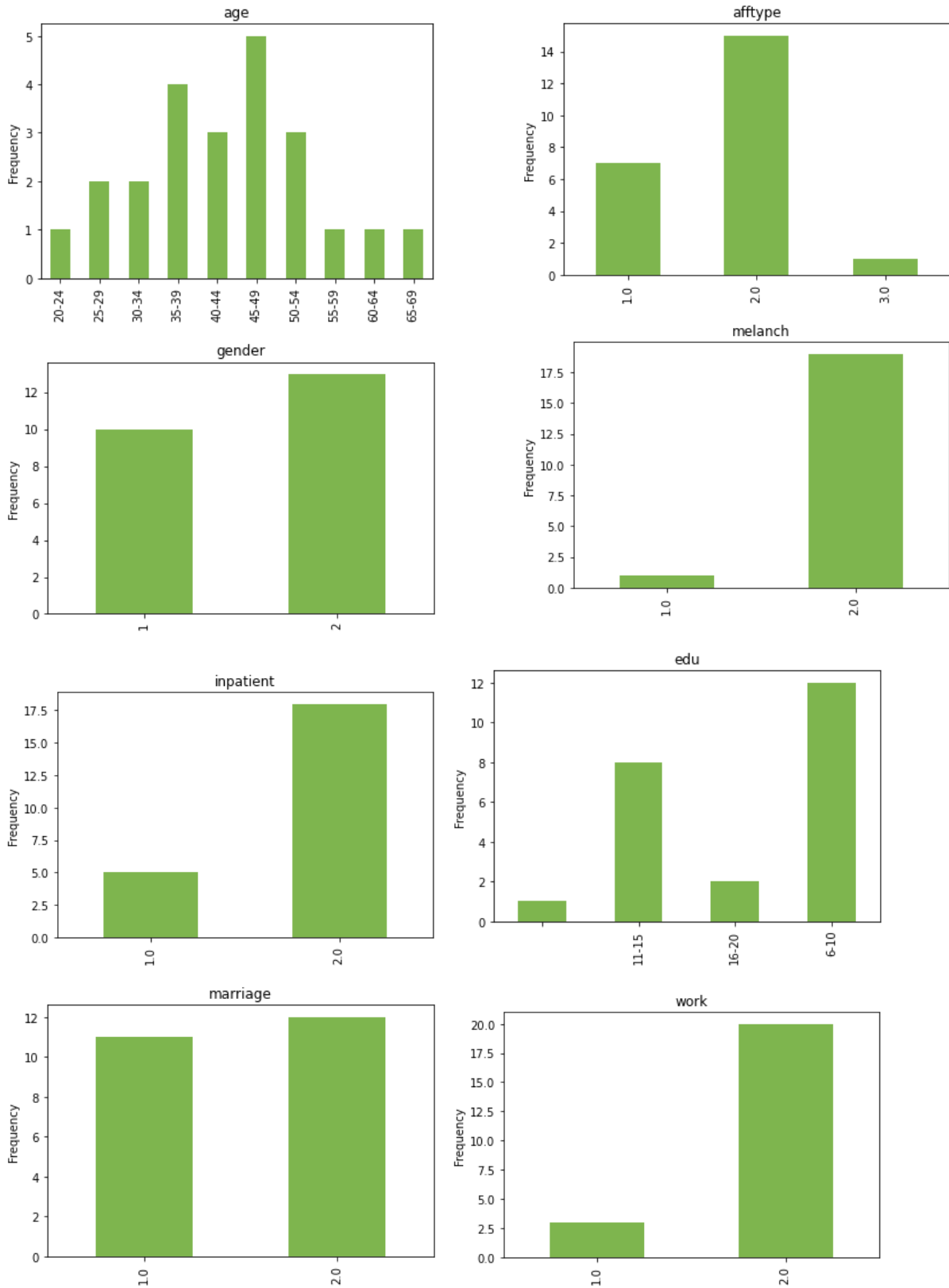


Figure 1. Categorical data of scores.csv including conditions individuals data

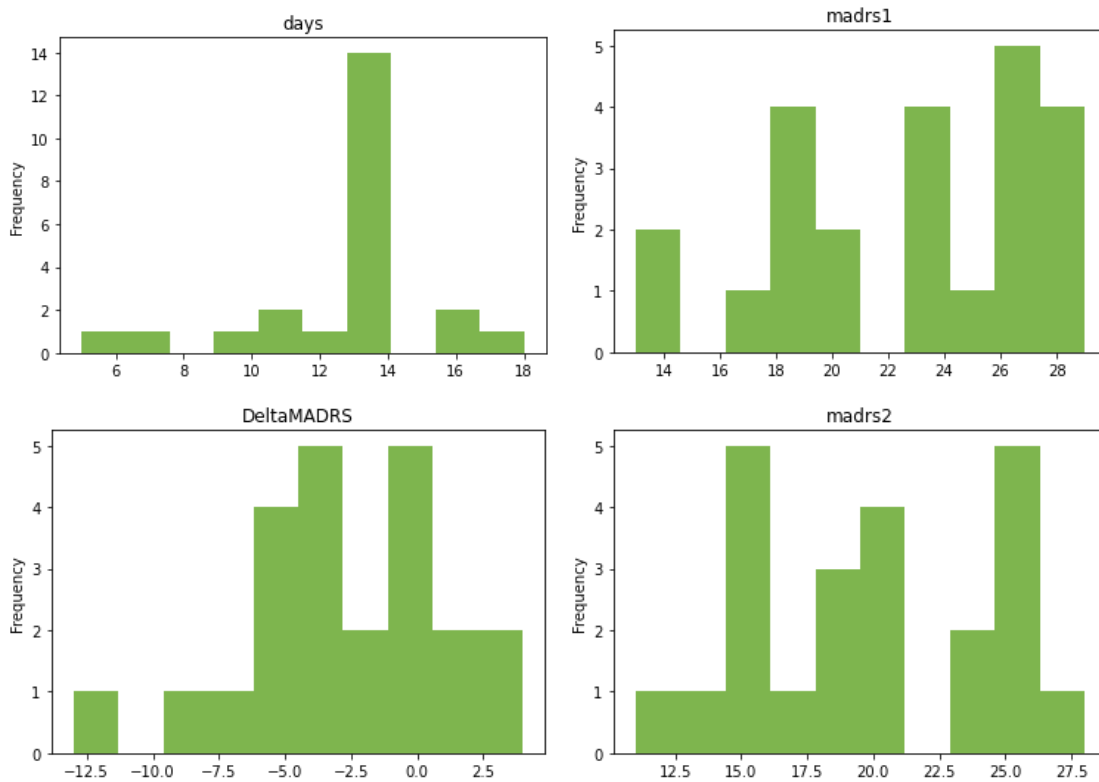
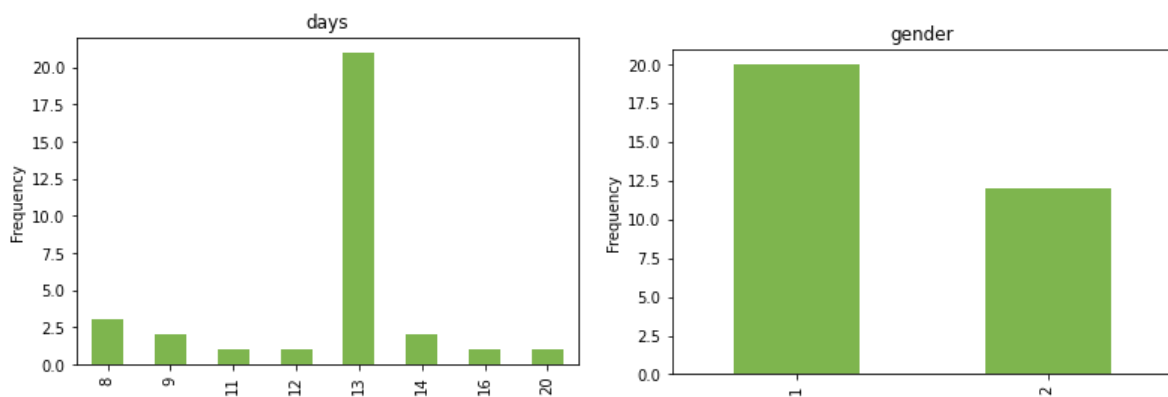


Figure 2. Numerical data of scores.csv including conditions individuals data

In scores data for the Control group, there is only basic information data available, since they are individuals without the depression condition, thus further tests are not applicable. The data available are days, gender and age that can be viewed in figure 3 below.



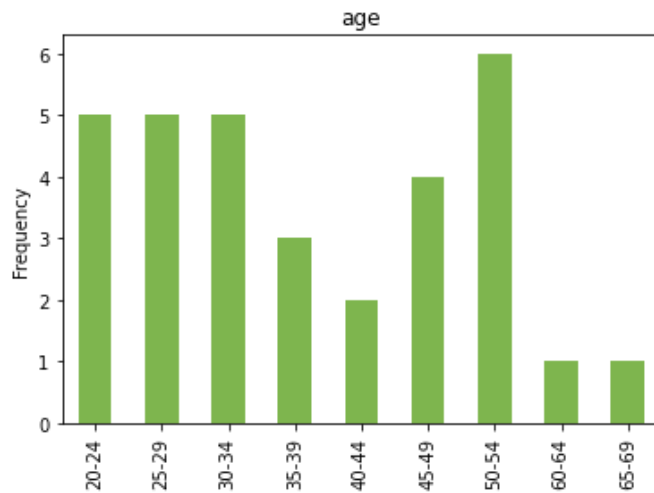


Figure 3. Available data of control individuals in scores.csv

By comparing the data that is available from scores.csv, we can analyze the Control and Condition individuals participating in the studies, according to days, gender and age. The length of the studies for both groups are pretty similar with the most having around 13 days long. The control group is dominated mostly by females, however the condition group is slightly dominated by males. Whether males are more prone to depression than females from this result is more to be analyzed in the future.

The age observation between both groups show that the Condition group or individuals with depression age is the highest of adults in their 40s.

