Jurnal Aplikasi Sains Data
Journal homepage: https://jasid.upnjatim.ac.id/index.php/jasid

Analysis of Music-Listening Habits and Mood Changes During Daily Activities Using Multiple Linear Regression

Erida Dwi Cahyanti[1], Miftakhul Aulia Zuhrotun Nisa'[2], Alya Aziza Khairunnisa[3] , Andini Mei Nuraini[4] , Nadifa Alya Nur Atika Sari[5], and Dyah Linda Aulia[6].

[1,2,3] Department of Data Science, UPN "Veteran" Jawa Timur, Indonesia, [4] Department of Science Education, Universitas Surabaya, Indonesia, [5] Department of Fisheries Product Technology, Universitas Airlangga, Indonesia, [6] Department of Sports Science, Universitas Negeri Malang, Indonesia.

[1]25083010050@student.upnjatim.ac.id, [2]25083010066@student.upnjatim.ac.id, [3]25083010069@student.upnjatim.ac.id, [4]25030654111@mhs.unesa.ac.id, [5]nadifa.alya.nur-2025@fpk.unair.ac.id, [6] dyah.linda.2506216@students.um.ac.id.

ABSTRACT

Music is commonly used during daily activities for entertainment, concentration, and mood regulation. However, the association between music-listening habits and perceived mood changes may vary depending on individual preferences and listening contexts. Previous research has examined the association between music listening and perceived stress and mood in daily life [1]. This study aims to describe music-listening patterns and examine the association between selected listening habits and reported mood changes using Exploratory Data Analysis (EDA) and multiple linear regression. Survey data from 150 respondents were analyzed through data preprocessing, descriptive statistics, data visualization, Pearson correlation analysis, and regression modeling. The initial regression model included 11 predictors and yielded an R^2 of 0.235 and an adjusted R^2 of 0.174. Diagnostic tests indicated non-normal residuals and heteroskedasticity. Based on the initial model's statistical significance results, two predictors were retained for the final model: the percentage of daily activities perceived as more comfortable with music and the indicator for listening without a specific time. HC3 heteroskedasticity-consistent standard errors were applied to obtain more robust statistical inference under heteroskedasticity. The final model achieved an R^2 of 0.1503 and an adjusted R^2 of 0.1387, with both predictors statistically significant at the 0.05 level. These findings indicate that perceived comfort when listening to music and listening-time patterns are associated with reported mood changes. However, the model's limited explanatory power and the observational survey design mean that the findings do not establish causal relationships.

Keywords: exploratory data analysis, music-listening habits, mood changes, multiple linear regression, HC3 robust standard errors

I. INTRODUCTION

Music is part of everyday life and is frequently used while studying, working, travelling, exercising, and relaxing [2]. In addition to providing entertainment, music may be associated with emotional experiences, including changes in mood and perceived comfort during daily activities [3]. The role of music in emotional regulation has attracted attention because individuals may select different types of music depending on their emotional state, preferences, and activity context [4]. Research involving

* *Corresponding author.*
E-mail address: 25083010050@student.upnjatim.ac.id
http://doi.org/10.33005/jasid.v2i2.84

Indonesian university students has also examined music as a relaxation medium and its potential relationship with stress and mood [5], [6].

Previous studies have investigated the relationship between music and psychological conditions using different populations and research designs. Arcita et al. examined music as a relaxation medium for new students and reported that respondents perceived music as helpful in improving mood and providing calmness [5]. Abdullah et al. investigated the use of Javanese music to reduce stress among university students and reported a significant reduction in stress scores after the intervention [6]. These findings provide relevant background for studying music-related experiences. However, the present study focuses on respondents' reported music-listening habits during everyday activities and their perceived degree of mood change, rather than evaluating a controlled music intervention [7].

Exploratory Data Analysis (EDA) provides a systematic approach to understanding data through descriptive statistics, graphical visualization, and the examination of relationships between variables [8]. EDA can help identify the dominant characteristics of respondents, the distribution of listening habits, and possible associations among numerical variables before statistical modeling is performed [9]. Previous Indonesian research has demonstrated the use of EDA and visualization to explore datasets and communicate patterns in a more interpretable form [10]. Combining EDA with regression analysis allows the present study to describe the collected survey data and examine whether selected listening habits are statistically associated with reported mood changes.

This study aims to describe music-listening habits among 150 respondents and examine the relationship between selected listening characteristics and perceived mood changes through multiple linear regression [11]. The analysis considers several factors, including listening frequency, perceived comfort when accompanied by music, listening duration, listening-time patterns, and selected categorical characteristics. Regression diagnostics are also conducted to evaluate the model and determine whether robust standard errors are needed [12]. The findings are intended to provide a descriptive and statistical overview of music-listening experiences in daily activities, while recognizing that survey-based associations do not establish causal relationships.

