**HDSC Spring ’23 Premiere Project Presentation by Team Scipy**

**Prediction of the Therapeutic Effect of Music on Mental Health**

Music has a profound impact on mental health and well-being. It can elicit powerful emotional responses, provide comfort, and offer an outlet for self-expression. Some questions to be asked can be how frequent you listen to music, the type of music you listen to. Music is a powerful art form that touches our hearts, stimulates our minds, and enriches our lives. Its ability to communicate emotions, unite people, and provide a means of self-expression makes it an essential part of human culture and a source of joy and inspiration for millions of people worldwide.

Music therapy, or MT, is the use of music to improve an individual's stress, mood, and overall mental health. MT is also recognized as an evidence-based practice, using music as a catalyst for "happy" hormones such as oxytocin.

**Aims and Objectives**

The aim of studying the therapeutic effects of music on mental health is to investigate and understand how music can be used as a therapeutic tool to promote psychological well-being, alleviate symptoms of mental disorders, and enhance overall mental health.

Flow Process

The flow chart below illustrates the steps taken:





**Data Collection**

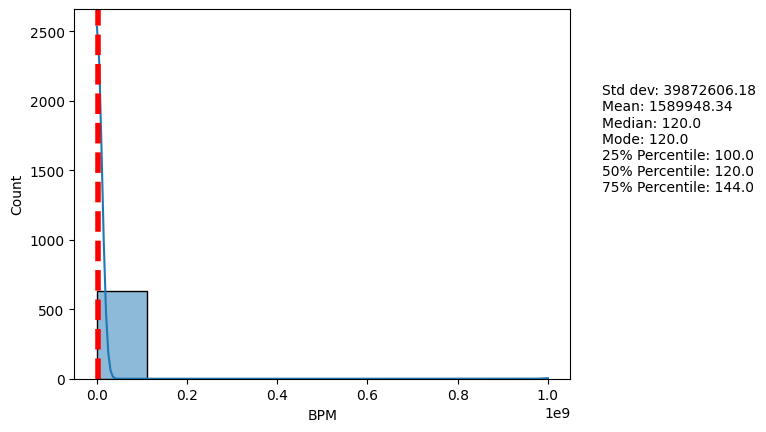
The data used for this task is an open-source data gotten from Kaggle, it can be accessed via the link below:

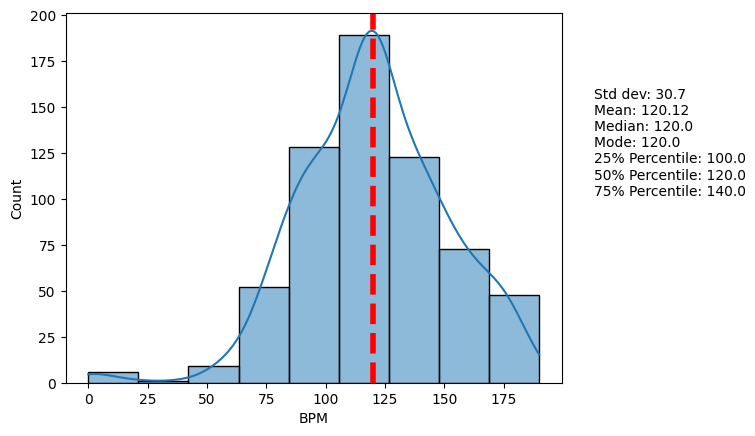
<https://www.kaggle.com/datasets/catherinerasgaitis/mxmh-survey-results>

**Data Preparation**

**Removing Outliners:** This has some nasty outliers that need to be replaced, soby exploring values around the mode and median we can select a suitable threshold to set max values and fill in missing values with the mean.

