Music and Mental Health: Comparative Analysis of Survey Data

Leon Holub[1], Kateřina Křížanová[1], David Marek[1], Petr Nosek[1]

[1]Tomas Bata University in Zlín, Faculty of Applied Informatics, Department of Software Engineering, Czech Republic

December 12, 2025

Abstract

Purpose – This study investigates the impact of everyday music listening habits—including genre preference, listening frequency, and listening context—on indicators of mental health such as depression, anxiety, insomnia, and perceived emotional effects of music. It addresses a gap in the literature by examining multiple dimensions of musical behavior simultaneously using a large public dataset, offering new insights into how music-related patterns correspond with psychological well-being.

Design/Methodology/Approach – A quantitative approach was used, applying statistical analyses and supervised machine-learning models to the publicly available MxMH survey dataset (N = 736). The data include demographic information, musical preferences, listening behaviors, and self-reported mental-health scores. Analyses involved one-way ANOVA, t-tests, Pearson correlations, and predictive modeling with logistic regression and random forest classifiers, supported by ROC curve evaluation.

Findings – The results show significant differences in depression levels across music genres, a very weak relationship between listening frequency and depression, and no meaningful association with anxiety. Predictive models demonstrated only marginal ability to infer mental-health risk from musical preferences alone, though demographic and behavioral variables contributed modestly. Listening to music while working did not influence depression levels, and perceived emotional benefits of music were consistent across age groups. These findings indicate partial support for the hypotheses, while also revealing substantial variability not explained by musical factors.

Originality/Value – This research provides novel insights into how diverse aspects of music listening relate to mental health, highlighting both the potential and the limits of using everyday musical behavior as an indicator of psychological well-being. By combining statistical testing with predictive modeling, the study contributes to a more nuanced understanding of the role of music in emotional functioning.

Keywords: music listening; mental health; depression; anxiety; predictive modeling; genre preferences

1 Introduction

Music plays a significant role in everyday life and is frequently used as a tool for emotion regulation, mood enhancement, stress reduction, and personal expression. Prior research shows that musical engagement can influence emotional states, although the strength and nature of these effects vary between individuals. As music becomes increasingly embedded in daily routines, understanding its relationship to mental well-being has gained growing scientific relevance.

Existing studies have demonstrated that music can support emotional regulation and coping Saarikallio and Erkkilä, 2011 and that individuals with depressive tendencies often gravitate toward low-valence or melancholic music Miranda and Claes, 2007; K. Surana and Alluri, 2019. Large-scale analyses of listening behaviour further indicate that depressed listeners exhibit distinct engagement patterns, such as repetitive use of emotionally congruent music A. Surana *et al.*, 2020. At the same time, controlled experiments show that music can reduce anxiety and improve mood Gustavson *et al.*, 2021, and workplace studies report short-term benefits for affect and performance Lesiuk, 2005. However, most prior research relies on small or specific samples or examines only isolated aspects of musical behaviour, leaving limited understanding of how different listening habits jointly relate to mental-health indicators in everyday contexts.

This study addresses this gap by conducting a quantitative analysis of the publicly available *Music and Mental Health* dataset. The dataset provides detailed information on musical behaviours—including favourite genres, listening frequency, and music use during work—alongside self-reported measures of anxiety, depression, insomnia, and obsessive–compulsive tendencies. Using this broad, international sample, the study investigates five research questions examining how musical preferences, listening habits, and demographic factors relate to mental-health outcomes. The goal is to contribute to a more comprehensive understanding of how everyday music engagement corresponds with psychological well-being.

2 Research Questions

Based on the identified research gap regarding the connection between musical behaviour and mental health, this study formulates five research questions. Each question targets a specific aspect of how music-related habits may relate to psychological well-being.

- **RQ1: Does the level of depression differ between listeners of different music genres?**
  This question examines whether individuals with different favourite genres show statistically significant differences in self-reported depression levels.
- **RQ2: Is the frequency of listening to music related to anxiety or stress?**
  The objective is to determine whether higher daily exposure to music correlates with changes in anxiety or stress indicators.
- **RQ3: Can the risk of a mental disorder be predicted based on music preferences?**
  This question evaluates whether musical behaviour and demographic variables can serve as useful predictors of elevated mental health risk.
- **RQ4: Does listening to music while working influence the level of depression?**
  Here, we analyse whether individuals who listen to music during work or study differ in depression scores compared to those who do not.
- **RQ5: Does age play a role in how people perceive the effect of music on their mental health?**
  This question investigates whether the perceived effect of music (*improve*, *no effect*, *worsen*) varies across different age groups.

2.1 Hypotheses

- **H1**: Listeners of melancholic music genres will report higher depression scores than listeners of more uplifting or positive genres.
- **H2**: Higher daily music listening frequency will be associated with lower levels of stress or anxiety.
- **H3**: Musical preferences and listening behaviour will contain enough information for a machine-learning model to predict mental-disorder risk above chance level.

3 Method

This section describes the methodological approach used in this study. It summarises the dataset, the data-cleaning procedures, the statistical methods applied to address the research questions, and the machine-learning models used for predictive analysis. Each methodological step is presented in a structured way to ensure clarity and reproducibility.

3.1 Dataset

This study utilized the publicly available *Music and Mental Health (MxMH)* dataset published on the Kaggle platform by **Catherine Rasgaitis** (University of Washington, USA). The dataset was originally collected through an online survey distributed via Reddit forums, Discord servers, social media channels, and physical advertisements in public spaces such as libraries and parks. Data collection took place between **July 27 and November 8, 2022**. All responses are self-reported, fully anonymous, and released under the **CC0: Public Domain** license, allowing unrestricted use for research and educational purposes.

Table I: Dataset summary

Source	Kaggle – *Music and Mental Health (MxMH)* dataset
Author	Catherine Rasgaitis (University of Washington)
Collection period	July 27 – November 8, 2022
Collection method	Google Form survey shared online and via printed materials
Geospatial coverage	Worldwide
Number of respondents	736 participants
Number of variables	33 (demographics, musical behavior, psychological indicators)
License	CC0: Public Domain
Ethics	Anonymous responses, no personally identifiable information, GDPR-compliant

The dataset includes three main groups of variables:

- **Demographic and behavioral data** – age, primary streaming service, average listening time per day, and whether the respondent listens to music while working, plays an instrument, or composes music.
- **Musical preferences** – favourite genre (*Fav genre*) and frequency of listening to 16 specific genres (e.g., Classical, Rock, Pop, Jazz, EDM, Hip Hop, Metal, Lo-fi, Video-game music). Each frequency variable uses four categorical levels: *Never, Rarely, Sometimes, Very frequently.*
- **Psychological indicators** – continuous self reported scores for *Anxiety*, *Depression*, *Insomnia*, and *OCD* ranging from 0 to 10, where 0 represents "I do not experience this" and 10 indicates "I experience this regularly or to an extreme." The variable *Music effects* is represented as an ordered categorical variable with levels Worsen, No effect, and Improve.

Numeric variables (*Age*, *Hours per day*, *Anxiety*, *Depression*, *Insomnia*, *OCD*, *BPM*) are stored as floats, while categorical attributes (e.g., genres, frequency categories, Yes/No responses) are encoded as strings. Before statistical processing, missing or inconsistent values were removed, and categorical responses were standardized for consistency across entries.

Potential dataset limitations include the self-reported nature of the responses, unequal distribution of musical genres, and the lack of clinical validation for psychological scores. Nevertheless, the MxMH dataset provides a unique, diverse sample suitable for quantitative analysis of the relationship between musical preferences and mental health indicators.

