# BRSM
# Project Report

**Name:** Divyansh Tiwari
**Roll No:** 2020111002
**Team:** Psyduck

## Introduction

Music has long been acknowledged as a potent force in human existence, capable of eliciting emotions, memories, and bodily responses. But, in addition to its entertainment appeal, music has been shown to significantly influence mental health. Music's therapeutic effects are rapidly recognized and investigated by researchers and doctors alike, from lowering stress and anxiety to boosting mood and cognitive performance.

A significant amount of data has arisen in recent years to support the use of music as a supplemental treatment for a variety of mental health issues such as depression, anxiety, and post-traumatic stress disorder. Music therapy, in which trained professionals utilize music as a tool to assist individuals in achieving therapeutic goals, has also gained acceptance as a beneficial intervention for those suffering from mental health concerns.

Music may be a powerful tool for fostering mental well-being and enhancing the overall quality of life, whether it's listening to your favorite song to improve your mood, playing an instrument to relieve stress, or engaging in group music therapy sessions.

Anxiety, depression, and other mental health issues are on the rise in today's world. We hypothesize that there is a link between a person's mental health and the features of the music they hear. As a result, we want to study the association between the two using the Kaggle dataset (link is provided in the last section of this report) "Music and Mental Health" and analyze the results.

## Methodology

**A brief outline of the methodology adopted by the team:**
We did the following steps throughout the course of the project:

- First up, we explored the dataset. During this phase of our project, we studied the features comprising the dataset. A brief explanation about the dataset is: it contains features that describe the user like their age, the preferred music streaming platform, whether or not they listened to music while working or not, etc. Then there were features that described the musical preferences of the user like their favorite genre, the BPM of the music that they listen to, and the frequency with which they listen to the specific genres of music.

Later, there were features that explained the mental health conditions of the users like their anxiety, depression, etc through scores and later they had a feature describing whether or not they benefitted from listening to the music or not.
- We, then, performed the data preprocessing steps which included the handling of Nan values and converting the ordinal/categorical variables to numeric variables.
- Now, we proceeded onto visualizing the dataset so as to get a better understanding of the features and the data spread which could help us to choose the features for performing various statistical tests.
- We then proceeded to apply some statistical tests to the data and built and tweaked some models around it to get an even better understanding of the outcome of the project.

**The analysis that was done by me:**

Our project was divided into three phases: Data exploration and visualization, statistical tests, and building models. We, as teammates, collaborated with each other in all three phases. However, we can draw boundaries as per the contribution to the project. During the first phase of the project, the lead was taken by me. Subsequently, during the later phases of the project, I contributed my fair share with certain statistical tests and exploratory models. The analysis that has been done by me is explained in detail below.

**Data preprocessing:**

In a research study, the data we collect and the data that is helpful in interpretation are far apart from each other. There are certain pre-processing steps required by our project and were carried out by me which are explained below:

- The dataset that we took from Kaggle had features that were both, numerical and categorical. I converted the categorical data into numerical data in order to facilitate the execution of the test that specifically requires the data to be of numerical nature. This was mainly used in ANOVA testing.
- The dataset also had missing values in a variety of features like the BPM, frequencies, etc. Such missing values were handled by the method of grouping. The missing values that occurred, for example, in the BPM feature, were grouped by their favorite genre of music and the mean for that group of data was used to fill up the missing values. Also, the missing values in categorical/ordinal data were handled using grouping (where ever possible) and mode of the data.
- The features that were not useful for the project like timestamps were dropped.
- Were had a large number of features in our dataset and dealing with such a large number of features increased the complexity of the project and execution time which is why we used Principle Component Analysis to reduce the dimensionality of the data. During the

project, we used different numbers of principal components to check for the best output, and the final values, against which, the best output was received were frozen.
- During the execution of certain models like Decision Trees, we re-tweaked the data to convert the mental health scores into a categorical variable with values as 'Low', 'Medium', and 'High'.