Methods

There are several methods that are applied in this project including observations of MADRS scores, comparison between MADRS scores and data features, scatter plot comparison for each pair of numerical features in conditions data with regression line method, and finally Time Series Analysis for the Condition and Control data.

	days	madsr1	madsr2	DeltaMADRS
count	23.000000	23.000000	23.000000	23.000000
mean	12.652174	22.739130	20.000000	-2.739130
std	2.773391	4.797892	4.729021	3.968253
min	5.000000	13.000000	11.000000	-13.000000
25%	12.500000	18.500000	16.000000	-5.000000
50%	13.000000	24.000000	21.000000	-3.000000
75%	14.000000	26.000000	24.500000	0.000000
max	18.000000	29.000000	28.000000	4.000000

Figure 4. Numerical Features of Conditions in scores data descriptions

First methods are observations of MADRS scores. The difference between madsr2 and madsr1 is stated as DeltaMADRS in Figure 4 above and can also be seen in figure 6 below.

	days	madsr1	madsr2	DeltaMADRS
days	1.000000	0.122678	0.027726	-0.115285
madsr1	0.122678	1.000000	0.653090	-0.430772
madsr2	0.027726	0.653090	1.000000	0.402082
DeltaMADRS	-0.115285	-0.430772	0.402082	1.000000

Figure 5. Numerical Features of Conditions in scores data correlation matrix

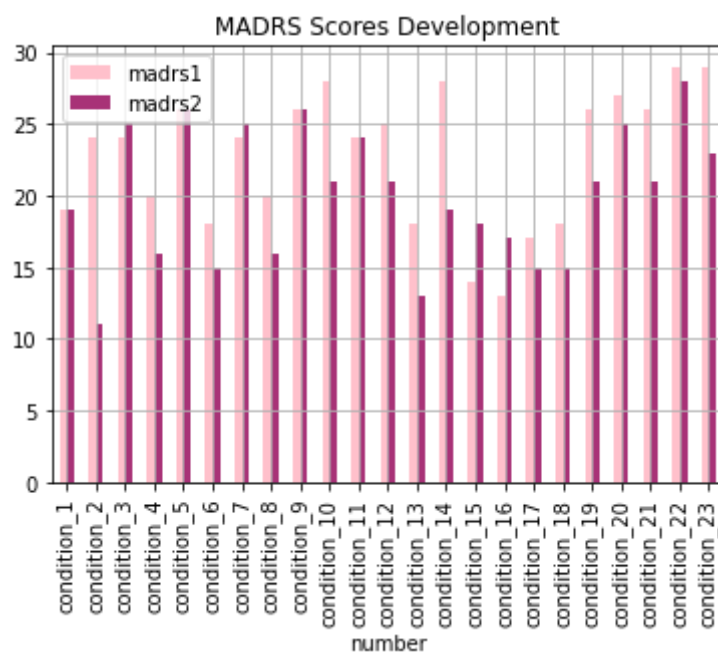


Figure 6. MADRS1 and MADRS2 Scores for all condition/depressed patients

The next visualizations are the comparison between MADRS scores and categorical data features utilizing boxplots. Figure 7 below shows the plots of DeltaMADRS compared to each of the categorical data: age, inpatient, gender, education, afftype, marriage, melancholia, and work.

