# Part A – Data Science Project

Executive Summary

Question Asked, **"Is there a connection between music therapy and mental health?”**

This aim of this project is the explore the connection between music therapy and self-assessed mental health, particularly identifying patterns in how music therapy may influence stress, anxiety, and OCD symptoms.

The dataset was sourced from Kaggle. Initial steps included the removal of null and invalid entries and the anonymisation of patient details to ensure privacy. My approach includes exploratory data analysis (EDA) using Python, supported by visualisations to examine correlations.

In addition to EDA, I plan to apply K-means clustering to segment patients based on their mental health responses and therapy usage. This technique is well suited to the dataset as it can group individuals with similar profiles, such as anxiety levels and hours of music therapy exposure, revealing hidden patterns that may not be visible through basic analysis. These insights can help healthcare professionals target specific subgroups and tailor mental health interventions accordingly.

The results of this analysis will help inform researchers and practitioners about the effectiveness of music therapy and its potential integration into broader mental health strategies.

GitHub Link <https://grace-taylor21.github.io/DSPortfoliov2/>

## EDA

Using Python, code included in Figure 1, I’ve ran basic EDA to check the quality of the data being used.A screen shot of a computer program

Description automatically generatedA screen shot of a computer code

Description automatically generated

B

A

C

D

Part A & B- This returns that there are 736 entries in the data, in 33 columns and the number of non-null entries in the below columns.

A screenshot of a computer program

Description automatically generated

The first 5 rows are printed to give an immediate view of the structure of the data, ensuring that all categorical information is in the right format, and provides a faster view of any immediate issues with analysing the data.A screenshot of a computer

Description automatically generated

The describe function is providing insights into the numerical columns, such as number of non null values in each column, the average values of the columns, measure the spread of the data (including the min, max and standard deviation).

A screenshot of a computer

Description automatically generated

Part C – Distribution graphs are produced to provide visual context to the distribution columns for example most patients sit within the younger range 18-25, the hours per day predominantly sits within 0-5 hours, anxiety is leaning towards the higher end of the scale.

A group of blue and white bars

Description automatically generated

Part D – correlation heat mapping to quickly identify any mild to strong correlations:

A screenshot of a graph

Description automatically generated

The disorders show positive correlations with each other, albeit to a moderate level.

## K Means Clustering

Firstly, using the elbow method I will plot the Within Cluster sum of Squares to look for the elbow point to determine the number of clusters, however this method relies on personal interpretation of an ‘elbow’ and may be unreliable. (Schubert, 2023) in future analysis I would explore Silhouette Clustering.

A computer screen shot of a program

Description automatically generated

A graph with a line

Description automatically generated

As the gradient of the curve is steep between clusters 1 to 4 and then flattens suggests the appropriate number of clusters would be 4. Any clusters after this would yield diminishing returns in how well the model fits.

With my clustering number I run the below python script to generate my K means Cluster output graph:

A computer screen shot of a program

Description automatically generated

A graph with different colored dots

Description automatically generated

Creating the following bar plots in python:

A screen shot of a computer program

Description automatically generated

A collage of graphs and charts

Description automatically generated

From the Bar graphs the following observations can be made:

* Video Game Music – Most commonly listened to by a younger audience and appears detrimental to all 4 types of Mental health.
* Jazz/R&B/Metal – Shown to improve all Mental health groups
* K Pop/Country – Only impacted insomnia
* Rock/Hip Hop – Split reactions, with improvements in Anxiety and OCD but negative effects in Insomnia and significantly worsen effects in Depression.
* EDM – Again split reactions with improvements seen in Insomnia and Depression with the adverse shown in Anxiety and OCD.
* Pop – Showed negative impacts on all mental health
* Rap – Improvements in Anxiety, Depression and Insomnia with the largest number of hours listened clocked.
* Classical – Showed positive effects on Insomnia but negative in Anxiety and severely negative in Depression and OCD.
* Folk – Positive impacts on all areas except insomnia
* Lofi/Gospel – only genres to only have positive implication on mental health. Although they are on opposite ends of the age groups.
* Latin – Positive impacts on Depression and Anxiety
* Anxiety – All participants experience anxiety although participants that prefer Rock, Jazz, K-Pop, Hip Hop, Pop and folk music have higher self-reported levels of anxiety, 6+. Which may be linked to their energetic nature.
* Insomnia – Overall the levels of insomnia are fairly low (below 4), however those participants favouring Metal, Lofi or Gospel experience higher levels of Insomnia.
* OCD – Seen predominately in listeners of Rap and Lofi, this may indicate an obsessive behaviour link.
* Depression – High levels of reported depression seen in listeners of Lofi, Hip Hop and Rock music.