**Data Visualisation:**

Throughout the course of the project, we had phases in which we were exploring the dataset, applying statistical tests, and building models. In all of these phases, the visualizations which helped us look at the outcome were made. The visualization techniques that were used were:

- Correlation matrices: We used the correlation matrices to understand the correlation between the variables. This also helped us during the dimensionality reduction step using PCA to group the features that were highly correlated separately and use their reduced principal components to perform the analysis.
- Histograms: Histograms were used to see the spread of data across the dataset. We made histograms to study the spreads of the data in accordance with the features like BPM, Age, and Hours Per Day (average time spent in listening to music).
- Box plots: We used box plots for a variety of reasons. First up, we used them during the preliminary data analysis to handle the outliers. During the exploratory data analysis and the statistical testing phase, we made box plots to study the relationship between the features such as Age vs Primary Streaming Service, Hours spent listening against the genre of music, Age vs Genre, etc.
- Pie charts: We used pie charts in order to study the spread of the population as per their music streaming platform preferences. We got interesting results around it. We made some plots revealing the spread of the population exploring their musical background. Also, we made plots that revealed their music-related preferences such as whether they listened to music while working or not. We looked at the spread of the population against their favorite genre of music. We made a pie chart to see how the population benefitted from listening to music in different conditions like listening while working.
- Cat plots: We used cat plots to study the relationship between variables like the genre of music and the BPM. Cat-plot provides a more intuitive way to look at the data spread and hence we used it for our study.
- Bar plot: We extensively used bar plots to study the spread of data when looked at through the mental health condition and their favorite genre of music. We also studied the spread of data of people with different levels of severity of mental health conditions against different mental health conditions and the average number of hours of listening to music. We defined a general mental health score using the average of mental health

scores for different mental health conditions. A multi-bar plot was made to look at the spread of the population's mental health score against their favorite genre.

**Statistical Tests:**

The statistical test that I did towards the various hypothesis that we made for this project were:

- T-test:
    - One potential approach to examining the relationship between music and mental health is to conduct t-tests comparing various mental health variables (such as anxiety, depression, insomnia, and OCD) with a variable measuring the impact of music on mental health. The objective of this analysis is to determine if there is a significant difference in the mean scores for mental health between those who believe that music has an impact on their mental health and those who do not.
    - The null hypothesis posits that there is no significant difference in mean mental health scores between individuals who report that music affects their mental health and those who do not. Conversely, the alternative hypothesis suggests that there is indeed a significant difference in the mean scores for mental health between these two groups.
    - By exploring the potential relationship between music and mental health through statistical analysis, we may be able to gain insight into the therapeutic potential of music as a complementary treatment for various mental health conditions. This type of research may also help to inform the development of music-based interventions for mental health, which could improve the lives of individuals struggling with mental health issues.
- Correlation test:
    - We performed correlation tests on the variables to study how the data was correlated in our project study.
- Cronbach-alpha:
    - We used this to test the internal consistency of our dataset.
- Kruskal-Wallis Test:
    - Since our data comprised of features that were both, categorical, and numerical, we performed the Kruskal-Wallis test on them as the numerical data was supposed to be non-normally distributed, in order to study the correlation of the features.

**Machine Learning Models:**

In our project study, we built different models so as to explore which one would help us in the direction of our approach better. One such exploratory model, KNN, was built for me. Also, I collaborated with the Logistic Regression models which were one of the essential models used in our study.

**The analysis that was done by my teammates:**
The other statistical analysis that was carried out was ANOVA testing and the other models that were used in the project were Logistic regression and Decision Tree Classifier.