3.1.1 Variable Overview

Table II: Overview of dataset variables

Column Name	Description
Timestamp	Date and time when the Google Form response was submitted.
Age	Respondent's age in years (numeric).
Primary streaming service	Respondent's primary platform for listening to music (e.g., Spotify, YouTube Music, Apple Music, Pandora, Other).
Hours per day	Average number of hours per day the respondent listens to music.
While working	Indicates whether the respondent listens to music while studying or working (Yes/No).
Instrumentalist	Indicates if the respondent regularly plays a musical instrument (Yes/No).
Composer	Indicates if the respondent composes or creates music (Yes/No).
Fav genre	The respondent's favourite or most frequently listened-to genre.
Exploratory	Whether the respondent actively explores new artists or genres (Yes/No).
Foreign languages	Whether the respondent regularly listens to music with lyrics in a language they are not fluent in (Yes/No).
BPM	Beats per minute of the respondent's favourite genre (numeric).

Column Name	Description
Frequency [Classical]	Frequency of listening to classical music (*Never, Rarely, Sometimes, Very frequently*).
Frequency [Country]	Frequency of listening to country music.
Frequency [EDM]	Frequency of listening to electronic dance music (EDM).
Frequency [Folk]	Frequency of listening to folk music.
Frequency [Gospel]	Frequency of listening to gospel music.
Frequency [Hip hop]	Frequency of listening to hip hop music.
Frequency [Jazz]	Frequency of listening to jazz music.
Frequency [K pop]	Frequency of listening to Korean pop music.
Frequency [Latin]	Frequency of listening to Latin music.
Frequency [Lofi]	Frequency of listening to lo-fi music.
Frequency [Metal]	Frequency of listening to metal music.
Frequency [Pop]	Frequency of listening to pop music.
Frequency [R&B]	Frequency of listening to R&B music.
Frequency [Rap]	Frequency of listening to rap music.
Frequency [Rock]	Frequency of listening to rock music.
Frequency [Video game music]	Frequency of listening to video game or soundtrack-style music.
Anxiety	Self-reported level of anxiety on a scale from 0 (none) to 10 (extreme).
Depression	Self-reported level of depression on a scale from 0 (none) to 10 (extreme).
Insomnia	Self-reported level of insomnia on a scale from 0 (none) to 10 (extreme).
OCD	Self-reported level of obsessive–compulsive disorder on a scale from 0 (none) to 10 (extreme).
Music effects	Ordered categorical variable with levels: Worsen, No effect, Improve.
Permissions	Respondent's consent to publish data (Yes/No).

3.2 Data Cleaning and Preparation

All analyses were conducted in Python (3.x) using `pandas`, `numpy`, `matplotlib`, and `scipy/scikit-learn`. A dedicated preprocessing script was implemented to ensure that all steps are fully reproducible.

First, technical metadata fields *Timestamp* and *Permissions* were removed, as they do not contain information relevant to the research questions. The raw dataset was then filtered to retain only responses with valid entries in key analytical variables: *Depression*, *Fav genre*, *Music effects*, *Hours per day*, *Anxiety*, *Insomnia*, *While working*, and *Age*. Rows containing missing values, empty strings, or placeholder values (e.g. "N/A") in any of these fields were excluded.

The variable *Music effects* was transformed into an ordered categorical variable (Worsen ¡ No effect ¡ Improve). All other variables were kept in their original formats for subsequent analysis, with further recoding or grouping applied only within the corresponding analytical procedures described in the Statistical Analysis section.

This scripted preprocessing pipeline ensures that the same cleaned version of the dataset can be reproduced consistently from the original Kaggle source.

3.3 Statistical Tests

To address the research questions and evaluate the relationships between musical behaviour and mental health indicators, several statistical methods were applied. These included one-way ANOVA for comparing group means, independent samples t-tests for binary comparisons, Pearson correlation for assessing linear associations between continuous variables, and supervised machine-learning techniques (logistic regression and random forest) accompanied by ROC analysis to evaluate predictive performance. Each method was selected on the basis of the data type and the specific hypothesis associated with each research question.

3.3.1 ANOVA (One-way ANOVA)

One-way ANOVA was used to test for statistically significant differences in mean depression scores across multiple categorical groups. This method is appropriate when comparing the means of three or more independent groups and assumes approximate normality and homogeneity of variances.

ANOVA was used in two parts of the study:

- **RQ1:** To compare depression levels across different favourite music genres.

- **RQ5:** To examine potential differences in psychological variables across age groups.

In both cases, ANOVA allowed us to evaluate whether variability between group means exceeded what would be expected from random variation alone.

3.3.2 t-test (Independent Samples t-test)

Independent samples t-tests were employed when analysing differences between two groups. This method tests whether the means of two independent populations differ significantly and is appropriate for binary comparisons.

The t-test was applied in:

- **RQ4:** To assess whether individuals who listen to music while working differ in depression levels from those who do not.

This test enabled a straightforward evaluation of whether the presence of music during work is associated with measurable differences in depressive symptoms.

3.3.3 Pearson correlation

Pearson's correlation coefficient was used to quantify the strength and direction of linear relationships between continuous variables. This method assumes interval-scale measurement and approximately linear associations.

Pearson correlation was used in:

- **RQ2:** To assess whether daily music listening time is related to anxiety or depression levels.
- **RQ5:** To explore whether age is associated with any psychological variables included in the dataset.

The resulting correlation coefficients provided a simple and interpretable measure of association, complementing the group-based comparisons conducted through ANOVA and t-tests.

3.3.4 Model Training and ROC Analysis

To address RQ3, supervised classification models were trained to evaluate whether musical preferences and listening behaviours can predict mental-health risk. Logistic regression and random forest classifiers were implemented to compare linear and non-linear decision boundaries. Model performance was assessed using standard classification metrics, including accuracy, precision, and recall.

Receiver Operating Characteristic (ROC) analysis was used to further evaluate model quality. The ROC curve plots the true positive rate against the false positive rate across multiple decision thresholds, while the Area Under the Curve (AUC) provides a threshold-independent measure of discriminatory ability. AUC values above 0.5 indicate performance better than chance, with higher values representing stronger predictive capacity.

This framework enabled a systematic comparison of predictive features and model robustness, supporting the evaluation of Hypothesis H3.

3.4 Machine Learning

This section describes the machine learning methods used to address RQ3, including model selection, feature construction, the definition of the target variable, and evaluation procedures. All analyses were implemented in Python using `scikit-learn`.

3.4.1 Logistic Regression (LR)

Logistic Regression is a linear classification method that estimates the probability of a binary outcome using the logistic function. The model is interpretable, computationally efficient, and commonly used as a baseline in psychological and behavioural prediction studies. In this project, LR was trained with a maximum of 1000 iterations (`max_iter = 1000`) to ensure convergence of the optimization process.

3.4.2 Random Forest Classifier (RF)

The Random Forest classifier is an ensemble learning method based on aggregating multiple decision trees. RF is robust to noise, can capture nonlinear relationships, and provides model-agnostic estimates of feature importance. The implementation used 300 trees (`n_estimators = 300`) and default settings for other hyperparameters.

3.4.3 Target Variable (RQ3)

To address RQ3, a binary outcome variable *MentalDisorderRisk* was created from four psychological indicators included in the dataset: *Anxiety*, *Depression*, *Insomnia*, and *OCD*. All four variables are self-reported on a 0–10 scale as part of the questionnaire completed by participants. Following common conventions in mental-health screening, respondents scoring above 6 on at least one indicator (i.e., 7–10) were classified as "high risk", while all others were labeled as "low risk". This transformation enabled the application of supervised classification models.

3.4.4 Features Used for Prediction

The prediction models were trained on the following set of predictors: respondent age (*Age*), average daily listening time (*Hours per day*), listening behaviour (*While working*), perceived effect of music (*Music effects*), and favourite music genre (*Fav genre*). Categorical variables were transformed into one-hot encoded dummy variables using `pandas.get_dummies` (`drop_first = True`), while numerical variables were converted to numeric types and cleaned of missing values. These predictors represent the demographic, behavioural, and preference-related factors that may plausibly relate to psychological well-being.