## Conclusion

This project explored the relationship between music therapy and self-reported mental health outcomes, focusing on stress, anxiety, and OCD indicators. Through exploratory data analysis and clustering techniques, patterns emerged suggesting that music therapy may positively influence mental health—particularly among younger individuals and those engaging more regularly with it. The use of K-means clustering proved valuable in identifying subgroups with similar responses, laying the groundwork for more personalised therapeutic approaches.

While these findings are exploratory, they underscore the potential for integrating music therapy into broader mental health strategies. With further refinement—such as model validation and deeper statistical testing—these insights can help design more effective, evidence-based mental health interventions. Importantly, the results also have practical value for clinical stakeholders, who could use this information to prioritise patients for music-based therapies and shape holistic treatment plans. These insights may also support funding and policy decisions where integrative mental health approaches are being considered.

**Reflection and Limitations**

This project provided a valuable foundation for understanding the therapeutic role of music in mental health. However, several limitations should be acknowledged. The dataset relied on self-reported psychological assessments, which can introduce subjectivity and variability—an issue noted in the literature (Greenberg, 2017). Additionally, the lack of longitudinal data limits the ability to assess the long-term impact of music therapy. Future studies should aim to incorporate clinical assessments and track patient outcomes over time to increase reliability.

From a methodological perspective, K-means clustering enabled meaningful segmentation but is sensitive to the choice of k and assumptions around distribution and cluster shape. To strengthen future modelling, techniques such as silhouette analysis, hierarchical clustering, or DBSCAN could be employed to validate or refine cluster integrity.

While this formative project focused on exploratory data analysis, future iterations should incorporate predictive models and statistical testing (e.g. ANOVA, regression) to explore potential causal relationships. In addition, translating the findings into interactive dashboards could enable healthcare professionals, researchers, and policy teams to interpret and act on insights more effectively. Despite these limitations, this project marks a significant step toward evidence-based, data-driven approaches in mental health research.

**Horizon Analysis**

Horizon scanning in music therapy leverages data science methodologies to identify emerging trends, research advancements, and innovative therapeutic applications in mental health. By systematically analysing diverse sources—ranging from academic literature and clinical trials to social media discourse and patient feedback—data-driven techniques can uncover critical insights into the evolving role of music in mental well-being. For example, recent studies have demonstrated the potential of brain synchronisation between patients and therapists during music therapy sessions, providing a deeper understanding of its therapeutic mechanisms (University, 2019).

Additionally, advancements in predictive analytics and natural language processing (NLP) enable the identification of emerging therapy preferences, novel musical interventions, and the integration of AI-generated compositions or biofeedback technologies into therapeutic frameworks. Research further highlights the growing role of big data in music therapy, allowing for a systematic exploration of biopsychosocial and musical variables linked to behavioural and psychological improvements (Greenberg, 2017). Moreover, bibliometric analyses have been used to map research frontiers, guiding practitioners and policymakers in shaping future therapeutic approaches (Zhi, 2024).

By integrating horizon scanning into music therapy research, professionals can proactively adapt to emerging developments, ensuring that therapeutic interventions remain innovative, evidence-based, and responsive to the evolving needs of mental health care.