EXPLORING THE RELATIONSHIP BETWEEN MUSIC, CONCENTRATION, AND TASK PERFORMANCE

Project work submitted in partial fulfilment of the requirements for the degree

of

Master of Science

In

Data Science & Analytics

submitted by

Akanksha Kumari (Roll No. 23478123001)

Priti Saha (Roll No.23478123009)

Subarna Tapadar (Roll No.23478123020)

Sudesna Debnath (Roll No. 23478123024)

Under the guidance of

Dr. Sanjay Goswami

M.Sc. in Data Science & Analytics [2023-2025]

NSHM Institute of Computing and Analytics [NICA]

NSHM Institute of Computing and Analytics [NICA]

Master of Science in Data Science & Analytics

NSHM
KNOWLEDGE
CAMPUS

CERTIFICATE

This is to certify that *Ms. Akanksha Kumari (Roll No. 23478123001), Ms. Priti Saha (Roll No.23478123009), Ms. Subarna Tapadar (Roll No.23478123020)* and *Ms. Sudesna Debnath (Roll No. 23478123024)* have successfully completed the Project title ~

"Exploring the Relationship Between Music, Concentration, and Task Performance"

at NSHM Knowledge Campus, Kolkata (College Code: 234) under my supervision and guidance in the fulfilment of requirements of Fourth Semester, Master of Science (Data Science and Analytics) under Maulana Abul Kalam Azad University of Technology (MAKAUT), West Bengal.

Sanjay Goswami 16.06.25

Dr. Sanjay Goswami

(Professor)

M.Sc. Data Science at NSHM Knowledge Campus, Kolkata (College Code: 234)

DECLARATION

We declare that the work contained in this report is original and has been done by us under the guidance of our supervisor *Dr. Sanjay Goswami, Professor, NICA*. The work has not been submitted to any other Institute for any degree or diploma. We have followed the guidelines provided by the Institute in preparing the report. We have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute. Whenever we have used materials (data, theoretical analysis, figures, and text) from other sources, we have given due credit to them by citing them in the text of the report and giving their details in the references. Further, we have taken permission from the copyright owners of the sources, whenever necessary.

Akanksha Kumari

AKANKSHA KUMARI

Roll No.: 23478123001

Priti Saha

PRITI SAHA

Roll No.: 23478123009

Subarna Tapadar

SUBARNA TAPADAR

Roll No.: 23478123020

Sudesna Debnath

SUDESNA DEBNATH

Roll No.: 23478123024

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the report of the Project Work undertaken during our Master of Science Final semester. We owe special debt of gratitude to our project mentor Prof. Sanjay Goswami and Dr. Suparna Dhar. We would take this opportunity to thank faculty members for the constant support, guidance and motivation provided throughout the course of the project which enabled our endeavours and hard work to see the light of the day.

Last but not the least, we thank the Almighty for blessing us through the resources and giving us immense strength and willpower without which, the project would not have been made possible to the extent at which it has reached today.

ABSTRACT

Music is an integral part of human life, influencing both emotions and productivity. This study explores the dual perception of music as either a potential distraction or a powerful enhancer of efficiency. The project aims to uncover the impact of different musical genres, volume levels, and listening habits on focus, creativity, and task performance. It also examines how individual preferences determine whether music enhances or hinders productivity.

The survey was conducted using Google Forms, and data were collected on various aspects related to music and productivity. The responses were saved in Comma-Separated Values (.CSV) format for analysis. The study includes an Exploratory Data Analysis (EDA) consisting of three main components: Descriptive Analysis, Correlation Analysis, and Hypothesis Testing. These analyses were performed using Python libraries and Microsoft Excel (MS Excel) to identify patterns and extract meaningful insights.

Through both quantitative and qualitative methods, the study aims to provide a deeper understanding of music's role in professional and academic settings and to offer evidence-based recommendations for integrating music into daily routines to enhance focus and efficiency while minimizing distractions.

CONTENT

CHAPTER 1

INTRODUCTION

CHAPTER I
INTRODUCTION

1.1 Introduction

"Exploring the Relationship Between Music, Concentration, and Task Performance" is a research-based study that explores how music affects individuals' ability to perform various tasks, especially in environments that require focus, attention, and sustained mental effort. Music is a fundamental part of human life. It exists in every culture and has been used throughout history for storytelling, spiritual practices, celebration, comfort, and connection. However, in today's fast-paced, digitally driven world, music is also being used in a new way: as a tool for boosting productivity.

In modern times, music is no longer something we only hear during concerts, rituals, or celebrations. It has become a part of our daily environment. With the help of technology, music follows us wherever we go, from homes to offices, classrooms to gyms. We stream music through our phones, computers, smart speakers, and even through wearable devices like smartwatches. Whether commuting, cooking, studying, working, or relaxing, many people listen to music as a regular part of their routine *[Levitin, D. J. (2006)]*.

This ever-present role of music in our daily activities has raised an important question: Does music really help us become more productive, or is it just an enjoyable distraction? People use music in many different ways, some to block out noise, others to get into a "flow" state, and some simply to make dull tasks feel more enjoyable. However, the impact of music on work and learning is complex and often varies from person to person. The same music that helps one person focus might distract another *[Lesiuk, T. (2005)]*.

This study seeks to understand the true impact of music on productivity. It looks at the types of tasks music helps or hinders, the kinds of music that work best for different situations, and how personal preferences and personality traits shape the way we respond to music. In particular, it focuses on three key areas: task performance, concentration, and individual preferences.