II. METHODOLOGY

A. Research Design

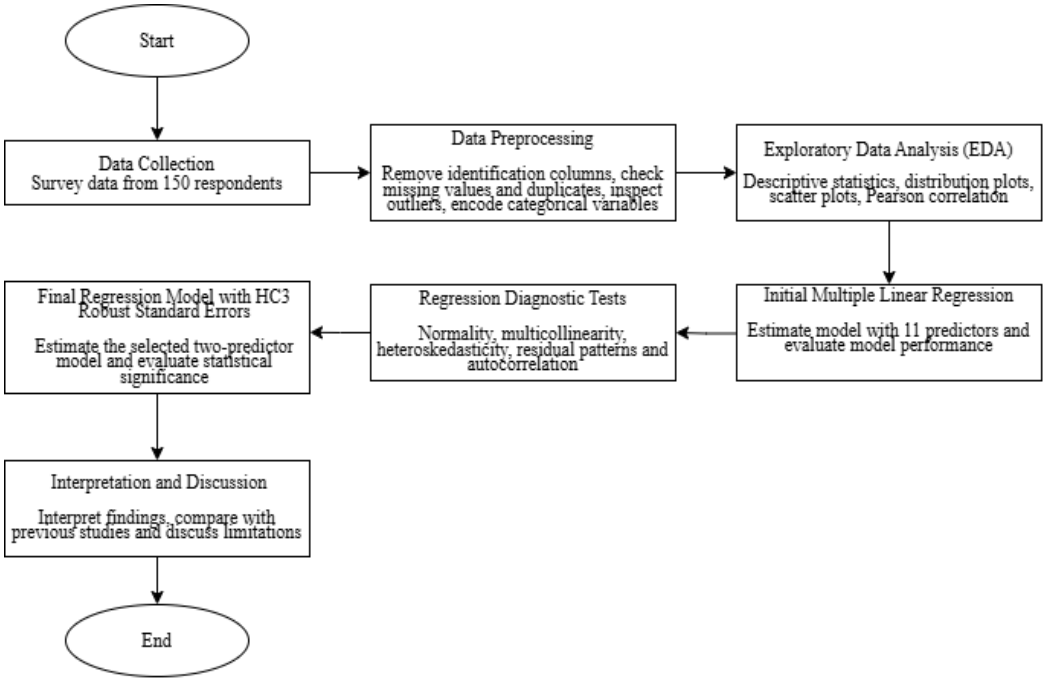

Fig. 1. Research Flowchart

The research workflow is illustrated in Fig. 1. The study began with data collection from 150 respondents regarding their music-listening habits and perceived mood changes during daily activities. The collected data were then preprocessed by removing personal identification columns, checking missing values and duplicate records, inspecting potential outliers using the Interquartile Range (IQR) method, and encoding categorical variables using one-hot encoding [13]. Subsequently, Exploratory Data Analysis (EDA) was conducted through descriptive statistics, data visualization, and Pearson correlation analysis to identify patterns and relationships among the variables [14]. An initial multiple linear regression model with 11 predictors was then estimated to examine the association between music-listening habits and reported mood changes. Regression diagnostic tests were performed to evaluate residual normality, multicollinearity, heteroskedasticity, and residual autocorrelation [15]. Because the diagnostic results indicated heteroskedasticity, the final model was estimated using HC3 heteroskedasticity-consistent standard errors and retained two predictors according to the selection procedure implemented in the notebook. Finally, the regression results were interpreted by examining the coefficient estimates, statistical significance, model performance, relevant previous studies, and study limitations.

B. Data Collection and Variables

The dataset consisted of 150 survey responses concerning music-listening habits during daily activities. The dependent variable was the reported degree of mood change after listening to music while performing activities, measured on a scale from 0 to 10, where higher values represented greater perceived mood change. The independent variables included the percentage of daily activities perceived as more comfortable with music, the difficulty of performing activities without music, music selection based on mood, listening duration, listening frequency per week, estimated mood-recovery time, gender, music-streaming platform, listening-time pattern, and the main purpose of listening to music. The variables were selected based on the questions available in the survey which were collected via google forms using convenience sampling targeting students and the predictor definitions used in the regression notebook.

C. Data Cleaning and Categorical Encoding

Data preprocessing began by removing columns containing personal identification information and other fields that were not required for the analysis [16]. The remaining dataset contained 150 observations and 19 columns. Missing-value checks showed no missing values in the retained variables, and the duplicate-record check identified no duplicate rows. Potential outliers were inspected using the Interquartile Range method. Because an observation flagged by IQR is not automatically an error, these values were treated as candidates for inspection rather than automatically removed. Categorical variables were subsequently converted into numerical dummy variables using one-hot encoding with the first category omitted as a reference category. The resulting encoded dataset contained 150 observations and 97 columns.

D. Exploratory Data Analysis

EDA was performed to summarize respondents' characteristics and identify patterns in their music-listening habits. Descriptive statistics were calculated for numerical variables, including means, standard deviations, minimum and maximum values, and quartiles [17]. Frequency distributions were used to summarize categorical variables such as gender, streaming platform, listening time, and listening purpose. Histograms and boxplots were used to inspect numerical distributions, while bar charts described categorical distributions [18]. Pearson correlation analysis was also performed to examine the direction and strength of linear associations among selected numerical variables, particularly between perceived comfort with music and reported mood change [19].

E. Multiple Linear Regression

Multiple linear regression was used to examine the association between music-listening characteristics and reported mood change. The general form of the multiple linear regression model is expressed as follows:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon$$

Where Y_i is the reported mood-change score of the i-th respondent, β_0 is the intercept, β_j is the regression coefficient of the j-th predictor, X_{ij} is the value of the j-th predictor for respondent i, k is the total number of predictors, n is the number of respondents (n = 150), and ε_i is the error term. For consistency, the following notation is used throughout this paper: Y denotes the reported mood-change score (scale 0–10); X_1 denotes the percentage of daily activities perceived as more comfortable with music (%); and X_{10} denotes the dummy variable for having no specific listening time (1 = no specific listening time, 0 = otherwise). The index 10 follows the position of this dummy variable among the 11 predictors of the initial model.