## Results

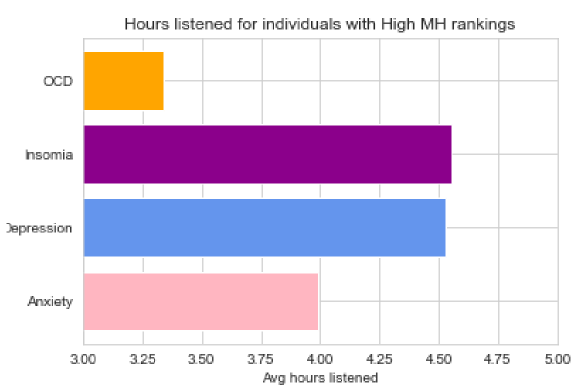
**Results from my part:**
The following are the results from my part and their interpretations:

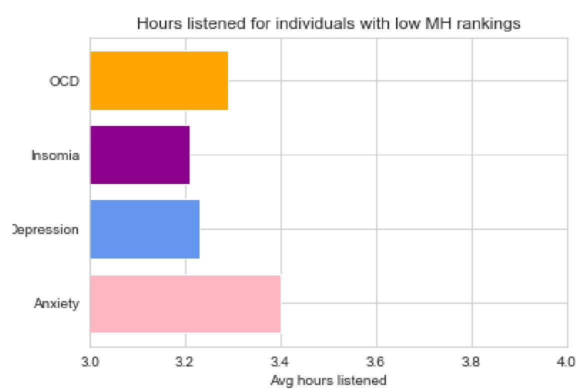
- From the Kruskal-Wallis test, we can see that the p-value is 0.237 and this means that there is no significant difference between the mental health scores across different favorite genres. We used this test, as the mental health score is a non-normally distributed variable and the favorite genre is a categorical variable.
- From the t-test:
    - By conducting t-tests and analyzing the p-values, we can gain insight into the potential impact of different genres of music on mental health variables such as anxiety, OCD, depression, and insomnia. Our analysis indicates that for anxiety and OCD, both classical and hip-hop genres do not have a statistically significant difference in mean mental health scores between those who listen to them and those who don't. This suggests that these genres may not have a significant impact on anxiety and OCD.
    - For depression, however, there is a statistically significant difference in mean mental health scores between those who listen to hip-hop music and those who do not. This suggests that hip-hop music may have a significant effect on depression. Similarly, for insomnia, there is a statistically significant difference in mean mental health scores between those who listen to classical music and those who do not, indicating that classical music may have a significant impact on insomnia.
    - It is worth noting that for all mental health conditions, except for depression, the p-values for both classical and hip-hop genres are greater than 0.05, suggesting that the effect of music genres on mental health may not be significant for most conditions. However, it is important to keep in mind that this is just a preliminary analysis, and further research is necessary to draw conclusive inferences.
    - Overall, our analysis provides some insight into the potential therapeutic benefits of different genres of music for specific mental health conditions. However, more

research is needed to fully understand the relationship between music and mental health.

- KNN model:
    - During the exploratory models, the accuracy came out to be 54.2%, which was not decent and hence we rejected it as the best model for Music and Mental Health classification task.
- Logistic Regression:
    - The logistic regression model was the best-performing model when it came to the Music and Mental Health regression task.
    - This was because the output corresponding to the model was in terms of the mental health scores and the regression task better suits the scenario as the aim is to predict the mental health status of the individual as close to their real condition as possible and not just putting the mental health condition into categories like ‘poor’, ‘average’, and ‘good’ as in decision tree classifiers.