3.4.5 Model Training and Evaluation (RQ3)

Both models (LR and RF) were trained on the same feature set to allow a direct comparison. The dataset was split into training and test subsets using a 70/30 stratified split (`train_test_split` with `test_size = 0.30`, `stratify = y`, `random_state = 42`). Model performance was evaluated using accuracy, F1-score, and the area under the ROC curve (AUC). ROC curves were generated for both models, and feature importance values from the Random Forest classifier were used to identify the most influential predictors.

3.4.6 Train/test split

To evaluate the predictive models developed for RQ3, the dataset was divided into separate training and testing subsets. This approach, known as a *train/test split*, allows for assessing how well a model generalises to unseen data. The training portion is used to fit the model, while the test portion provides an unbiased estimate of predictive performance.

In this study, a 70/30 split was applied: 70% of the data was used to train the logistic regression and random forest models, while the remaining 30% served as the independent test set. This proportion preserves a sufficiently large training dataset for learning while ensuring that a substantial portion is held out for evaluation, improving the reliability of the resulting performance metrics.

By separating training and testing data, we can evaluate performance measures such as accuracy, precision, recall, and AUC in a way that reflects the model's ability to generalise to new, unseen instances rather than its fit to the training data. This is particularly important for RQ3, where the goal is to determine whether musical preferences and behavioural variables can meaningfully contribute to predicting mental-health risk.

Although the train/test split is effective for reducing overfitting, it introduces variability because the results may depend on how the dataset is partitioned. Future work may consider techniques such as cross-validation to provide more stable and generalisable performance estimates.

3.5 Tools and Environment

All analyses in this project were conducted in Python 3.10, using a modular codebase designed to ensure clarity, reproducibility, and separation of concerns. The computation environment was managed via a dedicated virtual environment, with all dependencies listed in the `requirements.txt` file, which is publicly accessible in the project's GitHub repository:
`https://github.com/Leon-Holub/AP9SI/blob/master/requirements.txt`.

Python Libraries. The analysis relied on a set of widely used scientific and data-processing libraries. Core data manipulation was performed using `pandas`, while `numpy` was used for numerical operations. Visualisations were created with `matplotlib` and `seaborn`. Statistical procedures and machine learning models (Random Forest and Logistic Regression) were implemented using the `scikit-learn` library. Supporting libraries included `scipy` for statistical tests, `tqdm` for progress tracking, and `requests` and `kagglehub` for dataset retrieval. The complete specification of package versions is available in the Git repository under `requirements.txt`.

Project Structure. The project follows a clear modular structure:

- `main.py` — orchestrates the execution of all analyses by calling functions defined in supporting modules.
- `FileLoader.py` — responsible for dataset loading and initial preprocessing.
- `PlotCreator.py` — contains visualisation utilities centralised for consistent styling across all graphs.
- `ResearchQuestions.py` — includes all analytical functions addressing RQ1–RQ5 (ANOVA tests, correlations, predictive modelling, and group comparisons).
- `plots/` — stores all generated visual outputs, including subfolders such as `age_music_effect/` and `Q3_predikce/`.
- `requirements.txt` — specification of all required Python packages for reproducibility.

Execution Workflow. The pipeline is designed so that executing `main.py` runs the full analysis sequence: loading the dataset, computing descriptive statistics, performing hypothesis tests for each research question, training predictive models for RQ3, and exporting all resulting plots. This ensures a fully automated and replicable workflow.

Overall, the development environment and tooling were selected to support reproducible, transparent analysis while maintaining a modular codebase suitable for extending or rerunning individual research steps.

3.5.1 Computational Environment for Predictive Modelling

The predictive models developed for RQ3 (Random Forest and Logistic Regression) were executed on a local workstation. The system specifications are provided below to ensure transparency and reproducibility of the computational setup. Although the models used in this study are not computationally intensive, documenting the hardware configuration supports the repeatability of the analysis.

Hardware configuration.

- **Machine:** HP Z400 Workstation
- **Processor:** Intel(R) Xeon(R) CPU W3565 @ 3.20 GHz (4 cores)
- **Installed RAM:** 24 GB DDR3
- **Graphics:** NVIDIA Quadro K620 (2 GB)
- **Storage:** 1.82 TB HDD + 112 GB SSD
- **Operating System:** 64-bit Windows (x64 architecture)

Relevance to the analysis. The machine learning models employed in this study (Random Forest and Logistic Regression from the `scikit-learn` library) are CPU-based and do not require GPU acceleration. The provided hardware configuration was fully sufficient for all computations, including model training, hyperparameter use of defaults, feature importance extraction, ROC curve generation, and cross-validation procedures. Training and evaluation times were consistently short (under several seconds per run), confirming that the analysis is reproducible on any standard modern workstation.

4 Results

This section presents the empirical findings obtained from the statistical analyses conducted to address the five research questions. The results are reported in the order of the research questions and include group comparisons using ANOVA and t-tests, correlation analyses examining relationships between musical behaviours and psychological variables, and predictive modelling outcomes. Each subsection summarises the relevant statistical evidence, supported by visualisations and descriptive statistics where appropriate.

4.1 RQ1: Do depression levels differ across music genres?

The analysis examined whether self-reported depression levels vary across listeners of different favorite music genres. A one-way ANOVA was conducted on depression scores across all genres with sufficient sample size (n ¿ 2). The test indicated a statistically significant overall effect of genre on depression levels, $F(15, \ldots)$=2.158, p=0.00806, suggesting that at least some genres are associated with higher or lower depression scores compared to others.

Descriptive statistics show a noticeable variation between genres. The highest mean depression levels were observed among Lofi listeners (M=6.60), **Hip hop** (M=5.88) and **Metal** (M=5.07), while listeners of **Gospel** (M=2.67), **R&B** (M=3.83) and **Latin** (M=4.50). Median values follow a similar pattern, with Lofi and Hip hop listeners exhibiting medians of 8.00 and 7.00 respectively, contrasting with the much lower median of 1.00 among Gospel listeners. The general variability between genres was moderately high, with standard deviations typically around 2.6 to 3.2.

Figure 1 visually illustrates these differences, showing that some genres—particularly Lofi, Hip hop, and Metal—tend to have higher central depression scores and wider distributions. In contrast, genres such as Gospel, R&B, and Classical show lower central tendencies and generally narrower ranges. Although not all pairwise differences were tested, the ANOVA results confirm that depression levels are **not uniformly distributed** between genres and that genre preferences are associated with a significant variation in reported depressive symptoms.