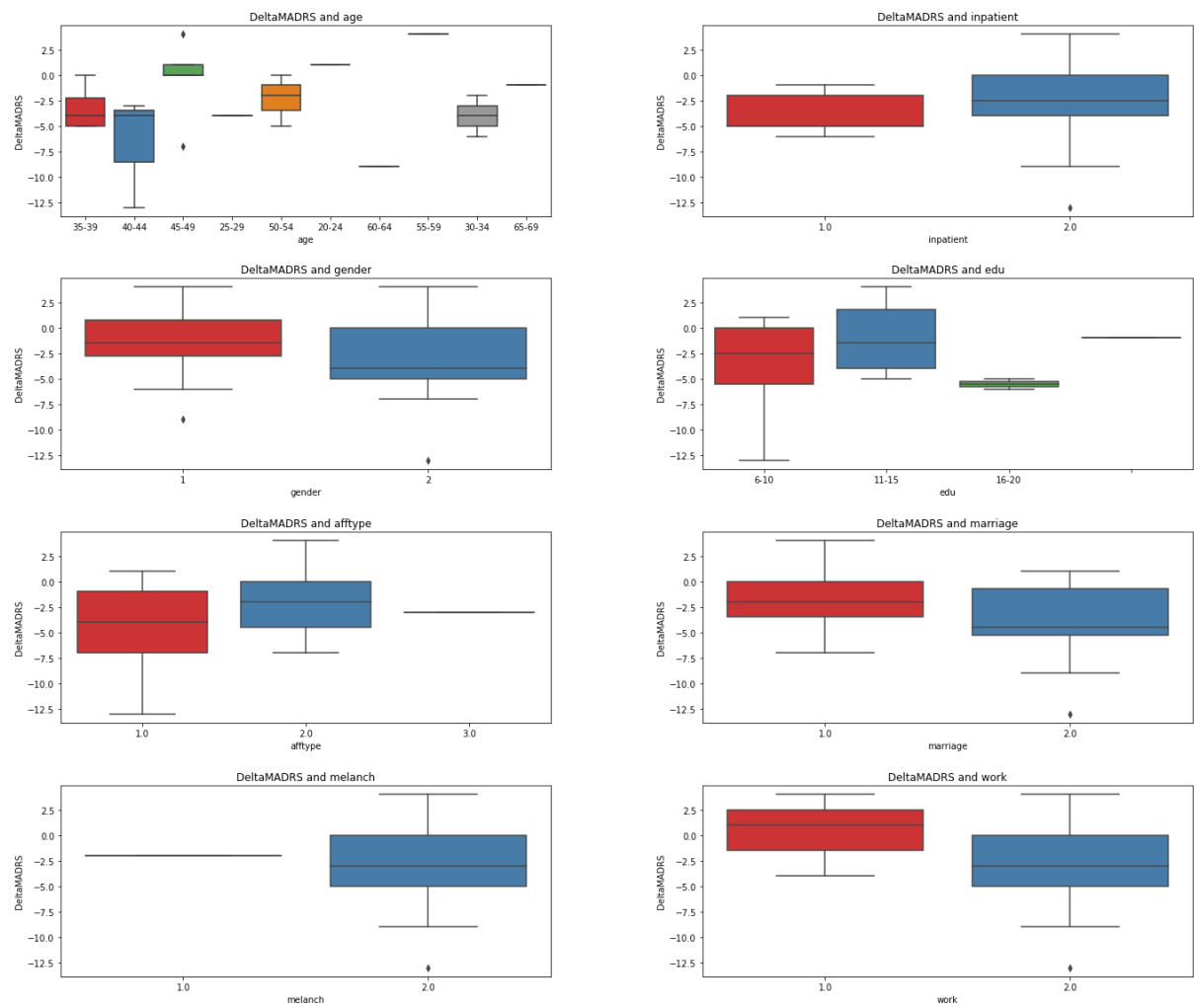


Figure 7. Comparison between DeltaMADRS scores and Categorical features of Condition Data

Figure 8 below shows the plots of MADRS1 scores compared to each of the categorical data.