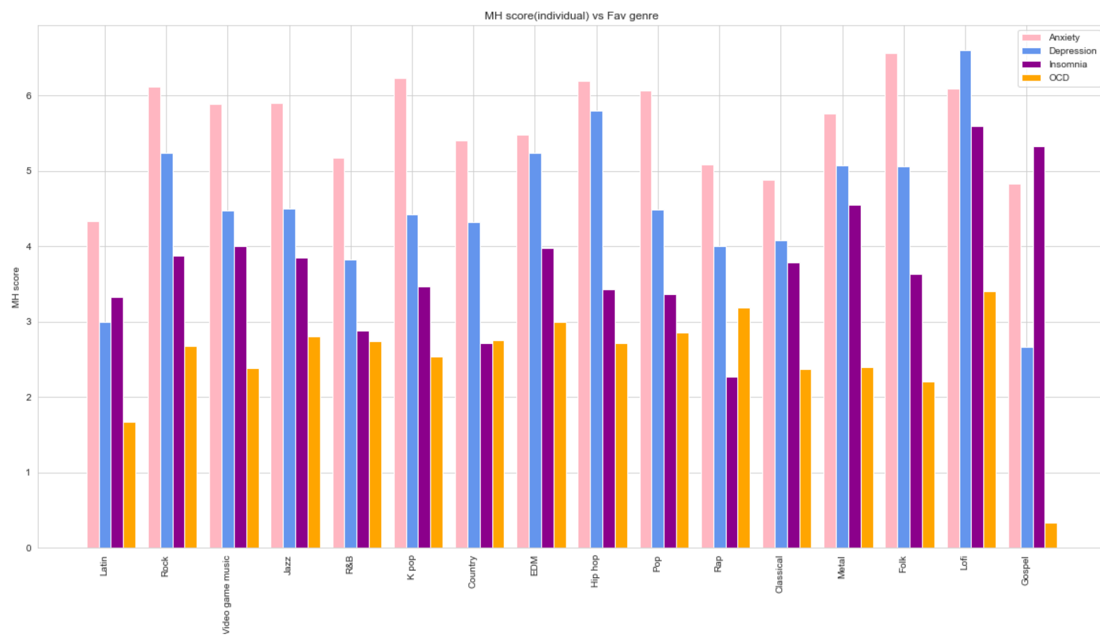
Now, the following are some inferences from the visualization plots:



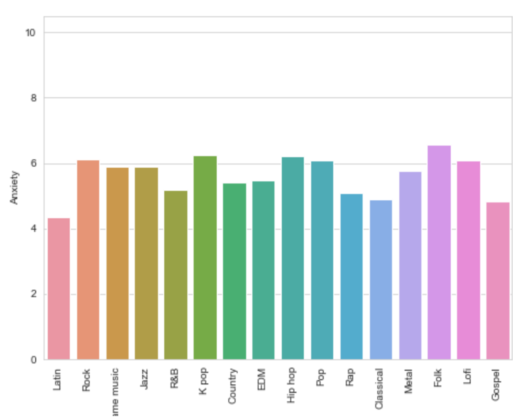



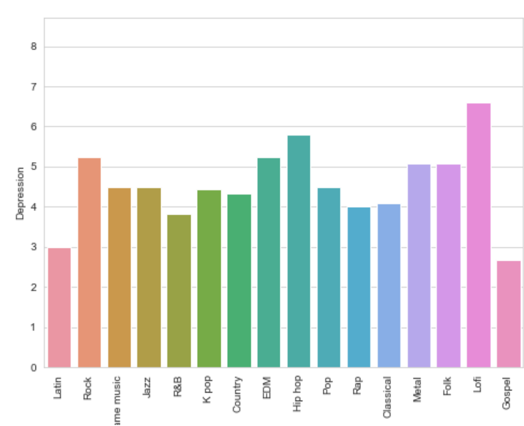

A low MH ranking is defined as a score of 0 or 1. Individuals with modest insomnia showed a significant decrease in average hours of listening when compared to their extreme counterparts. This was equally true for those who had mild levels of depression.

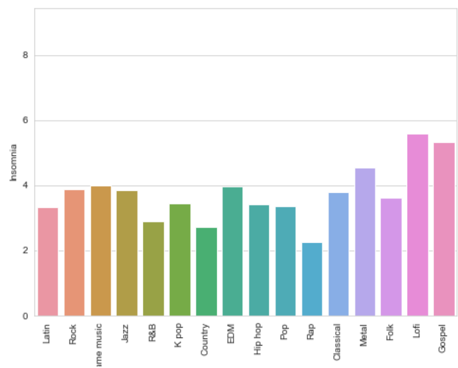
An MH ranking of 9 or 10 is considered severe. Extreme insomniacs had somewhat more listening time than people with extreme ranks in other categories. People with severe OCD have a somewhat shorter listening time. These distinctions, however, may be regarded as insignificant.