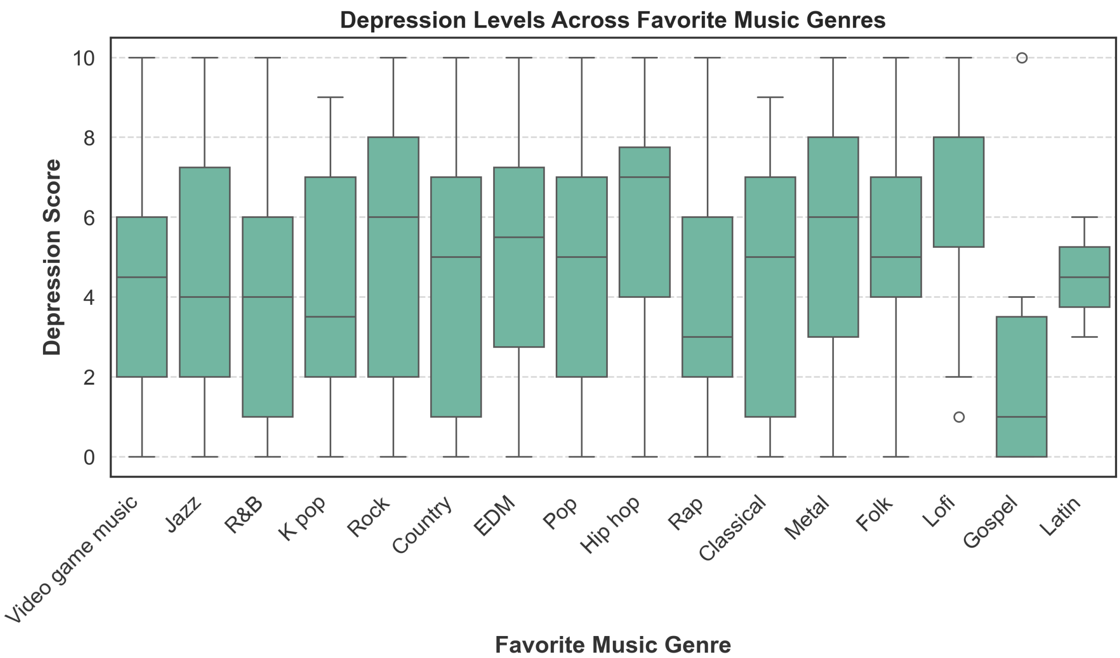


Figure 1: Depression levels across favorite music genres.

4.2 RQ2: Does music listening frequency relate to anxiety or depression?

To evaluate whether the amount of time respondents spend listening to music is associated with psychological indicators, Pearson correlation coefficients were computed between *Hours per day* and the variables *Anxiety* and *Depression* (Figure 2). The analysis revealed a very weak positive correlation between listening frequency and anxiety levels (r = 0.041), which was not statistically significant ($p = 0.27774$). Thus, no measurable relationship was observed between listening duration and anxiety.

For depression, the correlation was likewise very weak and positive, but in this case statistically significant ($r = 0.099$, $p = 0.00778$). This suggests that respondents who spend more time listening to music tend to report slightly higher levels of depressive symptoms; however, the effect size is minimal and does not allow for meaningful predictive conclusions.

Overall, the results indicate no substantive association between listening frequency and anxiety, and only a very small but statistically detectable association with depression, which should not be interpreted as evidence of a causal effect.

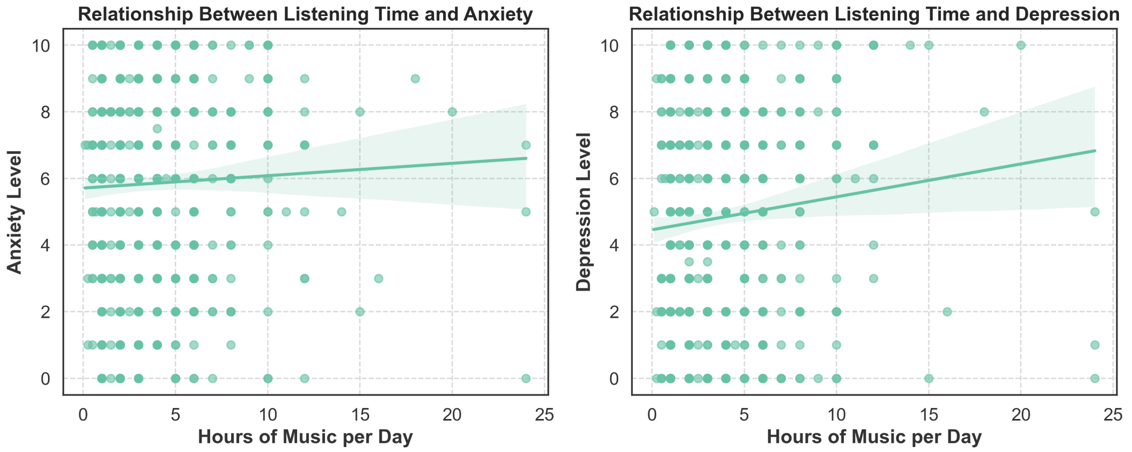


Figure 2: Relationship between daily music listening time and psychological indicators.

4.3 RQ3: Can mental disorder risk be predicted from music preferences?

This analysis evaluated whether the binary outcome *MentalDisorderRisk* (see Methods Section 3) can be predicted from music preferences, listening habits, and age. Two supervised learning models—Logistic Regression and Random Forest—were trained on the same feature set.

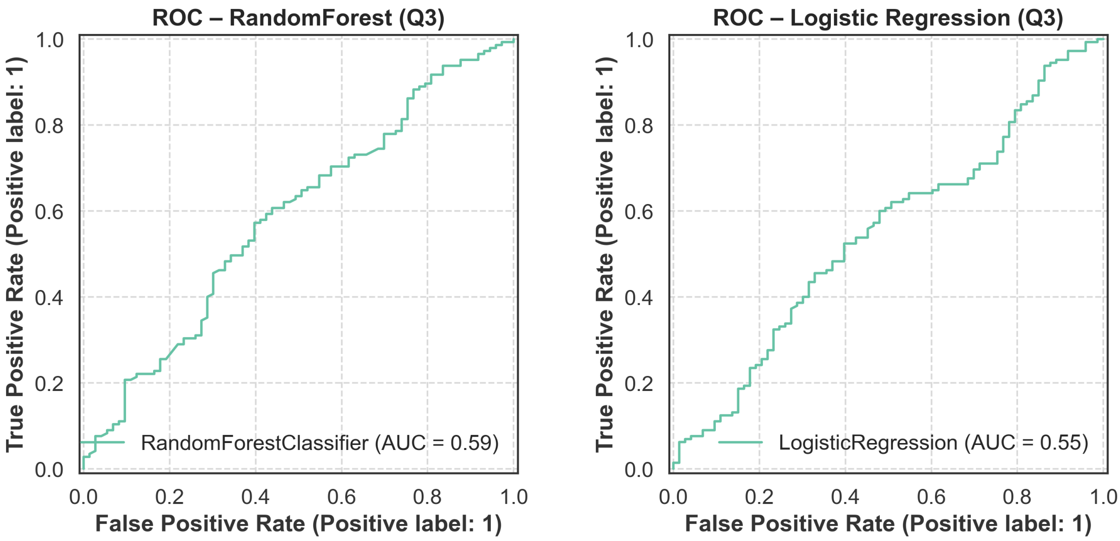
Model performance was generally weak. The Random Forest classifier achieved an accuracy of 0.56, F1-score of 0.55, and ROC AUC of 0.56, while Logistic Regression reached an accuracy of 0.57, F1-score of 0.56, and ROC AUC of 0.58. These results indicate only marginal improvement over random guessing. The ROC curves shown in Figures 3a and 3b confirm this, as both curves remain close to the diagonal with limited class separability.

Table III: Performance metrics for both classification models used in RQ3.

Model	**Accuracy**	**F1-score**	**ROC AUC**
Random Forest	0.56	0.55	0.56
Logistic Regression	0.57	0.56	0.58

Despite the low predictive performance, the Random Forest feature importance analysis offers insight into the relative contribution of each variable (Figure 4). *Age* emerged as the strongest predictor, followed by *Hours per day*. These findings suggest that younger respondents and individuals who listen to music for longer periods tend to fall more frequently into the high-risk category. Moderate contributions were observed for *Music effects* and *While working*. In contrast, favourite music genres showed only minimal predictive relevance, indicating that genre preference alone does not provide meaningful information about mental-health risk.

Overall, the results indicate that music preferences by themselves are insufficient for reliable mental-disorder risk prediction. Although some behavioural and demographic variables show modest associations with the target, the model performance suggests that additional psychological or contextual factors would be required to achieve meaningful predictive accuracy.



(a) Random Forest (Q3)

(b) Logistic Regression (Q3)

Figure 3: ROC curves for both classification models used in RQ3.