The initial model included 11 predictors selected from the encoded dataset. Model performance was evaluated using the coefficient of determination (R^2), adjusted R^2, the overall F-statistic, and individual coefficient tests. To achieve a more parsimonious specification, a predictor selection procedure was implemented based on statistical significance in the initial model: predictors with p-values below 0.05 were retained for the final model. The final model retained the percentage of activities perceived as more comfortable with music and the indicator representing respondents who reported having no specific listening time. The selection procedure was exploratory and should not be interpreted as a guarantee that these are the only relevant predictors.

F. Regression Diagnostic Tests

Regression diagnostics were conducted on the initial model using residual plots, the Shapiro–Wilk test, the Variance Inflation Factor (VIF), the Breusch–Pagan test, and the Durbin–Watson statistic. The Shapiro–Wilk test indicated non-normal residuals ($p=0.0017$), while the Breusch–Pagan test indicated heteroskedasticity ($p<0.001$). VIF values ranged from 1.034 to 3.846, indicating no severe multicollinearity, and the Durbin–Watson statistic was 2.0801, showing no strong indication of residual autocorrelation. HC3 heteroskedasticity-consistent standard errors were applied to the final model for three reasons. First, the Breusch–Pagan test indicated heteroskedasticity ($p < 0.001$), under which conventional OLS standard errors may be biased and make t-tests unreliable. Second, among the HC estimators, HC3 adjusts for observation leverage and has better finite-sample properties than HC0–HC2, which suits a moderate sample of n = 150 [4]. Third, HC3 changes only the standard errors, not the coefficient estimates, so the interpretation of the coefficients stays the same. HC3 does not correct residual non-normality or model misspecification [4].

III. RESULT AND DISCUSSION

A. Respondent Characteristics and Descriptive Statistics

The dataset contained 150 respondents. The descriptive results showed that respondents listened to music frequently during their daily activities. The mean listening frequency was 5.83 days per week, while the mean listening duration during activities was 3.63 hours per day. The average reported mood-change score was 7.87 on a scale from 0 to 10. The average percentage of daily activities perceived as more comfortable when accompanied by music was 78.41%. These descriptive findings indicate that respondents generally reported frequent music use and relatively high perceived mood changes.

Table I. Descriptive Statistics Of Selected Numerical Variables

Variable	Mean	Standard deviation	Minimum	Maximum
Age (years)	19.55	1.25	17	23
Music-listening frequency based on mood (scale)	4.29	0.89	1	5
Difficulty performing activities without music (scale)	2.67	1.05	1	5
Listening frequency (days/week)	5.83	1.56	0	7
Number of songs listened to per day	14.53	8.73	2	36
Activities perceived as more comfortable with music (%)	78.41	17.12	10	100
Music-listening habit (years)	4.75	2.98	0	12
Reported mood change (scale 0–10)	7.87	1.34	3	10
Listening duration during activities (hours/day)	3.63	2.51	1	17
Estimated mood-recovery time (minutes)	28.28	26.83	1	180

Note: Values are taken from the descriptive-statistics output of the analysis notebook. In the regression equations, the reported mood change (scale 0–10) is denoted Y and the percentage of activities perceived as more comfortable with music is denoted X_1. The reported mood-change variable measures respondents' perceived change rather than an experimentally measured difference between pre-listening and post-listening mood.

Gender distribution showed that 100 respondents (66.67%) were female and 50 respondents (33.33%) were male. Spotify was the most frequently reported primary streaming platform, selected by 107 respondents (71.33%), followed by YouTube or YouTube Music with 34 respondents (22.67%). In terms of listening time, 85 respondents (56.67%) reported having no specific listening time, while 53 respondents (35.33%) most frequently listened at night. Regarding listening purpose, 97 respondents (64.67%) selected mood improvement or motivation as their main purpose, followed by 36 respondents (24.00%) who selected reducing boredom, traffic-related monotony, or loneliness.

Table II. Distribution Of Selected Categorical Variable

Variable	Category	Frequency	Percentage %
Gender	Female	100	66.67
Gender	Male	50	33.33
Primary platform	Spotify	107	71.33
Primary platform	Youtube / Youtube Music	34	22.67
Primary platform	Apple Music	7	4.67
Primary platform	OuterTune	1	0.67
Primary platform	MP3 download	1	0.67
Listening time	No specific time	85	56.67
Listening time	Night	53	35.33
Listening time	Morning	7	4.67
Listening time	Afternoon	3	2.00
Listening time	Evening	2	1.33
Main listening purpose	Mood improvement / motivation	97	64.67
Main listening purpose	Reducing boredom / monotony / loneliness	36	24.00
Main listening purpose	Avoiding environmental noise	9	6.00
Main listening purpose	Improving focus / concentration	8	5.33

Note: Percentages are calculated using 150 respondents. Percentages within each variable may not sum to exactly 100% because of rounding.