The images below shows before and after removing the outliners

****

****

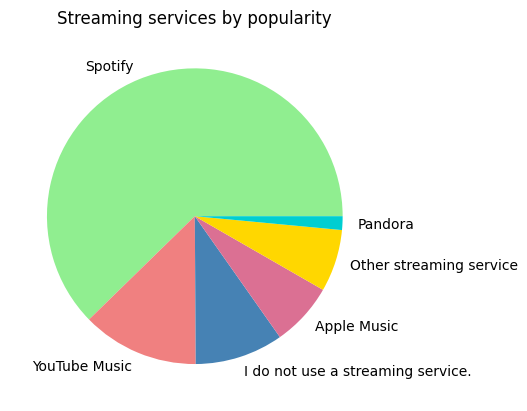
**Data Cleaning**: Here we changed some of the variables so make them shorter and easier to analyze. For example under the primary streaming service column, responses like ‘Other streaming services’ and ‘I do not use a streaming service’ were changed to ‘Other’ and ‘None’

Also some categorical variables were changed to numeric like the responses to how frequently you listen to music; (‘Never’, 'Rarely', 'Sometimes', 'Very frequently') to (0, 1, 2, 3).

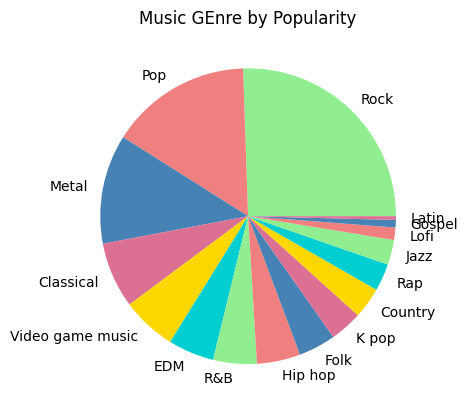
**Handling Missing Values:** Missing values in numeric columns like Age and BPM columns were replaced by the mean. While the categorical values were replaced with the mode.

**Exploratory Data Analysis**

After data cleaning some visualizations were carried out on the dataset to get some insights about it.



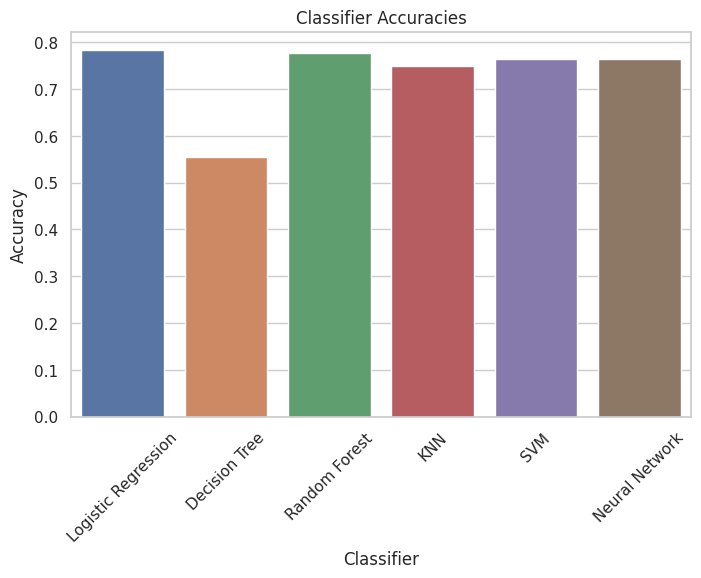
This chart shows that the most popular streaming service used is Spotify by a very large margin.



And the most popular genre is rock.

**Model Training, Evaluation, and Validation**

1. **Model Training**: The dataset was split into a training dataset and a testing dataset with a split percentage of 80:20. 80% was used to train the data using different regression models. Models used were Logistic Regression, Decision Tree Accuracy, Random Forest Accuracy, KNN, SVM and Neural Network.
2. **Model Evaluation:** The evaluation metric used is sklearn accuracy score since this is a categorization problem and the results is as follows:Logistic Regression Accuracy: 0.7837837837837838, Decision Tree Accuracy: 0.5675675675675675, Random Forest Accuracy: 0.7702702702702703, KNN Accuracy: 0.75, SVM Accuracy: 0.7635135135135135,NN Accuracy: 0.7432432432432432
3. Logistic Regression has the highest accuracy



**4. Model Deployment:** Link to the github: https://github.com/OmobaCoder/ScipyTeamProject

**Conclusion and Recommendation**

Hyperparameter tuning can improve results, for example, adjusting the Max BPM feature down from 200 to 190 resulted in an increase in model accuracy from 77 to 78 % for logistic regression as well as slight improvements in other models, also more training data, If we increased training data to 90:10 split, accuracy went to 80% for example.