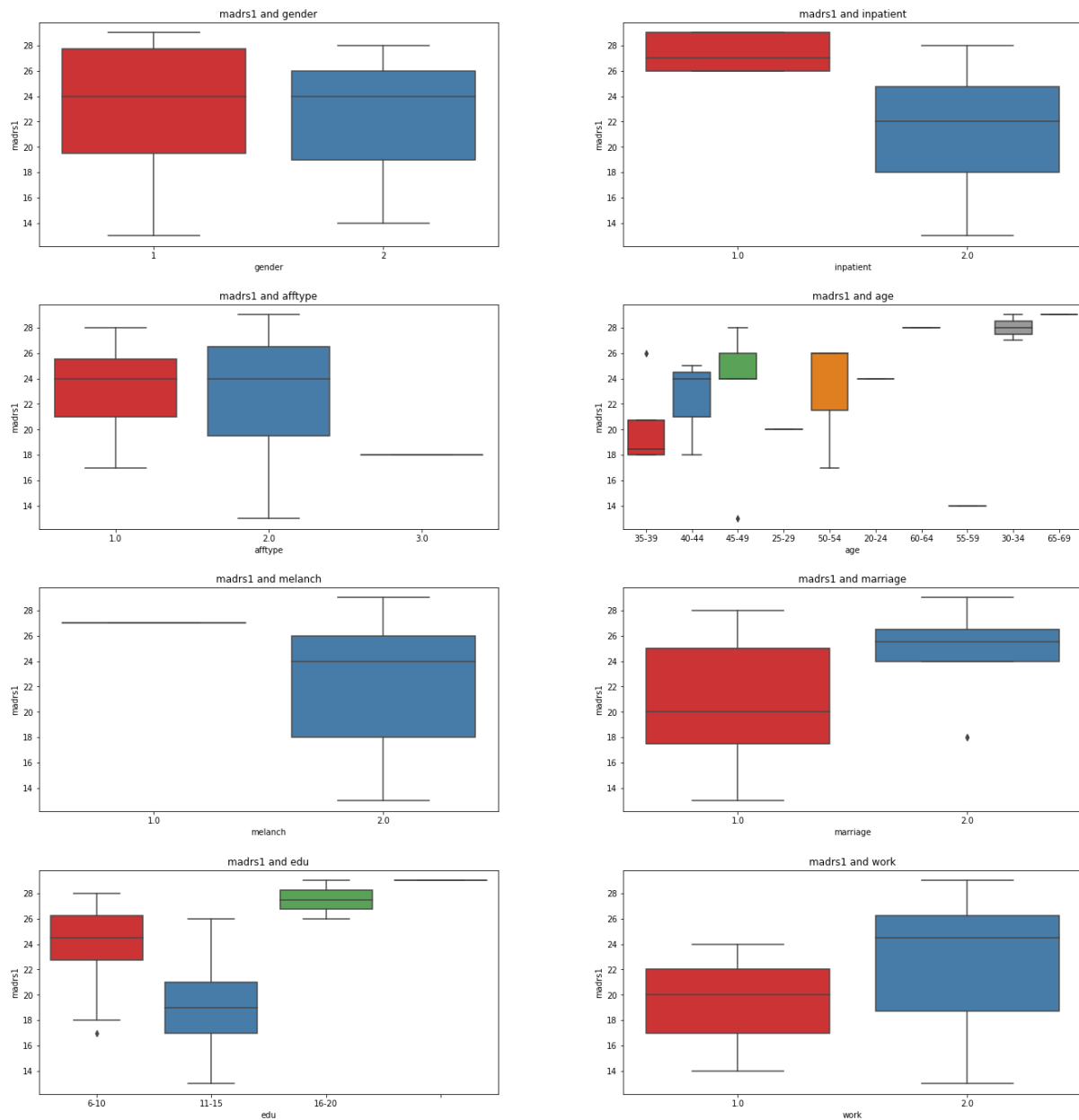


Figure 8. Comparison between MADRS1 scores and Categorical features of Condition Data

Figure 9 below shows the plots of MADRS1 scores compared to each of the categorical data.

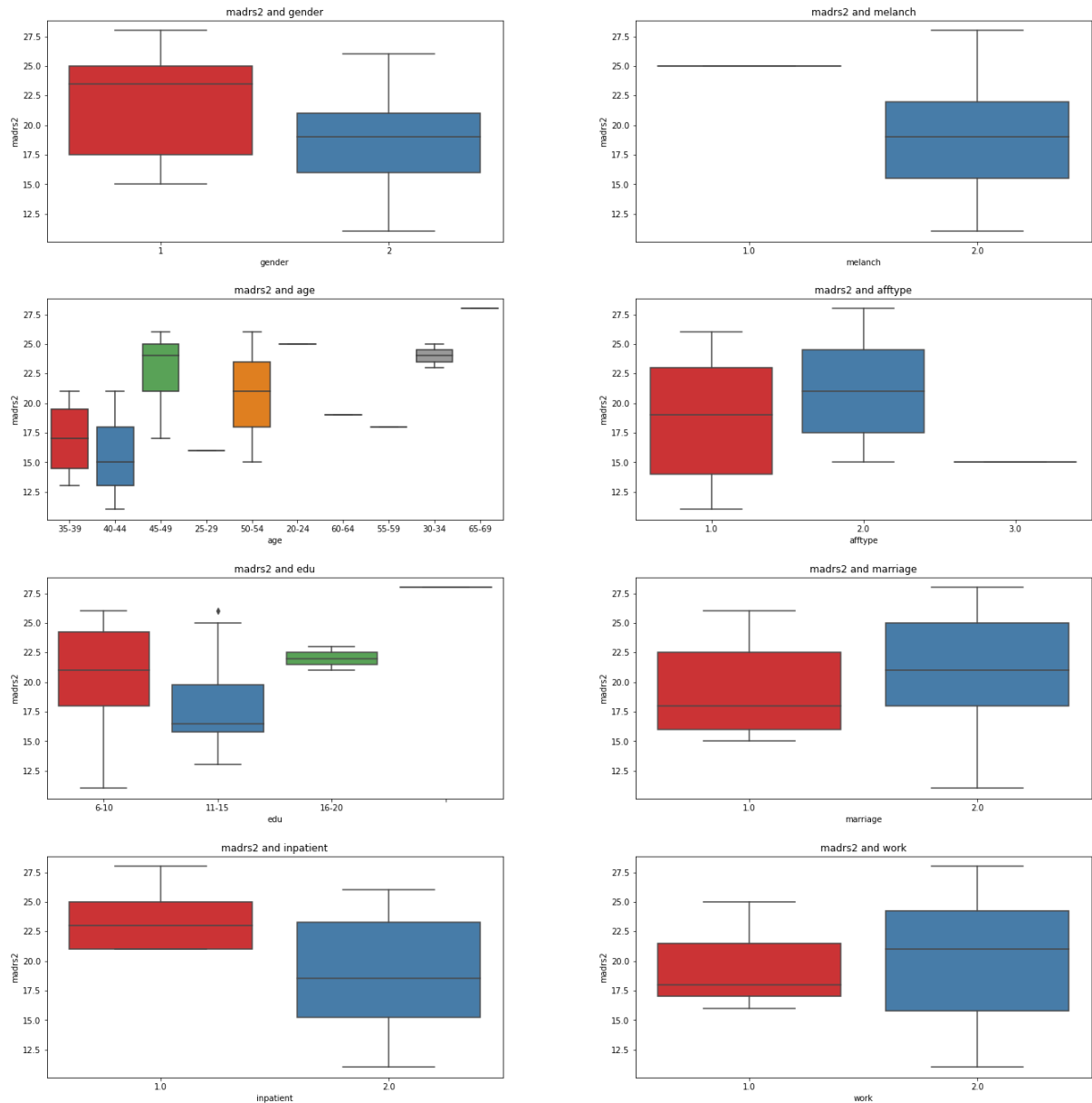


Figure 9. Comparison between MADRS2 scores and Categorical features of Condition Data

Scatter plot comparison is applied and visualized for each pair of numerical features in conditions data to understand the relationships between the features better. A regression line is also applied to the scatter plot as a way to view the data movement more precisely and a way to predict where the data heads to.