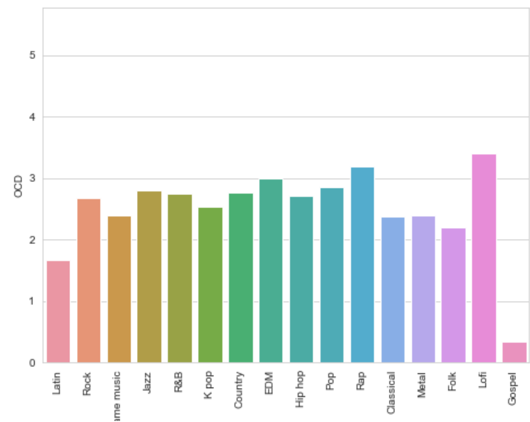



The graph above tells us that the Insomnia and Depression rates were the highest in the people whose favorite genre of music was LoFi.






Elevated depression and anxiety scores are quite prevalent, with average values of 5 and 6, respectively. Popularity appears to trend in the same way between each ranking (i.e., from rankings 1 to 2, the popularity of these rankings increases for both anxiety and depression). Interestingly (and concerningly), a given person is more likely to rate depression as a 10 than a 0.

Outside of the 0 rankings, insomnia is more prevalent and more evenly distributed. Insomnia rankings, on the other hand, show a declining tendency in popularity as the ranks rise. OCD is the least prevalent disorder, with 0 as its mode. Similarly to the insomnia rankings, as OCD ranks rise, so does its popularity.

Rock is the most popular genre, being selected as an individual's favorite genre 1 out of 4 times. In terms of popularity, rock is followed by pop and metal. Lofi, Gospel, and Latin music were selected less than 0.14% of the time and do not appear in the above pi chart.

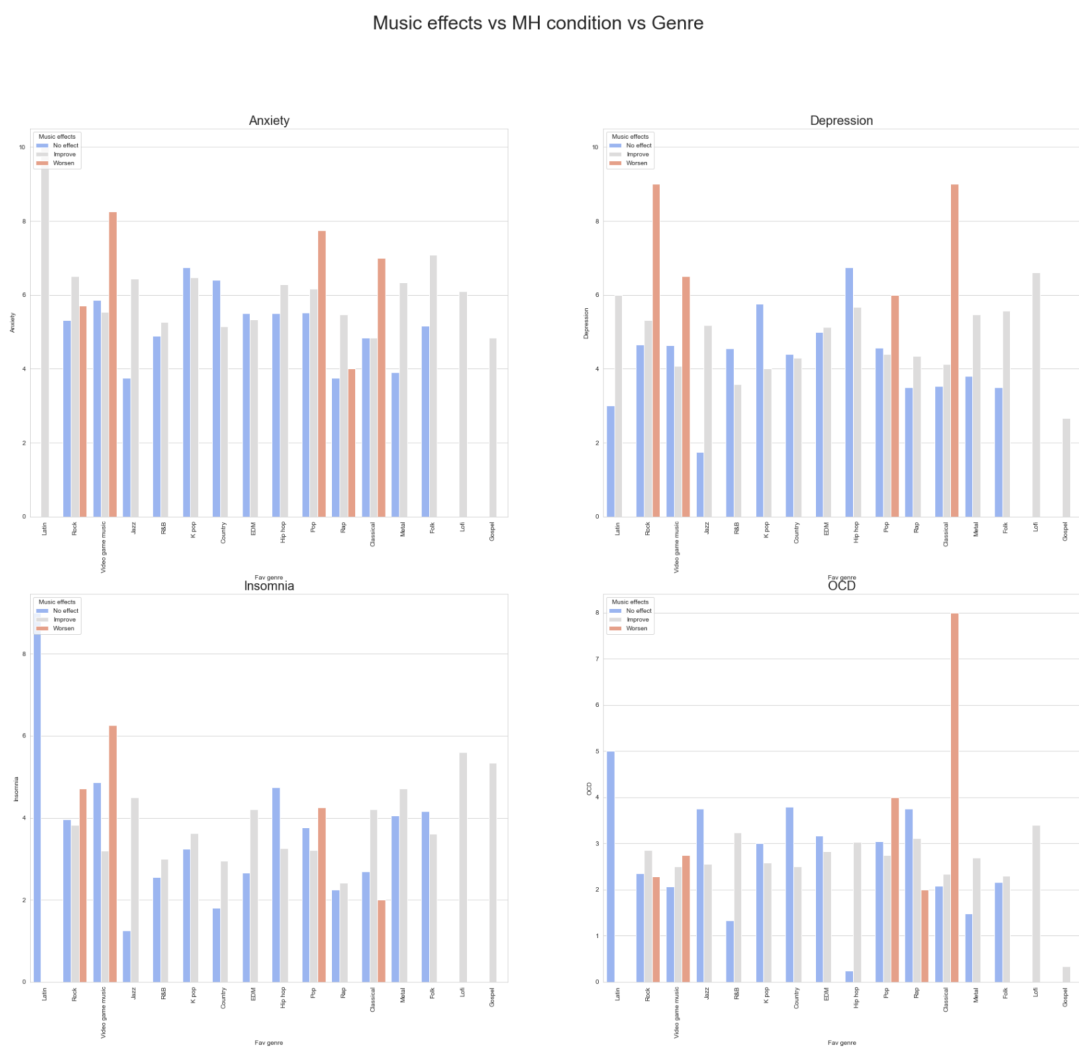
For a genre breakdown by listening frequency, please see the "(In Depth) Genres by Popularity" figure.