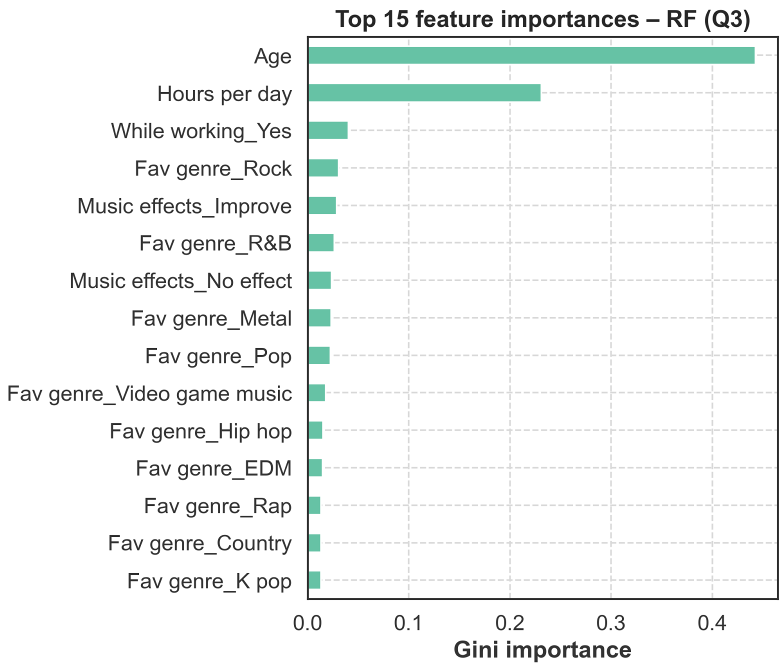


Figure 4: Top 15 feature importances for the Random Forest model (Q3).

4.4 RQ4: Does listening to music while working influence depression levels?

This research question investigated whether individuals who listen to music while working or studying exhibit different depression levels compared to those who do not.

To test this, the dataset was split into two groups based on the *While working* variable (Yes/No), and an Independent Samples t-test was performed on the *Depression* scores.

The descriptive statistics revealed that respondents who listen to music while working had a slightly higher mean depression score ($M = 4.89$) compared to those who do not ($M = 4.48$). However, the t-test results indicated that this difference is not statistically significant ($t = 1.455$, $p = 0.147$).

Figure 5 visualizes the distribution of depression scores for both groups. While the interquartile range (IQR) is similar for both categories, the median depression score appears virtually identical, reinforcing the lack of a significant statistical effect.

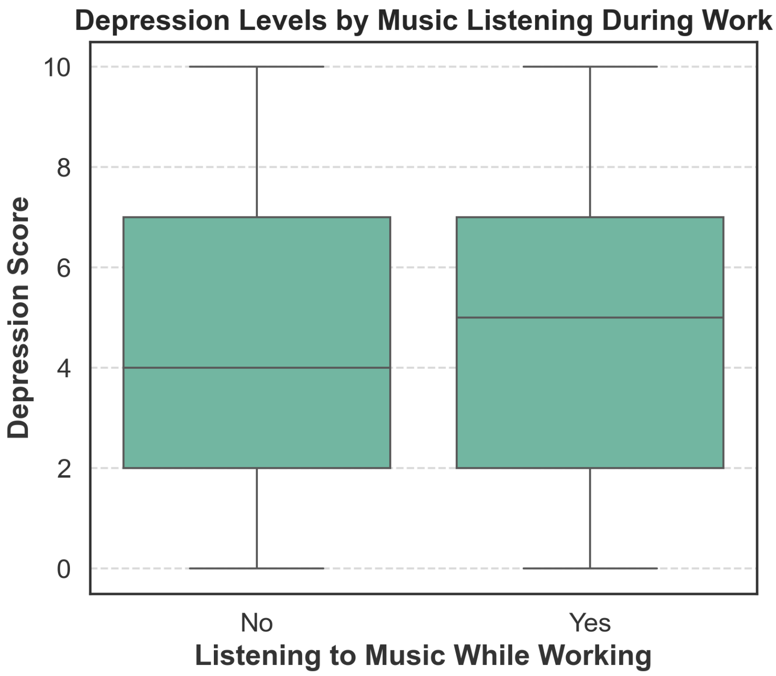


Figure 5: Depression levels while working and listening to music or not.

4.5 RQ5: Does age affect how people perceive the effects of music?

To examine whether the perceived effect of music on mental well-being differs by age, we compared responses to the *Music effects* question (*Worsen*, *No effect*, *Improve*) across six age groups: < 20, 20–29, 30–39, 40–49, 50–59, and 60+. Summary statistics for each group are shown in Table IV.

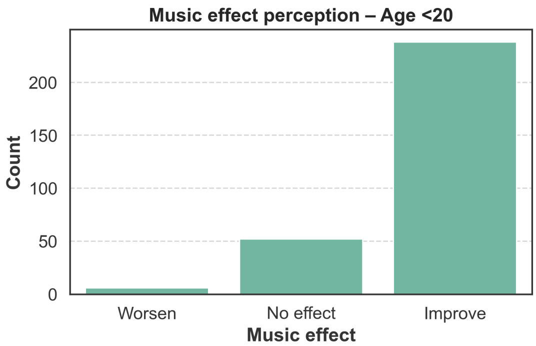
As illustrated in Figure 6, the overall pattern is highly consistent across all age groups. In every group, most respondents reported that music *improves* their mental state. This is also reflected in the numerical results: all age groups have a median value of 1.00, which corresponds to the *Improve* category.

The mean values differ slightly between groups, but the differences are small. The strongest positive perception appears in the 60+ group (M = 0.84), followed by respondents under 20 (M = 0.78). Middle-aged groups show slightly lower averages (e.g., 30–39: M = 0.57), but the overall trend remains the same: most people report that music helps them feel better. Some of the variation is likely due to the different sizes of age groups (for example, $N = 296$ in < 20 compared to only $N = 21$ in the 50–59 group).

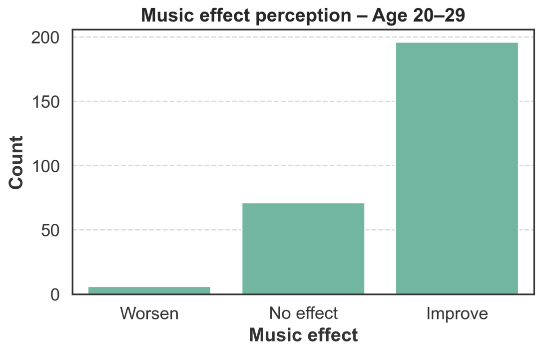
In summary, the results do not show significant age differences. Regardless of age, most of the respondents perceive music to have a positive effect on their mental well-being.