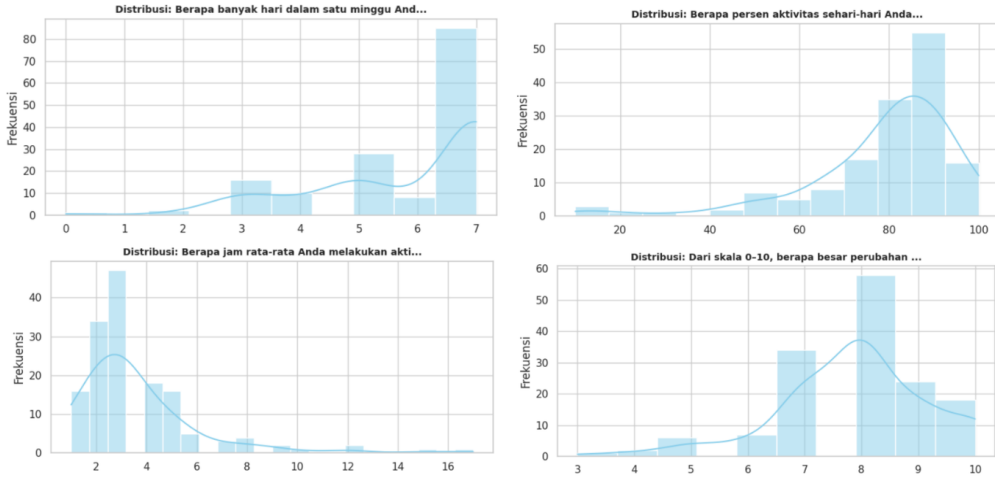

Fig. 2. Distributions of selected numerical variables: (a) listening frequency (days/week), (b) activities perceived as more comfortable with music (%), (c) average listening duration during activities (hours/day), and (d) reported mood change (scale 0–10). Curves indicate kernel density estimates.

Fig. 2 presents the distributions of selected numerical variables, complementing the descriptive statistics in Table I. The mood-change scores were concentrated toward the upper end of the 0–10 scale, with a mean of 7.87 and a median of 8. This indicates that many respondents reported relatively substantial mood changes. However, these descriptive results alone do not establish whether music directly caused the reported changes.

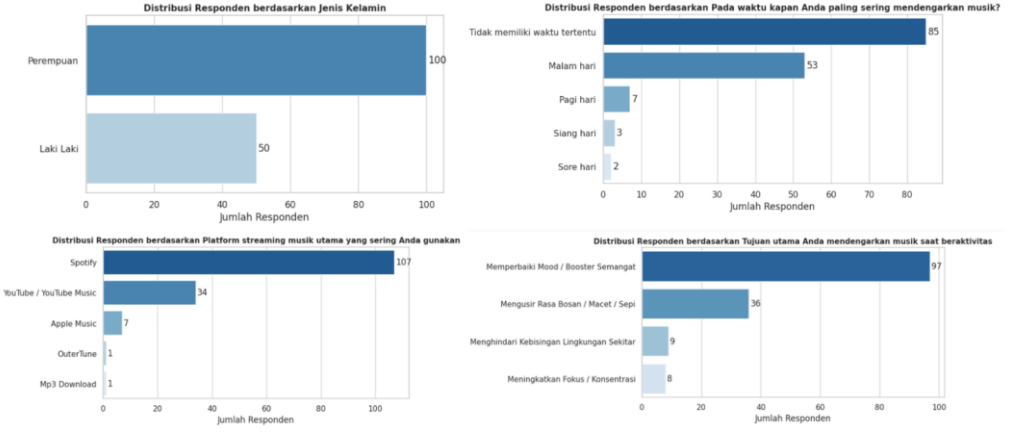

Fig. 3. Respondent distribution by (a) gender, (b) listening-time pattern, (c) primary music-streaming platform, and (d) main listening purpose.

The categorical distributions in Fig. 3 show that Spotify was the dominant platform and that listening was commonly reported as occurring without a specific time or at night. Mood improvement or motivation was the most frequently reported listening purpose. These patterns describe the respondents in this dataset and should not automatically be generalized to all university students or music listeners.

B. Exploratory Data Analysis and Correlation

Pearson correlation analysis was conducted to examine linear relationships among selected numerical variables. The correlation between the percentage of activities perceived as more comfortable with music and reported mood change was approximately 0.355. This represents a positive association of modest magnitude: respondents reporting greater comfort when listening to music also tended to report greater mood changes. The correlation between difficulty performing activities without music and mood change was approximately 0.260, while the correlation between listening frequency per week and mood change was approximately 0.161. These relationships were weaker than the association involving perceived comfort.

Fig. 4. Pearson correlation heatmap of nine selected numerical variables related to music-listening habits and mood change.

The heatmap helps compare the direction and relative strength of associations across variables. The positive correlation between perceived comfort and mood change is consistent with the descriptive pattern observed in the dataset. However, correlation does not demonstrate causation, and the magnitude of a correlation does not account for other variables that may influence respondents' experiences. Therefore, the correlation results are interpreted as preliminary evidence of association rather than proof of a direct effect.

C. Initial Multiple Linear Regression Model

The initial multiple linear regression model included 11 predictors. The model produced an R^2 of 0.235 and an adjusted R^2 of 0.174. The overall F-statistic was 3.863, with a p-value of 7.19×10^{-5}. Since the p-value was below 0.05, the null hypothesis that all predictor coefficients were jointly equal to zero was rejected. This indicates that the initial model was statistically significant overall. Nevertheless, the adjusted R^2 indicates that the model explained approximately 17.4% of the variation in reported mood change after accounting for the number of predictors.