Music, regardless of genre, has a good influence on mental health, according to the great majority of respondents. The "Effects of Music on Mental Health" figure supports this.

Individuals who chose Gospel and Lofi as their favourite genres all agree that music is beneficial.

Individuals that chose Video Game Music as their favourite genre had the most varied answer distribution. Approximately 40% of these respondents did not believe music to be useful at all, with 10% expressing a negative influence on their mental health.

The only other respondents who deemed music harmful were those who chose Classical, Pop, Rap, or Rock music as their favourite genre.



It is important to pay attention to the most popular music genres, as illustrated in the "Top genre breakdown" figure. It is reasonable to expect that genres like Rock, Pop, and Metal will have higher value counts.

While Latin and Gospel music are not popular favorite genres, they are included in the second list, which suggests that they may have a significant correlation with lower mental health scores. This is partially supported by the m_vs_gfreq plots generated by the function. However, since there are fewer Latin and Gospel music lovers, these results may be more susceptible to noise.

Country and Rap genres also appear more frequently among those with lower mental health scores. In contrast, EDM appears slightly more frequently among those with higher mental health scores, which is consistent with the notion that listening to EDM more frequently is associated with higher mental health scores in all categories.

It is also interesting to note that Lofi does not appear in the second list, despite appearing three times in the first list. This may be explained by the fact that Lofi listeners tend to suffer more from insomnia, as shown in the "Relation between Insomnia & Genre Frequency" figure.

The other plots can be seen in the GitHub link that we have attached.

**Other results:**

Various statistical methods were employed in the analysis, including t-tests, Chronbach alpha, ANOVA, and Kruskal Wallis tests. However, their numerical results are not presented here due to space limitations. In addition to logistic regression, a decision tree classifier was also used to build a prediction model based on the data. The accuracy of the decision tree classifier was determined to be 57%.

## Conclusion

Drawing conclusions about mental health status from the data is not a straightforward task. Anxiety and depression appear to be more prevalent than OCD and insomnia, and there is a correlation between anxiety and depression, which is to be expected.

As for the effect of music on mental health, most listeners believe that music has a positive impact on their mental health, although there are some who believe that rock or video game music can have a negative effect. However, there does not seem to be a significant relationship between mental health and either the frequency of listening to music or preferred music genres. The correlation matrix analysis does not provide sufficient information to determine which specific music features contribute significantly to mental health improvement, as it is primarily based on self-evaluations of the listeners themselves.

The following is the github: https://github.com/DivyanshTiwari23/BRSM_Project

The following is the link to the dataset:
https://www.kaggle.com/code/yannansu/music-and-mental-health-eda