Table IV: Summary of perceived music effects across age groups

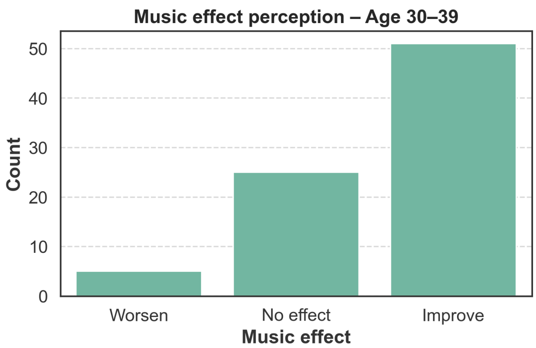
Age group	Mean	Median	N
< 20	0.78	1.00	296
20–29	0.70	1.00	273
30–39	0.57	1.00	81
40–49	0.69	1.00	29
50–59	0.67	1.00	21
60+	0.84	1.00	25



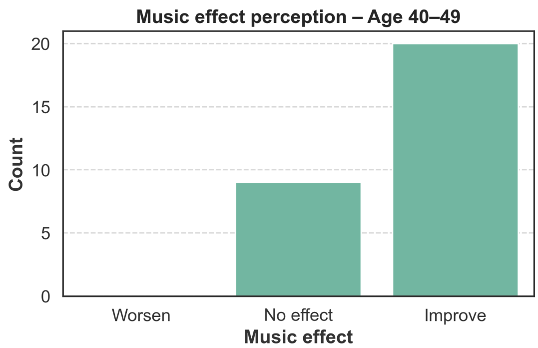
(a) Age <20



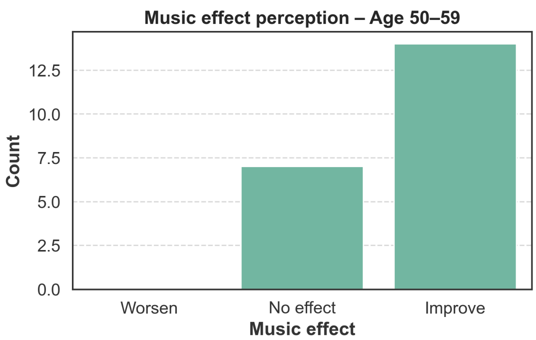
(b) Age 20–29



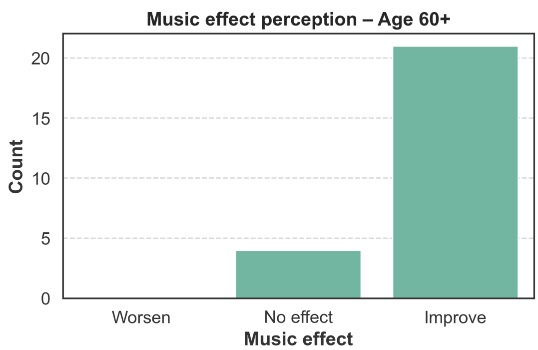
(c) Age 30–39



(d) Age 40–49



(e) Age 50–59



(f) Age 60+

Figure 6: Distribution of music effects across different age groups.

5 Discussion

In this section, interpret the findings in the context of the research questions. Compare the results with prior studies, discuss any unexpected outcomes, and address the broader implications for the field. Limitations and potential future research directions should also be included.

5.1 Interpretation of Findings

5.1.1 RQ1: Do depression levels differ across music genres? - Leon

The findings for RQ1 indicate that depression levels differ significantly across music genres, suggesting that preferred musical styles are associated with meaningful variation in emotional states. Genres such as Lofi, Hip hop, and Metal showed comparatively higher depression scores, whereas Gospel, R&B, and Classical listeners exhibited substantially lower levels. These patterns are consistent with previous research demonstrating that individuals with depressive symptoms tend to gravitate toward low-energy, melancholic, or emotionally intense music, often using it as a form of mood regulation or coping K. Surana and Alluri, 2019; Miranda and Claes, 2007. Conversely, uplifting or spiritually oriented genres such as Gospel have been linked to more positive emotional outcomes, which may explain the lower depression scores observed in this group.

One plausible explanation is that individuals experiencing psychological distress selectively choose music that reflects or validates their current emotional state. Prior work has shown that depressive adolescents and young adults often rely on emotionally charged or darker genres as a coping mechanism, using music to process or regulate negative affect Miranda and Claes, 2007. This aligns with our results, where the highest depression scores appeared among genres commonly described as introspective, atmospheric, or emotionally heavy. Similarly, Surana and Alluri K. Surana and Alluri, 2019 found that streaming patterns of depressed users consistently included more melancholic and low-arousal tracks, reinforcing the link between emotional distress and genre preference.

At the same time, the relatively high within-genre variance observed in our dataset suggests that genre preference alone cannot account for individual differences in mental health. Music listening is a multifaceted behavior influenced by personality traits, cultural background, social context, and listening motivations. For example, some individuals may listen to intense music for stimulation rather than emotional validation, while others may choose relaxing genres as an avoidance strategy. These complexities highlight the need for caution in inferring psychological states solely from musical taste.

Additionally, several genres in our sample had limited representation (e.g., Latin, Gospel), which constrains the robustness of between-group comparisons. Future research would benefit from more balanced samples, a deeper analysis of sub-genres, and the inclusion of psychological constructs such as coping style, emotional regulation strategies, or trait affectivity. Longitudinal or experimental studies could further clarify whether genre selection influences mood over time or whether mood primarily drives genre choice.

5.1.2 RQ2: Does music listening frequency relate to anxiety or depression?

The present study examined whether the amount of time people spend listening to music is associated with levels of anxiety or depression. Our findings showed no significant relationship between listening frequency and anxiety, alongside a very weak but statistically significant positive correlation with depression. Although the association with depression reached statistical significance, the effect size was tiny, suggesting that listening time alone is not a meaningful predictor of depressive symptomatology.

These results align with literature indicating that music engagement and mental health are linked in complex and often bidirectional ways. Research reviewed by Gustavson et al. (2021) shows that individuals experiencing internalizing symptoms—including depression and anxiety—may increase their music use as a form of emotion regulation rather than music use causing worsening symptoms Gustavson *et al.*, 2021. Their scoping review highlights that people with elevated depressive or anxious tendencies often turn to music to cope with heightened emotional states, including for mood regulation, distraction, or emotional expression. This interpretation aligns with our data: the weak positive correlation may reflect that individuals with slightly higher depressive symptoms tend to make greater use of music in their daily lives, rather than suggesting that increased listening causes harm.

Support for this interpretation also emerges from Surana et al. (2020), who analyzed naturalistic music listening behaviors of users at varying levels of depression risk. Their findings show that individuals at higher risk for depression tend to engage more with music overall, often using it repetitively and gravitating toward emotionally congruent content (e.g., low-valence or sad music) as part of a ruminative coping style A. Surana *et al.*, 2020. Importantly, Surana et al. observed that greater musical engagement was associated with emotion-focused coping, consistent with our interpretation that increased listening time may be a response to, not a cause of, elevated depressive symptoms.

An unexpected outcome of our analysis was the complete absence of any meaningful association with anxiety. Although prior meta-analytic work (see, e.g., Panteleeva et al., 2017; Gustavson et al., 2021) suggests that music listening can reduce *state* anxiety in controlled settings, our results indicate that habitual listening frequency in daily life does not reliably relate to *trait* anxiety levels. This distinction between intervention-based effects and naturalistic habits highlights the importance of context. The frequency of listening may not reflect anxiety management strategies as clearly as variables related to *how* and *why* individuals choose specific music – factors emphasized in earlier correlational research on music-based emotion regulation strategies.

The broader implications of these findings suggest that simple metrics, such as listening time, are insufficient for understanding the nuanced relationship between music and mental health. Literature increasingly emphasizes that outcomes depend on users' motivations, emotional states, and engagement style (e.g., healthy vs. unhealthy regulation, rumination, social context). Our study reinforces this view by demonstrating that overall listening frequency carries very limited explanatory value for anxiety or depression.

Limitations. The present analysis includes several limitations. The cross-sectional nature of the dataset prevents causal inference, and the reliance on self-reported psychological measures may introduce reporting biases. Furthermore, listening time does not capture the qualitative elements of music engagement – such as emotional valence, lyrical content, or genre preferences – that other studies have shown to be more predictive of mental health outcomes.

Future research. Future work should integrate richer behavioral and content-based metrics, such as those analyzed by Surana et al. (2020), including acoustic features, emotional profiles of tracks, and temporal listening patterns. Longitudinal designs that capture fluctuations in mental state alongside detailed music-use behavior would help differentiate whether increased listening reflects coping, emotional reinforcement, or independent lifestyle factors. Finally, incorporating validated measures of music-based emotion regulation strategies may provide better explanatory models than listening duration alone.