Table III. Summary Of The Initial Regression Model

Metric	**Result**
Number of observations	150

Number of predictors	11
R^2	0.235
Adjusted R^2	0.174
F-statistic	3.863
Probability of F-statistic	7.19×10^{-5}

The initial coefficient results indicated that the percentage of activities perceived as more comfortable with music was positively associated with mood change ($\beta = 0.0243$, $p < 0.001$). The indicator for listening without a specific time was negatively associated with mood change ($\beta = -0.4244$, $p = 0.047$). The other predictors did not reach the 0.05 significance level in this model. Based on the selection procedure implemented in the notebook, the two significant predictors were retained for further analysis. This model-selection approach is exploratory, and the resulting coefficients should be interpreted with consideration of the selection process.

After evaluating the initial multiple linear regression model, diagnostic tests were conducted to assess whether the model satisfied the relevant regression assumptions. These tests examined residual normality, multicollinearity, heteroskedasticity, and residual autocorrelation. The diagnostic results were subsequently used to determine the appropriate approach for estimating the final regression model.

D. Regression Diagnostic Tests

The diagnostic results indicated that not all assumptions of the initial regression model were satisfied. The Shapiro–Wilk test produced a statistic of 0.9688 and a p-value of 0.0017, providing evidence against the null hypothesis of normally distributed residuals. The Breusch–Pagan test produced a statistic of 40.6853 with a p-value reported as 0.0000 after rounding, indicating evidence of heteroskedasticity. In contrast, the Variance Inflation Factor values ranged from approximately 1.034 to 3.846 among the predictors, which did not indicate severe multicollinearity under the commonly used threshold of 5. The Durbin–Watson statistic was 2.0801, which was close to 2 and did not suggest a strong first-order autocorrelation pattern based on the simple screening criterion used in the notebook.

Table IV. Diagnostic Result For The Initial Regression Model

Diagnostic	**Statistic**	**p-value / range**	**Interpretation**
Shapiro–Wilk normality test	0.9688	0.0017	Evidence against residual normality
Breusch–Pagan test	40.6853	0.0000*	Evidence of heteroskedasticity
VIF	-	1.034–3.846	No severe multicollinearity indicated
Durbin–Watson	2.0801	-	Close to 2; no strong autocorrelation indicated by the screening criterion

Note: The Breusch–Pagan p-value is displayed as 0.0000 because of rounding; it should be reported as $p < 0.001$, not as an exactly zero probability. The Durbin–Watson interpretation is a descriptive screening result rather than a formal test using specified critical bounds.

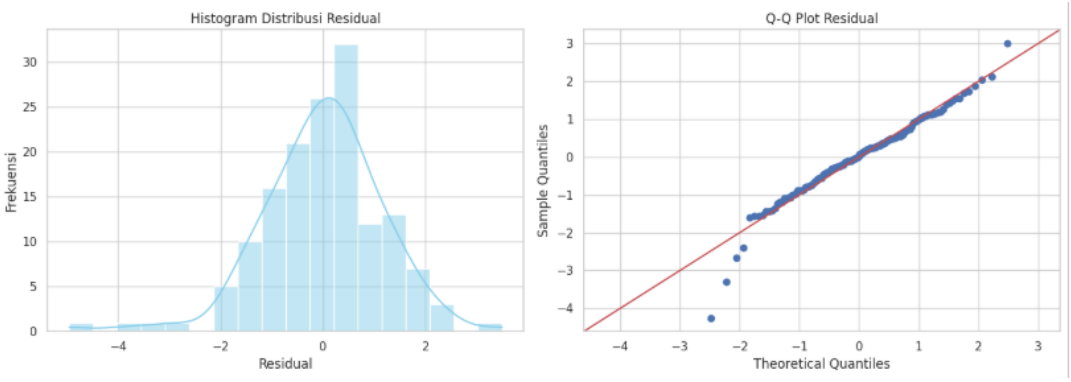

Fig. 5. (a) Histogram with kernel density curve and (b) normal Q–Q plot of the initial regression residuals.

The normal Q–Q plot should be examined together with the Shapiro–Wilk test. Deviations of the residual points from the reference line indicate that the residual distribution does not perfectly follow a normal distribution. This is consistent with the statistical test, which returned a p-value below 0.05. The finding does not automatically invalidate every coefficient estimate, but it should be considered when interpreting conventional regression inference, particularly in combination with the evidence of heteroskedasticity.

