

Behavioral Research: Statistical Methods

Project Proposal

Music and Mental Health

Team Name: Psyduck

Team Members:

- Aryamaan Basu Roy (2020101128)
- Divyansh Tiwari (2020111002)
- Pratyush Mohanty (2020101005)

Initial Topic Selection:

The topic that has been selected is “Music and Mental Health,” and the dataset for it has been selected from Kaggle. The current scenario is witnessing the rise of mental health problems such as anxiety, depression, etc. The same increase has been observed in terms of participation on music streaming platforms such as Spotify, YouTube Music, and others. As a result, we intend to investigate the relationship between the two elephants in the room and conduct an analysis of the findings.

Brief Description of what we intend to study:

We intend to look at the relationship between various aspects of music and their impact on or correlation with individuals' mental health status throughout the course of this study.

We believe that a research study on the relationship between music and mental health could explore a range of different questions, including:

- What are the specific mental health benefits of listening to or playing music, and how do these benefits vary depending on the type of music, the individual's preferences, and mental health history?
- Can music be a therapeutic tool to help individuals manage mental health conditions such as depression, anxiety, etc? If so, what genre of music could help with the same?
- How does the frequency of listening to music impact the mental health of an individual?
- Does listening to music have an impact on the productivity of an individual, or does performing these two actions simultaneously affect their mental health?
- How do the genre of music and the music speed (in terms of BPM) affect the mental health of the individual?

Above mentioned are some of the questions that we intend to study through this project. We might also look for patterns that do not fall in the scope of these questions to get any exciting results between the subjects.

How do you plan to collect data?

We intend to use [music and mental health survey results](#) that are available in the public domain. The dataset would be derived from Kaggle, which is an online community platform for providing useful datasets for machine learning and data science projects.

With its GPU-integrated notebooks, we would perform a substantial amount of pre-processing on the data. The dataset contains many potentially useful features regarding the demographics (like age, language, time, etc.) of the participants in relation to the type of music they listen to.

Our dataset contains a large and significant amount of features that would enable us to perform meaningful analysis. The dataset contains 736 inputs of participants along with 33 features collected without any bias in selection. Thus, we can claim that we have an unbiased dataset.

For each entry, the dataset contains 33 features out of which some of the most important are age, hours per day, whether the music is heard by the person while working, favorite genre, Beats per Minute (BPM), and behavioral features like anxiety, depression, insomnia, OCD, etc.

How the analysis will be performed?

The following are the analysis steps that will be performed:

1. Data Preprocessing:

- a. In order to perform data mining, we need to clean, transform, and integrate the data to make it ready for analysis. We intend to improve the data quality and make it more suitable for the specific behavioral questions we intend to answer.
- b. The first preprocessing step that we will take is “dimensionality reduction,” especially on the 16 features defining the frequency of the music piece. There are frequency columns containing the frequency of different formats like EDM, metal, country, folk, Hip-Hop, etc. We hypothesize that the frequency of these formats in the music piece might correlate with the behavioral features.
- c. We would perform correlation analysis on the 16 features to find which features are similar. We can derive which features to merge given that they are similar enough to point in the same direction.
- d. For the 16 frequency features, we would use the SVD Transformation to perform dimensionality reduction on those features that have a high correlation coefficient.

- e. The above is a technique for portraying a matrix as a sequence of linear approximations that reveal the matrix's underlying meaning structure. The purpose of SVD is to determine the best set of criteria for predicting the outcome.
 - f. The next step would be to convert the categorical values into numerical values of 0,1,2 and 3. For example, the “Foreign Language” feature contains Yes/No. We would convert them into 0 and 1. The frequency feature contains categorical values of never, rarely, sometimes, frequently, and very frequently which will be converted to 0, 1, 2, 3, and 4 respectively.
 - g. The next step would be to normalize the data, now that we have converted the data into numerical values. Data normalization is a technique to transform the values of a dataset into a common scale. It helps to reduce data redundancy and improve data integrity.
- 2. Data Analytics:** This is the most important part as we visualize the correlations that we obtain. From our data analysis, we intend to obtain 2 functionalities:
- a. Meaningful correlations between features of music and behavioral characteristics of the participant.
 - i. For obtaining meaningful correlations and inferences, we can plot different charts and graphs to obtain the relation between each of the individual features to each individual behavioral characteristic.
 - ii. To find the correlation between all the features and behavioral characteristics, we will use linear regression. The corresponding θ values will give us the weight of each contributing feature.
 - b. Predict the behavioral characteristics of a new person on the basis of the features of the music that they listen to.
 - i. Based on the linear regression curve that we have obtained from the previous step, we will predict the anxiety, depression, insomnia, and OCD scores of the individual after performing preprocessing steps again on the new test data of the individual.
 - ii. On the basis of the above scores obtained for different behavioral characteristics, we can speculate the mental health of the participant.

What the data analysis is expected to teach us about the topic?

We hypothesize that music and increasing exposure to music have a relationship with the mental health conditions of the individuals. The following are *some of the expected correlations* that might be deduced from the data.

- We believe that music has a therapeutic effect on the mental health of an individual.
- Due to music being a source of dopamine, exposure to music alters or impacts the baseline dopamine levels of the brain and thereby influences the reward/pain balance of

the brain which can contribute to mental health problems. Therefore, there can be a relationship between the amount of time a person has spent listening to music and their mental health status. We hypothesize that listening to music for a certain amount of time can turn out to be therapeutic, however, excessive exposure to music can increase the severity of mental health problems.

- We hypothesize that the genre of music affects its impact on mental health conditions, either positively, or negatively. For example, music with a high tempo can negatively impact the mental health of individuals while music with less tempo can turn out to be therapeutic in nature.
- There can also be a reverse correlation between music and the mental health status of an individual. For example, an individual with depression might listen to sad music. Hence, we hypothesize a bi-directional relationship between music and mental health and we intend to learn about this relationship.
- While music has a relationship with the mental health of individuals, the genre of music which greatly affects individuals, both positively, and negatively, depends on the age of the individual and the prevalent genre of music of the time of their early years and the current time. For example, an elderly person might benefit from slow music and could experience a negative impact on fast music.
- A niche observation from the dataset is that there is a variable called “frequency [video game music]”. We could look at the relationship between such music and mental health conditions as it relates to the research topic of the relationship between the time spent playing video games and the mental health status of an individual.