5.1.3 RQ3: Can mental disorder risk be predicted from music preferences?

The results for RQ3 indicate that mental disorder risk, operationalized through a binary measure derived from self-reported psychological indicators, cannot be reliably predicted from music preferences alone. Both machine learning models—Logistic Regression and Random Forest—showed only marginally above-chance performance, with ROC AUC values between 0.56 and 0.58. These findings suggest that the relationship between musical behaviour and mental health is more complex than can be captured by demographic and preference-related variables alone. This perspective aligns with prior work showing that although music preferences correlate with personality and emotional tendencies, they rarely function as robust predictors of clinical mental health outcomes Greenberg, 2022; Vandewater *et al.*, 2020.

The weak performance of both models contrasts with popular assumptions and some preliminary findings suggesting that musical taste may reflect underlying emotional or psychological states. However, empirical studies indicate that while music can express or regulate mood, the link between musical behaviour and psychopathology remains indirect and context-dependent Miranda and Claes, 2007; K. Surana and Alluri, 2019. Our results are consistent with this view: certain variables—particularly age and daily listening duration—demonstrated meaningful importance scores, but they did not translate into strong predictive accuracy.

Several psychological mechanisms may help explain this outcome. Mental health conditions are influenced by multiple biopsychosocial factors—such as stress exposure, personality traits, social support, and coping strategies—that are not included in the dataset. Music preferences therefore represent only one limited aspect of an individual's emotional life. Research on emotional regulation through music shows that individuals use music differently depending on context, coping style, and emotional needs, meaning that the same genre can be associated with both adaptive and maladaptive regulation Saarikallio and Erkkilä, 2011. This variability makes genre-based prediction inherently unstable.

The feature importance results support broader psychological findings. Age emerged as the strongest predictor, with younger respondents more frequently classified as high-risk—a pattern consistent with documented increases in mental health difficulties among adolescents and young adults in recent years Twenge, 2019. Additionally, daily listening duration appeared as a relevant predictor, potentially reflecting the increased use of music as a coping strategy by individuals experiencing emotional distress. Yet these associations remain insufficient to produce clinically meaningful accuracy.

It is also noteworthy that favourite genres had minimal predictive relevance. While earlier results (e.g., RQ1) show genre-related variation in depression scores, these differences do not scale into reliable classification performance at the individual level. This distinction between group-level associations and individual-level prediction is well documented in psychological modelling, where large individual variability limits the predictive utility of simple behavioural indicators Yarkoni, 2020.

Several limitations should be acknowledged. The psychological indicators are self-reported and may not correspond directly to clinical diagnoses. The binary threshold used to define "high risk" simplifies the multidimensional nature of mental health conditions. Additionally, the dataset lacks variables known to be important for predicting mental health, such as personality traits, life stressors, or social support networks. Future research may benefit from integrating richer behavioural data, such as streaming patterns, acoustic features, or long-term listening histories, which have shown potential for mental health inference Montag *et al.*, 2021.

Overall, the findings for RQ3 indicate that although some behavioural and demographic variables show modest associations with mental-health outcomes, music preferences alone do not provide a sufficiently strong or stable basis for predicting mental disorder risk using standard machine learning methods.

5.1.4 RQ4: Does listening to music while working influence depression levels?

The analysis explored whether listening to music while working is associated with differences in self-reported depression levels. Our results showed no meaningful relationship: respondents who reported listening to music during work had similar depression scores to those who did not. This suggests that the presence or absence of music during work activities does not reliably predict depressive symptoms.

These findings stand in contrast to some earlier research showing that music listening in work contexts can influence emotional experience. Lesiuk (2005) found that employees who listened to music during computer-based tasks reported improved mood and lower negative affect throughout the workday Lesiuk, 2005. Her study suggests that music can create short-term emotional benefits in workplace settings, particularly by enhancing positive affect and reducing stress.

However, the absence of an association in our dataset may be explained by important differences in what is being measured. Lesiuk's study focused on *momentary* emotional states during specific tasks, whereas our measure of depression reflects a more stable, trait-like condition. Short-term mood improvements may not accumulate strongly enough to shift overall depression scores. This distinction between immediate affective responses and longer-term mental health outcomes is also highlighted in review work on music interventions.

Systematic reviews by Leubner and Hinterberger (2017) demonstrate that music can reduce depressive symptoms, but primarily in structured therapeutic or clinical contexts rather than in informal daily listening situations Leubner and Hinterberger, 2017. Their analysis emphasizes that intervention-based effects rely on intentional, repeated engagement with music tailored to treatment goals. In contrast, background music played while working may be too inconsistent or passive to influence depression in a measurable way.

Taken together, our results suggest that while music may temporarily improve mood during specific work tasks, it does not appear to have a broader or more stable effect on self-reported depression levels in everyday settings. Future research should examine how task type, listening context, and individual differences in emotional regulation influence whether music during work supports well-being or remains neutral in its psychological impact.

5.1.5 RQ5: Does age affect how people perceive the effects of music?

The goal of RQ5 was to examine whether people of different ages differ in how they perceive the effects of music on their well-being. Our results revealed a strikingly consistent pattern across all six age groups: the majority of respondents reported that music *improves* their mental state, whereas only a small minority endorsed *no effect* or *worsen*. Although the youngest (¡20) and oldest (60+)

participants showed slightly higher improvement scores, the overall differences between age groups were small, suggesting that the perceived benefits of music are widely shared across the lifespan.

This finding contrasts with experimental research showing that the emotional perception of music does change with age. Cohrdes et al. (2020) demonstrated that older adults generally perceive musical emotions—particularly negative, high-arousal expressions—as less intense than younger listeners Cohrdes *et al.*, 2020. Their results indicate that aging is associated with a shift toward perceiving music as more positive and less arousing, reflecting well-established age-related motivational tendencies toward maintaining positive affect. Based on this evidence, one might expect older adults in our sample to report stronger positive effects of music than younger adults.

However, our findings align more closely with lifespan research on *why* people listen to music rather than *how* they perceive musical emotions. Groarke and Hogan (2015) found that the functions of music listening differ substantially across age groups: younger adults rely on music for mood regulation, identity formation, and coping with stress, while older adults use music more for emotional reflection, social connection, and sustaining a positive affective tone Groarke and Hogan, 2015. Despite these differing functions, both younger and older listeners may arrive at the same overall conclusion—that music is beneficial for their well-being. This provides a plausible explanation for why our sample showed little variation in global evaluations of music's effects, even though underlying emotional processes may differ with age.

Together, these findings suggest that although age influences the *emotional perception* of music and the *functions* music serves, these differences do not necessarily translate into differences in broad self-reported judgments of music's impact on mental health. The perceived benefit of music appears to be a robust and widely shared experience across the lifespan.

Future work should employ more sensitive measures that capture subtle shifts in emotional intensity, regulation strategies, and listening motivations across age groups. Additionally, longitudinal research may help determine whether changes in the emotional processing of music with age eventually produce measurable differences in how its psychological effects are evaluated.

5.2 Evaluation of Hypotheses

5.2.1 H1: Genre and Depression Levels

Hypothesis H1 was only partially supported. While some melancholic genres showed higher average depression scores, the large variability within genres and several notable exceptions indicate that genre preference alone is not a strong or reliable predictor of depressive symptoms. The observed differences suggest that individual factors beyond genre choice play a substantial role in shaping depression levels.

5.2.2 H2: Listening Frequency and Emotional Well-Being

Hypothesis H2 was not supported. Listening frequency did not correlate with anxiety and showed only a very weak positive correlation with depression, contradicting the expectation that higher music listening frequency would relate to lower stress or anxiety. These findings imply that the amount of daily music exposure is not a meaningful indicator of emotional well-being on its own.