The last method of this project is the Time series analysis for the Condition and Control data of each individual. The data is preprocessed by combining all of the csv files of the activity recordings of 23 conditions/depressed individuals and 32 controls/non-depressed individuals. The data then is processed to obtain the activity mean, activity standard deviations, and the total zero activity counts.

Time series analysis is utilized here because the data is heavily related to the timestamp. Thus, this method is applicable since it is time interval data points.

Results

The overall descriptions and visualization in Figure 4 and figure 6 show that the MADRS score is overall lower at the end of the measurement. This is an interesting finding because it means that the patients do have improvements after the studies, having less severe symptoms of depression.

The following figure 10 is the scatter plot visualization. These plots show that there are positive and negative correlations between the numerical features of condition data in MADRS scores. The *madr2* and *DeltaMADRS* seem to have positive correlations whereas *madr1* and *DeltaMADRS* have negative correlations.

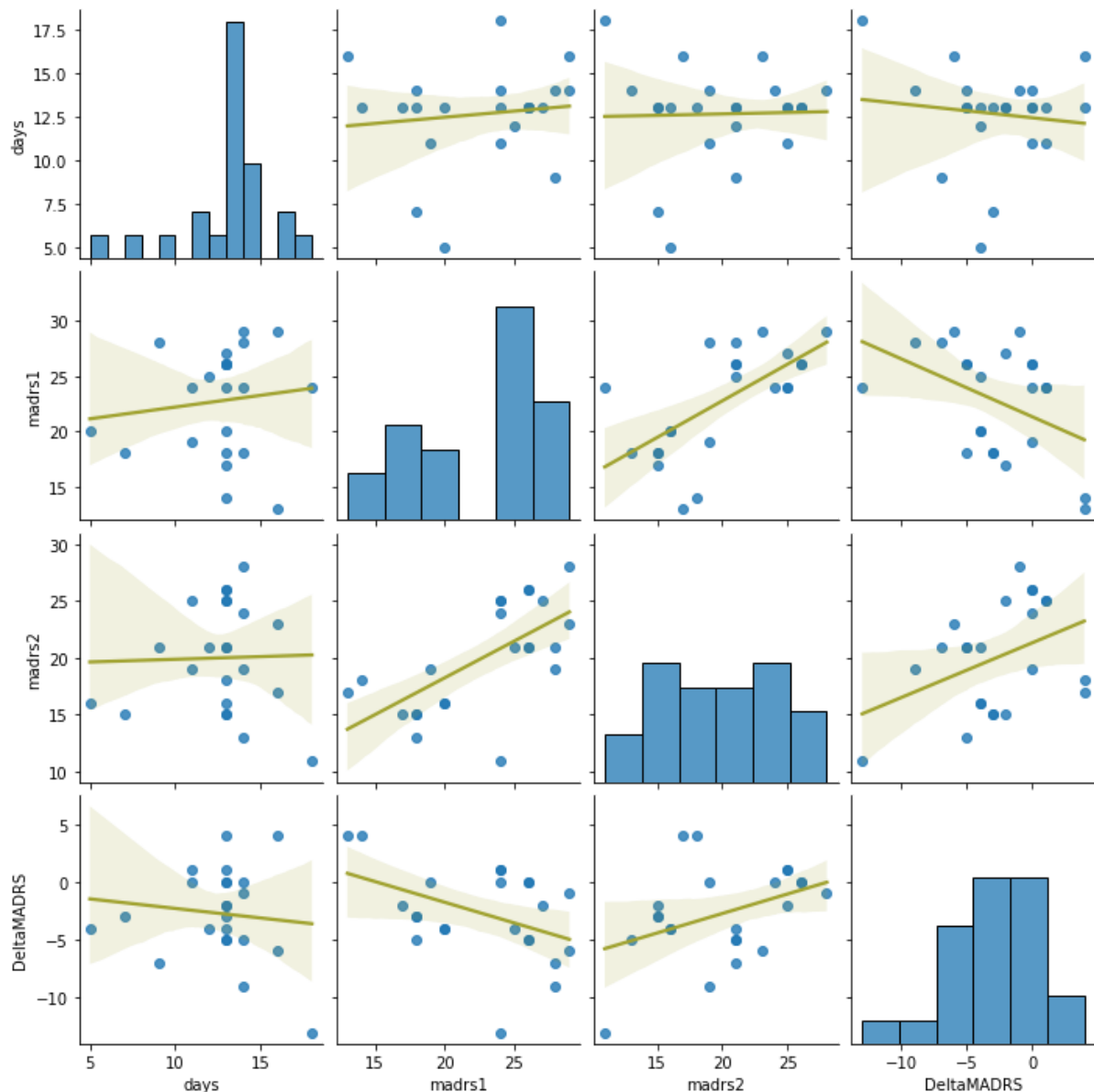


Figure 10. Scatter plot for each pair of Numerical features in Condition data with regression line

Figure 11 below shows two plots of Time Series Analysis of Zero Activity Count between depressed patient (Condition 22) and non-depressed patient (Control 9).