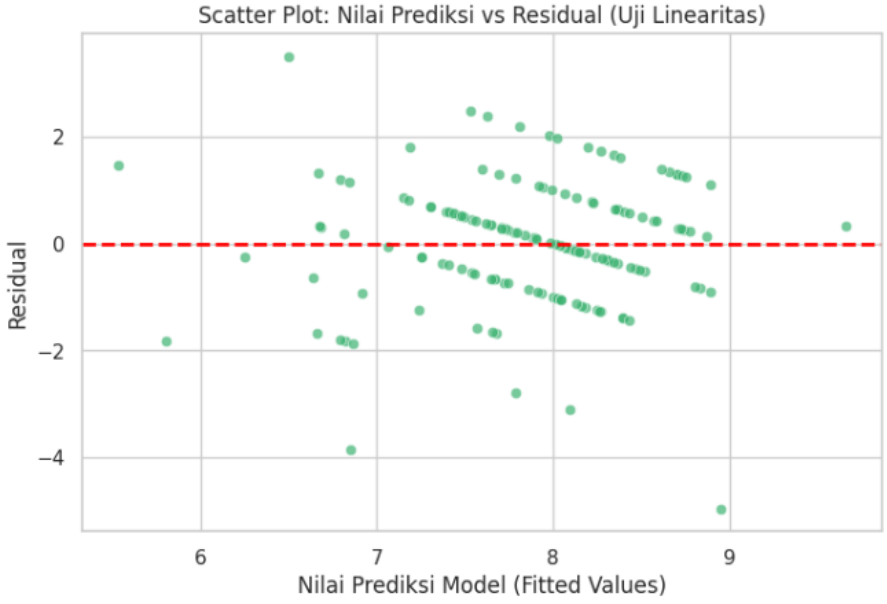

Fig. 6. Residuals versus fitted values for the initial regression model.

The residual-versus-fitted plot is used to inspect whether residuals show systematic patterns or changing dispersion across fitted values. Although the notebook's interpretation describes the residuals as scattered around zero without an obvious curved pattern, the Breusch–Pagan test indicates that the residual variance is not constant. Consequently, the graphical inspection should not be used alone to conclude that homoskedasticity holds. To make statistical inference more robust to this issue, the final model was estimated using HC3 robust standard errors.

The diagnostic results indicated evidence of residual non-normality and heteroskedasticity in the initial regression model. In particular, the Breusch–Pagan test indicated that the residual variance was

not constant. Therefore, the final regression model was estimated using HC3 heteroskedasticity-consistent standard errors to improve the robustness of statistical inference under heteroskedasticity. The final model and its coefficient estimates are presented in the following subsection.

E. Final Regression Model with HC3 Robust Standard Errors

The final regression model retained two predictors: the percentage of daily activities perceived as more comfortable with music and the indicator for listening without a specific time. HC3 heteroskedasticity-consistent standard errors were applied because the initial model showed evidence of heteroskedasticity. The final model produced an R^2 of 0.1503 and an adjusted R^2 of 0.1387. The overall F-statistic was approximately 6.1626, with a p-value of 0.002691. This indicates that the final model was statistically significant overall at the 0.05 significance level.

Table V. Final Regression Model With Hc3 Robust Standard Errors

Predictor	**Coefficient**	**Robust standard error**	**z-statistic**	**p-value**
Constant	5.9596	0.700	8.515	<0.001
Activities perceived as more comfortable with music (%)	0.0275	0.009	3.158	0.002
No specific listening time (dummy indicator)	-0.4220	0.201	-2.098	0.036

Note: The reported test statistics and p-values follow the HC3 model summary in the notebook. Dependent variable: reported mood change (Y, scale 0–10); n = 150; R^2 = 0.1503; adjusted R^2 = 0.1387; F = 6.1626 (p = 0.002691). The categorical coefficient is interpreted relative to the omitted reference category used during dummy encoding.

The estimated final regression equation, based on Equation (1), was:

$$\hat{Y} = 5.9596 + 0.0275X_1 - 0.4220X_{10}$$

Where Y represents the predicted mood-change score, X_1 represents the percentage of daily activities perceived as more comfortable with music, and X_{10} represents the dummy indicator for having no specific listening time (1 = no specific listening time, 0 = otherwise). As a consistency check, substituting the sample means (X_1 = 78.41 and X_{10} = 0.567, the proportion of respondents with no specific listening time) gives $\hat{Y} \approx 7.88$, which agrees with the observed mean mood-change score of 7.87 (Table I).

The coefficient of X_1 was positive and statistically significant ($\beta = 0.0275$, $p = 0.002$). Holding the other predictor constant, a one-percentage-point increase in activities perceived as more comfortable with music was associated with an estimated increase of 0.0275 points in the reported mood-change score. For example, a 10-percentage-point difference corresponds to an estimated difference of 0.275 points, according to this model. This is an association within the fitted model and should not be interpreted as a causal effect.

The coefficient for the no-specific-listening-time indicator was negative and statistically significant ($\beta = -0.4220$, $p = 0.036$). This means that respondents in this category had an estimated mood-change score 0.422 points lower than respondents in the omitted reference category, holding the other predictor constant. The comparison is determined by the dummy encoding and should not be interpreted as evidence that listening without a specific time causes lower mood changes.

The final model's R^2 of 0.1503 indicates that the two predictors accounted for approximately 15.03% of the observed variation in mood-change scores in the sample. The adjusted R^2 was 0.1387, or approximately 13.87%. Thus, most of the variation remained unexplained by the final model and may be related to other individual, contextual, or behavioral factors not included in the analysis. The HC3 adjustment improves the robustness of the estimated standard errors under heteroskedasticity, but it does

not eliminate all limitations associated with model specification, variable selection, or the non-normality of residuals [20].

The R^2 value should be interpreted in light of the purpose of this study. An R^2 of 0.1503 means that X_1 and X_{10} jointly account for about 15% of the variance in reported mood change. It is not a measure of prediction accuracy for an individual respondent, and the remaining variance should not be regarded as random noise. Reported mood change is a subjective outcome that is likely to depend on factors not measured in this survey, such as baseline mood, music genre and tempo, the type of activity being performed, and personal preference. Low to moderate R^2 values are common in survey-based studies of human perception and behavior; accordingly, the model is better suited to describing associations than to predicting mood change for individual listeners.