5.2.3 H3: Predictive Value of Musical Behaviour.

The third hypothesis stated that musical preferences and listening behaviour would contain enough information for a machine-learning model to predict mental-disorder risk above chance level. The obtained results offer only limited support for this assumption. Both the Logistic Regression and Random Forest classifiers achieved ROC AUC values slightly above random performance (approximately 0.56 to 0.58), indicating that the models capture a weak signal related to mental-health risk. However, the predictive strength is minimal, and the models do not provide robust or practically meaningful discrimination. Consequently, the evidence for H3 remains inconclusive: while the results do not contradict the hypothesis, they do not provide strong confirmation either.

5.3 Limitations

This study has several limitations that should be considered when interpreting the results. First, the dataset is cross-sectional, meaning that all variables were measured at a single point in time. As a result, the observed relationships cannot be interpreted causally. For example, although we found a weak association between music listening frequency and depressive symptoms, we cannot determine whether increased listening influences mood or whether individuals with higher depressive tendencies simply turn to music more often.

5.4 Threats to Validity

Several factors may limit the validity of the findings presented in this study. First, the dataset is based entirely on self-reported measures of musical behaviour and psychological symptoms. Although self-report is an appropriate method for capturing subjective experiences, it is also susceptible to biases such as inaccurate recall, social desirability, and individual differences in interpreting scale items. These factors may introduce noise into the data and reduce the precision of associations identified in the analysis.

Second, the dataset contains a relatively small number of respondents, and the distribution across age groups is highly uneven. Younger participants, particularly those under 30, represent the majority of the sample, whereas older age groups contain significantly fewer respondents. This imbalance limits the statistical power of comparisons involving age and may obscure meaningful differences in how older adults perceive the effects of music. As a result, conclusions drawn for RQ5 should be interpreted with caution, as the dataset may not adequately represent the full adult lifespan.

Finally, the cross-sectional nature of the data prevents causal interpretation. The observed relationships between music listening behaviours and psychological variables reflect associations at a single point in time and may be influenced by unmeasured confounding factors such as stress levels, personality traits, physical health, or recent life events. Without longitudinal or experimental data, it is not possible to determine the directionality of these relationships or assess potential short-term effects of music that may not be captured in global self-report metrics.

Together, these limitations suggest that while the present findings offer useful insights into patterns of music listening and perceived psychological effects, further research using larger, more balanced samples and more controlled study designs is needed to strengthen the validity of the conclusions.

6 Conclusion

This study examined the relationship between music listening habits and mental health indicators through five research questions addressing genre differences, listening frequency, predictive modelling, the role of music during work, and age-related perceptions of music's effects. By combining statistical testing, correlational analysis, and supervised machine learning, the study provides a multifaceted overview of how music-related behaviours relate to depression, anxiety, and subjective well-being.

The main contributions of this work lie in showing both the potential and the limits of using everyday music-listening patterns as indicators of mental health. First, although some music genres were linked to slightly higher depression scores, the large variation within each genre suggests that genre preference on its own is not a reliable sign of psychological distress. Second, listening frequency showed no meaningful association with anxiety and only a very weak relationship with depression, indicating that the amount of music people listen to is not a useful marker of emotional well-being. Third, the predictive modelling results suggest that musical preferences and listening behaviours contain only a small and uncertain signal related to mental-health risk. The models performed only slightly above random chance, and their ability to distinguish between risk levels was very limited. Together, these findings provide a balanced view of how much behavioural music data can contribute to understanding or anticipating mental-health outcomes.

Other analyses revealed that listening to music while working does not appear to influence depression levels in a measurable way, suggesting that the psychological impact of background music in everyday work contexts may be limited or highly dependent on situational factors not captured in this dataset. In contrast, perceptions of music's beneficial effects were remarkably stable across age groups, indicating that despite known age-related differences in emotional processing, individuals of all ages tend to view music as a positive influence on their mental well-being.

The findings of this study have several implications. They suggest that while music remains an important emotional resource for many people, its everyday usage patterns are not strongly linked to stable indicators of mental health. Future research should therefore incorporate richer behavioural measures—such as listening context, emotional intent, genre-specific acoustic features, and temporal patterns of music use—alongside longitudinal or experimental designs to better understand causal pathways. In addition, more representative and demographically balanced samples would allow for more robust inferences regarding age-related differences and subgroup-specific effects.

In summary, this study contributes to the growing body of work at the intersection of music psychology and mental health by clarifying how different aspects of music listening relate to depression, anxiety, and subjective well-being. While the associations observed here were generally weak, they provide a foundation for future investigations into how music interacts with emotional functioning and how it may be leveraged to support mental health in more targeted and intentional ways.

References

Cohrdes, Caroline, Wrzus, Cornelia, and Wald-Fuhrmann, Melanie (2020). "The sound of affect: Age differences in perceiving valence and arousal in music", *Psychology of Music*, Vol. 48.No. 6, pp. 864–881. DOI: 10.1177/0305735619828830.

Greenberg, David M. (2022). "The musical personality: How music shapes and reflects personality", *Psychology of Music*, Vol. 50.No. 3, pp. 743–760. DOI: 10.1177/03057356211017732.

Groarke, Jenny M. and Hogan, Michael J. (2015). "Enhancing wellbeing: An emerging model of the adaptive functions of music listening", *Psychology of Music*, DOI: 10.1177/0305735615591844.

Gustavson, Daniel E. *et al.*, (2021). "Mental health and music engagement: review, framework, and guidelines for future studies", *Translational Psychiatry*, Vol. 11.No. 370. DOI: 10.1038/s41398-021-01483-8.

Lesiuk, Teresa (2005). "The effect of music listening on work performance", *Psychology of Music*, Vol. 33.No. 2, pp. 173–191. DOI: 10.1177/0305735605050650.

Leubner, Daniel and Hinterberger, Thilo (2017). "Reviewing the Effectiveness of Music Interventions in Treating Depression", *Frontiers in Psychology*, Vol. 8, p. 1109. DOI: 10.3389/fpsyg.2017.01109.

Miranda, Dave and Claes, Michel (2007). "Music listening, coping, peer affiliation and depression in adolescence", *International Journal of Adolescence and Youth*, Vol. 13.No. 2, pp. 113–133.

Montag, Christian, Elhai, Jon D., and Hall, Brian J. (2021). "Towards using digital footprints in music streaming data to infer mental health", *Computers in Human Behavior*, Vol. 120, p. 106749. DOI: 10.1016/j.chb.2021.106749.

Saarikallio, Suvi and Erkkilä, Jaakko (2011). "Music as emotional self-regulation throughout adulthood", *Psychology of Music*, Vol. 35.No. 1, pp. 88–109. DOI: 10.1177/0305735610385459.

Surana, Aayush, Goyal, Yash, and Alluri, Vinoo (2020). "Static and Dynamic Measures of Active Music Listening as Indicators of Depression Risk", *arXiv preprint arXiv:2009.13685*, **available at:** https://arxiv.org/abs/2009.13685.

Surana, Karan and Alluri, Vinoo (2019). "Online Music Consumption: Characterizing Depression Using User Logs". *Proceedings of the 20th International Society for Music Information Retrieval Conference*, pp. 810–816.

Twenge, Jean M. (2019). "Age, cohort, and mental health trends among adolescents and young adults", *Journal of Abnormal Psychology*, Vol. 128.No. 2, pp. 110–124. DOI: 10.1037/abn0000410.

Vandewater, Sophia, He, Mu, and al., et (2020). "Predicting mental health problems using digital behaviour markers", *Current Opinion in Psychology*, Vol. 36, pp. 91–96. DOI: 10.1016/j.copsyc.2020.05.004.

Yarkoni, Tal (2020). "The generalizability crisis in psychology", *Behavioral and Brain Sciences*, Vol. 45, e1. DOI: 10.1017/S0140525X20001685.