Depressed patient (Condition 22) seems to have generally higher zero activity levels compared to non-depressed patient (Control 9).

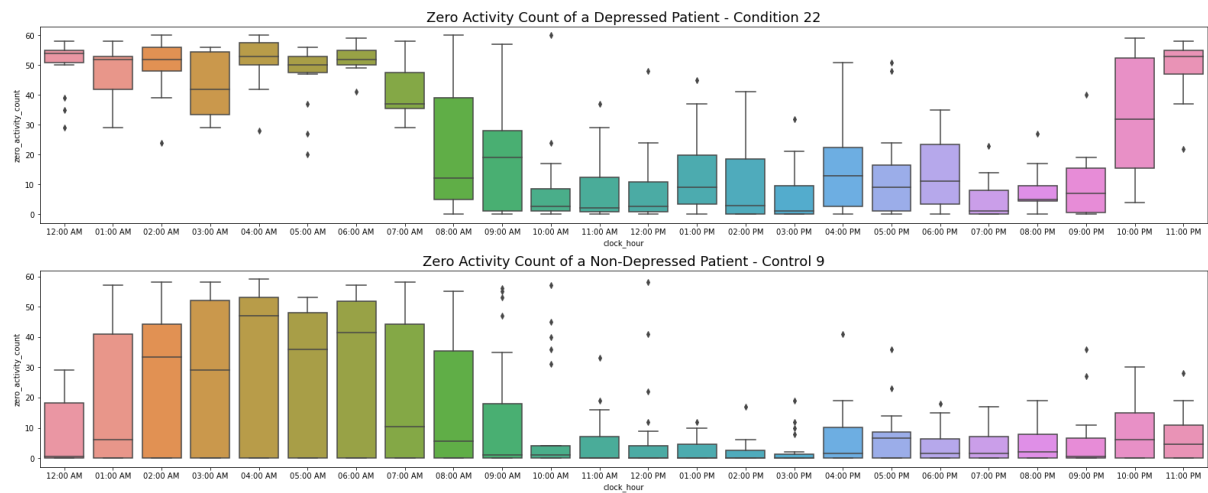


Figure 11. Time Series Analysis of Zero Activity Count between Condition and Control

Figure 12 below shows two plots of Time Series Analysis of Mean Activity Count between depressed patient (Condition 1) and non-depressed patient (Control 2). Depressed patient (Condition 1) has lower mean activity levels compared to non-depressed patient (Control 2).

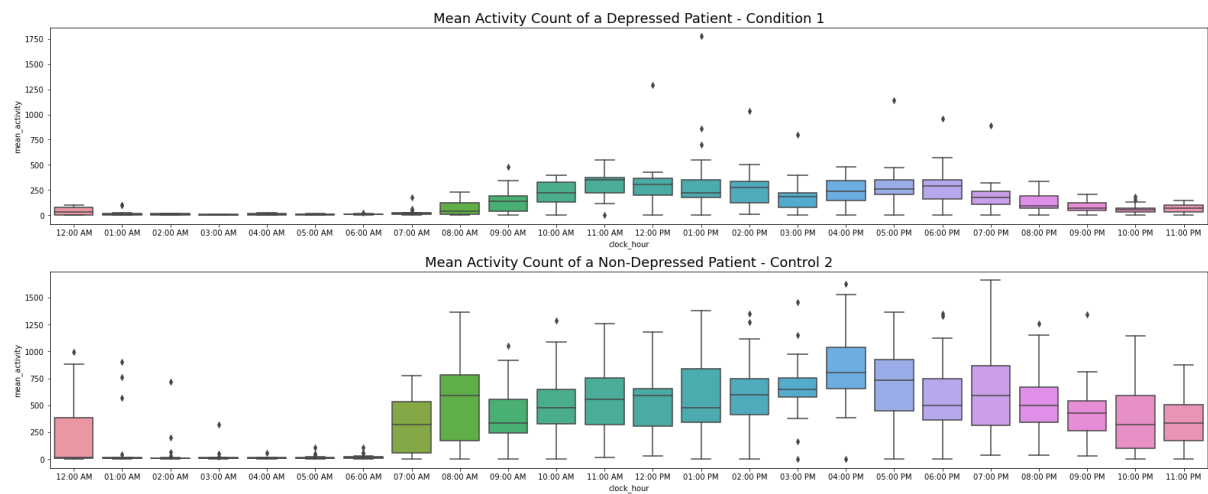


Figure 12. Time Series Analysis of Mean Activity between Condition and Control

Figure 13 below shows two plots of Time Series Analysis of Standard Deviations Activity Count between depressed patient (Condition 9) and non-depressed patient (Control 10). Depressed patient (Condition 9) has generally higher standard deviation activity count compared to non-depressed patient (Control 10)