To place the coefficients in context, the standard deviation of reported mood change was 1.34 points and that of perceived comfort was 17.12 percentage points (Table I). A one-standard-deviation increase in X_1 (about 17 percentage points) corresponds to an estimated increase of $0.0275 \times 17.12 \approx 0.47$ points in mood change, or roughly 0.35 standard deviations of the outcome. The coefficient of X_{10} (–0.4220) corresponds to roughly 0.31 standard deviations. Both effects are modest, which is consistent with the limited R^2 and with the correlation of approximately 0.355 reported in Section III.B.

The decline in explanatory power from the initial model ($R^2 = 0.235$; adjusted $R^2 = 0.174$) to the final model ($R^2 = 0.1503$; adjusted $R^2 = 0.1387$) should also be noted. Because the adjusted R^2 is lower in the final model, the nine predictors removed during selection jointly contributed some explanatory information, even though none was individually significant at the 0.05 level. The final model is therefore more parsimonious and easier to interpret, but it does not explain more variance than the initial model.

F. Discussion and Study Limitations

The findings show that perceived comfort when listening to music had a positive association with reported mood change in the final model. This result is consistent with prior Indonesian research describing music as a potential relaxation medium and examining its relationship with stress among students [5], [6]. However, the studies differ in design, measures, and objectives. The current study uses observational survey responses and evaluates perceived mood change during everyday activities, whereas an intervention study examines outcomes under a more controlled listening condition. Therefore, the present findings should be viewed as complementary evidence about respondents' music-listening experiences rather than a direct replication of previous intervention results.

Implications of perceived comfort. The positive association between X_1 and mood change suggests that respondents who find a larger share of their daily activities more comfortable with music also tend to report a greater mood change. One plausible interpretation is that mood change is greater when music fits the activity and the listener's needs, rather than when music is simply played more often or for longer; listening frequency ($r \approx 0.161$) and listening duration were not statistically significant in the initial model. The direction of the relationship cannot be established from these data, however. Respondents who experience stronger mood changes may also be more inclined to rate music as comfortable, and because both measures are self-reported they may reflect a shared general attitude toward music.

Implications of listening-time pattern. More than half of the respondents (56.67%) reported no specific listening time, and this group reported mood changes about 0.42 points lower than the reference category after accounting for perceived comfort. A possible explanation, which this study did not test, is that unscheduled listening functions more as background accompaniment and is less deliberately tied to mood regulation than listening at a particular time of day. The difference may equally reflect other characteristics that differ between the groups. Because this coefficient is only marginally significant ($p = 0.036$) and was retained through an exploratory selection procedure, it should be treated as tentative.

Taken together, the results suggest that, for people who want music to support their mood during daily activities, how well the music suits the activity may matter more than the amount of listening. Given the limited explanatory power of the model, this should be regarded as a hypothesis for further investigation rather than a recommendation.

Several limitations should be considered. First, the data were based on self-reported survey responses and may be influenced by recall, interpretation, and individual differences. Second, the sample may not represent all students or music listeners. Third, the final model explained a limited

proportion of variation in reported mood change ($R^2 = 0.1503$), indicating that other relevant factors were not included. Fourth, the predictor selection procedure relied on statistical significance in the initial model, which can affect the stability of estimates and the interpretation of subsequent significance tests. Finally, the diagnostic results showed residual non-normality and heteroskedasticity in the initial model. Although HC3 robust standard errors were used for the final model, additional analyses with theoretically justified predictors, alternative specifications, and validation on independent data would be useful before making broader claims. Future studies could also measure mood before and after listening, and include variables such as music genre, tempo, baseline mood, and the type of activity, as well as possible interaction effects (for example, between perceived comfort and listening-time pattern), using longitudinal or experimental designs.

IV. CONCLUSION

This study examined music-listening habits and their association with reported mood changes among 150 survey respondents using EDA and multiple linear regression. The descriptive analysis showed that respondents listened to music on an average of 5.83 days per week, reported a mean listening duration of 3.63 hours per day during activities, and had an average mood-change score of 7.87 out of 10. Spotify was the most frequently reported primary streaming platform, while mood improvement or motivation was the most common listening purpose. Pearson correlation analysis showed a positive association between perceived comfort with music and reported mood change.

The initial regression model was statistically significant overall but showed evidence of residual non-normality and heteroskedasticity. The final model, estimated with HC3 robust standard errors, retained two statistically significant predictors: the percentage of activities perceived as more comfortable with music and the indicator for listening without a specific time. The final model had an R^2 of 0.1503 and an adjusted R^2 of 0.1387, indicating limited explanatory power. In conclusion, the results suggest that perceived comfort and listening-time patterns are associated with respondents' reported mood changes, but they do not establish causation. Further research using broader samples and additional theoretically relevant variables is needed to better understand the relationship between music-listening habits and mood.

REFERENCES

[1] Feneberg, A. C., Stijovic, A., Forbes, P. A. G., Lamm, C., Piperno, G., Pronizius, E., Silani, G., & Nater, U. M. (2023). Perceptions of Stress and Mood Associated With Listening to Music in Daily Life During the COVID-19 Lockdown. JAMA network open, 6(1), e2250382. https://doi.org/10.1001/jamanetworkopen.

[2] Akbas, Ridvan & Vuoskoski, Jonna. (2026). Functions of Everyday Music Listening. DOI: 10.13140/RG.2.2.14894.27209.

[3] Bavarava, Abhi. (2024). The Impact of Music on Mood and Emotion: A Comprehensive Analysis. Journal of Advanced Research in Journalism & Mass Communication. 11. 12-21. DOI: 10.24321/2395.3810.202402.

[4] Jiang, Jiuhua. (2026). Understanding Music's Role in Emotion Regulation: An Integrated Music-Person-Context Framework. Lecture Notes in Education Psychology and Public Media. 120. 1-10. DOI: 10.54254/2753-7048/2026.ZJU31986.

[5] G. N. Arcita, A. N. F. Adisurya, and B. Setiaji, “The Influence of Music as a Relaxation for New Students,” Jurnal Penelitian Musik, vol. 6, no. 1, 2025, doi: 10.21009/JurnalPenelitianMusik.61.05.

[6] J. Abdullah, C. Suryaningrum, and S. Prasetyaningrum, “Musik Jawa untuk Mereduksi Stres pada Mahasiswa,” JIP (Jurnal Intervensi Psikologi), vol. 5, no. 2, pp. 239–251, 2013, doi: 10.20885/intervensipsikologi.vol5.iss2.art6.

[7] Krause, Amanda & North, A. & Hewitt, Lauren. (2013). Music-listening in everyday life: Devices and choice. Psychology of Music. 43. 155-170. 10.1177/0305735613496860. DOI: 10.1177/0305735613496860.

[8] Dhummad, Sukhdevsinh. (2025). The Imperative of Exploratory Data Analysis in Machine Learning. Scholars Journal of Engineering and Technology. 13. 30-44. DOI: 10.36347/sjet.2025.v13i01.005.

[9] Jebb, Andrew & Parrigon, Scott & Woo, Sang Eun. (2016). Exploratory data analysis as a foundation of inductive research. Human Resource Management Review. DOI: 27. 10.1016/j.hrmr.2016.08.003.
[10] I. N. Rizki, D. Prayoga, M. L. Puspita, and M. Q. Huda, "Implementasi Exploratory Data Analysis untuk Analisis dan Visualisasi Data Penderita Stroke Kalimantan Selatan Menggunakan Platform Tableau," Jurnal Informatika dan Teknik Elektro Terapan, vol. 12, no. 1, 2024, doi: 10.23960/jitet.v12i1.3856.
[11] Putri, Nuriyana & Rofiqoh, Rofiqoh & Wahyudin, Wahyudin. (2026). Correlation Between Listening to Music, Emotion Regulation, and Learning Focus Among English Education Study Program Students at Tadulako University. Journal of English Language and Education. 11. 1097-1104. DOI: 10.31004/jele.v11i4.3038.
[12] Abdel Bary, Mona. (2017). Robust regression diagnostic for detecting and solving multicollinearity and outlier problems: applied study by using financial data. Applied Mathematical Sciences. 11. 601-622. DOI: 10.12988/ams.2017.7253.
[13] Jovita, & Susanto, Susanto. (2026). Perbandingan Metode IQR vs Isolation Forest dalam Deteksi Outlier pada Data Klaim Asuransi di PT Askrindo Cabang Semarang. Jurnal Transformatika. 23. 268-277. DOI: 10.26623/transformatika.v23i2.13726.
[14] Isa, Indra Griha Tofik & Zulkarnaini, Zulkarnaini & Novianti, Leni & Agustri, Suzan. (2023). Exploratory Data Analysis (EDA) dalam Dataset Penerimaan Mahasiswa Baru Universitas XYZ Palembang. Smart Comp: Jurnalnya Orang Pintar Komputer. 12. 600-609. 10.30591/smartcomp.v12i3.4125.
[15] Ni, Haoran. (2026). Statistical Diagnostics and Reliability Evaluation of Linear Regression Models. Theoretical and Natural Science. 186. None-None. DOI: 10.54254/2753-8818/2026.37301.
[16] Schneijderberg, Christian & Wieczorek, Oliver & Steinhardt, Isabel. (2026). Data preprocessing. DOI: 10.4324/9781003496397-16.
[17] Kaur, Parampreet & Stoltzfus, Jill & Yellapu, Vikas. (2018). Descriptive statistics. International Journal of Academic Medicine. 4. 60. DOI: 10.4103/IJAM.IJAM_7_18.
[18] Rodrigues, Ariane & Barbosa, Gabriel & Lopes, Hélio & Barbosa, Simone. (2019). Comparing the Effectiveness of Visualizations of Different Data Distributions. 84-91. DOI: 10.1109/SIBGRAPI.2019.00020.
[19] Mcferran, Katrina & Garrido, Sandra & O'Grady, Lucy & Grocke, Denise & Sawyer, Susan. (2014). Examining the relationship between self-reported mood management and music preferences of Australian teenagers. Nordic Journal of Music Therapy. 24. 187-203. DOI: 10.1080/08098131.2014.908942.
[20] J. G. MacKinnon and H. White, "Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties," Journal of Econometrics, vol. 29, no. 3, pp. 305–325, 1985, doi: 10.1016/0304-4076(85)90158-7.