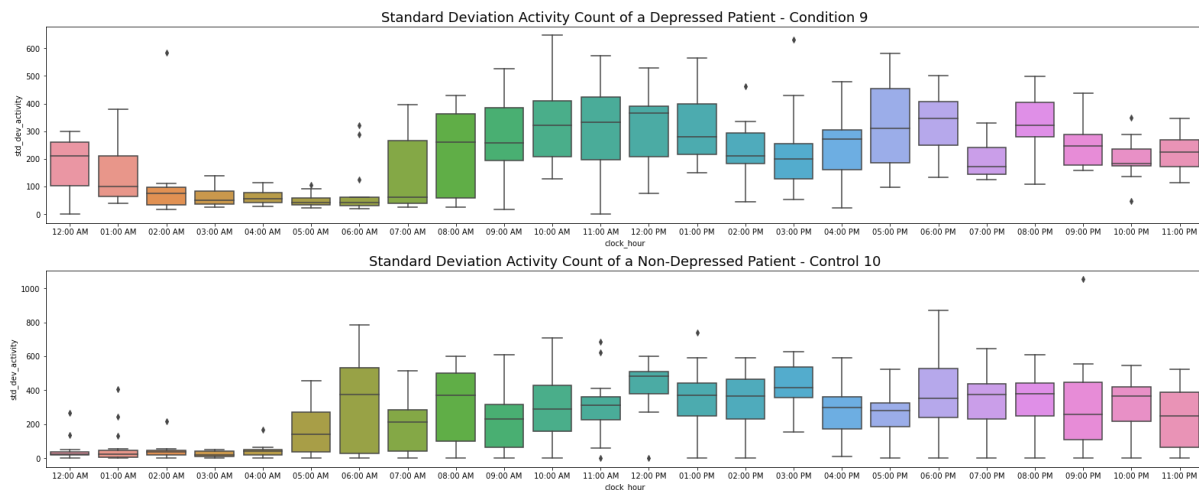


Figure 13. Time Series Analysis of Standard Deviation Activity between Condition and Control

Conclusion & Discussion

The results generally show that activity levels can be one of the indicators of depression symptoms. This however does not apply to all individuals, sometimes one that is not depressed might not have high activity levels because of various reasons, such as laziness, or disabilities. Therefore, these activity levels should be accompanied by MADRS scores and also inputs from medical professional analysis.

The overall length of the studies for both groups in total are around 13 days long. The control group is dominated mostly by females, however the condition group is slightly dominated by males. Whether males are more prone to depression than females from this result is more to be analyzed for the future analysis.

According to Mayo Clinic Staff [6], Men are half less likely to be diagnosed with depression. Depression rate is higher among women than men because women reach puberty earlier than men, this explains why women might develop depression also at an earlier age. However, according to an article by Nathan Swetlitz [7], men are most likely to suffer undiagnosed depression because the amount of men death by suicide is 3 to 4 times more than women. This might be related to how men are raised to express their emotions differently than women. Masculinities is one factor that is an obstacle for men to seek help when they feel they have mental health problems. This might also reveal that women are more willingly to participate in depression related studies than men.

There is limited time for this project, therefore more analysis in the future can be included. Different data from various sources could also be applied to the future research for better understanding and relations of the results.

References

1. *Depression (major depressive disorder)* (2022) *Mayo Clinic*. Mayo Foundation for Medical Education and Research. Available at: <https://www.mayoclinic.org/diseases-conditions/depression/symptoms-causes/syc-20356007> (Accessed: November 22, 2022).
2. *Depression* (no date) *World Health Organization*. World Health Organization. Available at: <https://www.who.int/news-room/fact-sheets/detail/depression> (Accessed: November 22, 2022).
3. Möbius (2021) *The depression dataset*, *Kaggle*. Available at: <https://www.kaggle.com/datasets/arashnic/the-depression-dataset> (Accessed: November 22, 2022).
4. *Depresjon* (no date) *Simula Datasets - Depresjon*. Available at: <https://datasets.simula.no/depresjon/#dataset-details> (Accessed: November 22, 2022).
5. *Project* (no date) *Project - CS-C4100 documentation*. Available at: <https://digitraceslab.github.io/dhnb/project.html> (Accessed: November 22, 2022).
6. *Women's increased risk of depression* (2019) *Mayo Clinic*. Mayo Foundation for Medical Education and Research. Available at: <https://www.mayoclinic.org/diseases-conditions/depression/in-depth/depression/art-20047725> (Accessed: December 22, 2022).
7. Swetlitz, N. (2021) *Depression's problem with men*, *Journal of Ethics | American Medical Association*. American Medical Association. Available at: <https://journalofethics.ama-assn.org/article/depressions-problem-men/2021-07> (Accessed: December 22, 2022).
8. *Montgomery-Åsberg Depression Rating Scale* (no date) *Montgomery-Åsberg Depression Rating Scale - an overview | ScienceDirect Topics*. Available at: <https://www.sciencedirect.com/topics/medicine-and-dentistry/montgomery-asberg-depression-rating-scale> (Accessed: December 22, 2022).
9. *Montgomery-Åsberg Depression Rating Scale* (2022) *Wikipedia*. Wikimedia Foundation. Available at: https://en.wikipedia.org/wiki/Montgomery%E2%80%93%C3%A5sberg_Depression_Rating_Scale (Accessed: December 22, 2022).