1.1.1 Music and Work Performance

Work performance refers to how effectively and efficiently a person can complete a task. It includes various factors such as task accuracy, completion speed, consistency, and the level of satisfaction or stress experienced during the activity. These elements can be influenced by many internal and external factors, one of which is music.

Music can have a powerful effect on a person’s mental and emotional state. The right kind of music might energize a person who is tired or calm someone who feels stressed. Upbeat music can make repetitive or boring tasks feel more interesting, while soft background music can reduce anxiety in high-pressure situations.

Different genres of music may be more helpful depending on the type of work. For example:

- Ambient or instrumental music may help with writing, reading, or studying.
- Energetic music with a fast beat may improve motivation during physical or repetitive tasks.
- Classical music may enhance memory and concentration in academic work.
- Songs with lyrics may be distracting during tasks that involve reading or writing.

This section of the study explores how various types of music affect task performance across different domains such as academic work, creative tasks, and daily chores. It aims to identify whether music supports better performance or if silence is more beneficial.

1.1.2 Music, Focus, and Productivity

Focus is one of the most essential components of productivity. It is the ability to concentrate on a task without getting distracted. In work and learning environments, maintaining focus can be challenging, especially in the presence of background noise, interruptions, or internal stress.

Music can serve as a tool to support concentration. Some people find that music helps them tune out distractions, enter a "flow" state, and stay engaged with the task at hand *[Schellenberg, E. G. (2005)].* Others may find that music distracts them from their work, especially if the music is loud, fast-paced, or has lyrics that demand mental attention *[Kämpfe, J. (2011)].*

Music that supports focus is often characterized by:

- Lack of lyrics (instrumental or ambient music)
- Steady rhythm
- Consistent tempo
- Soothing tones and low complexity

This section of the research examines how different types of music affect concentration during tasks requiring mental effort, such as problem-solving, writing, or studying. It also considers how people use music to control their environment, such as masking background noise or creating a calm space for work.

1.1.3 Personal Preferences and Individual Differences

One of the most important factors in understanding music's impact on productivity is recognizing that music affects everyone differently. Personal taste, emotional state, and personality all influence how someone responds to music while working. For example, an extroverted person might feel more energized with loud, upbeat music, while an introverted person might prefer quiet, soothing background music.

Other factors that influence music preference and effectiveness include:

- **Mood:** Some people choose music that matches their current mood, while others use it to shift their mood.
- **Task type:** People may choose different music for different types of tasks.
- **Familiarity:** Listening to familiar music might be comforting and helpful, while unfamiliar music may either distract or refresh the listener.
- **Cultural background:** Musical preferences are deeply influenced by culture, which affects the emotional response to certain sounds or instruments.

This part of the study looks at how individual characteristics affect the use and impact of music during tasks. It also explores whether people perform better with self-selected music compared to randomly assigned music *[Furnham, A. (2002)]*.

1.2 Why This Study Matters

Despite music being widely used during work and study, there is no clear agreement among researchers about whether it actually improves productivity. Some studies suggest that it helps people focus and work more efficiently *[Lesiuk, T. (2005)]*. Others show that it can be distracting, especially for complex tasks *[Perham, N. (2014)]*. This confusion creates a need for more research.

This study addresses that gap by exploring music's effects across different genres, task types, and individuals. It collects both data and personal feedback to better understand when music helps, when it doesn't, and why.

By examining not just the "what" but also the "why," this research aims to give practical advice to students, professionals, and anyone who uses music to get through their daily responsibilities.

1.3 Practical Applications

Understanding the true impact of music on productivity can benefit many areas of life:

- Students can use this knowledge to create better study habits.

- Teachers and trainers can design better learning environments.
- Workplaces can improve focus and employee satisfaction by offering options for music in open spaces.
- App developers can design smarter music recommendation systems for productivity.

In an age where people are constantly looking for ways to improve performance and manage stress, knowing how to use music as a tool can make a meaningful difference. Music, if used thoughtfully, has the power to shape moods, manage energy, and support focus. But it must be used in the right way, at the right time, and in the right context.

This study hopes to guide people in using music not just for entertainment, but for enhancement, to become not just background noise, but a helpful part of a more focused and productive life.

1.4 Aim Of The Survey

The aim of the survey is to explore how music influences productivity in different work and study environments.

It seeks to understand whether music enhances or hinders focus, efficiency, and task performance. The survey also aims to identify which types of music are most effective for various activities. Ultimately, it helps uncover how age or activity patterns influence music preferences and their impact on productivity.

1.5 Objectives of the project

To analyze how music affects focus, productivity, and individual preferences in work and study contexts using survey-based insights.

Objectives of the Survey:

I. To evaluate the impact of music on task performance across different types of activities.

II. To explore the role of music in supporting focus, attention, and sustained mental effort.

III. To understand how age or activity patterns influence music preferences and their impact on productivity.

IV. To examine how music affects productivity in different work and study environments.

CHAPTER 2

METHODOLOGY

CHAPTER II
METHODOLOGY

The purpose of the study is to analyse various categorical factors and examine whether any significant relationships exist between them using statistical methods. A total of 301 individuals were considered for the survey. The data was collected through primary means, ensuring firsthand insights relevant to the objectives of the study.

2.1 Sampling Procedure:

It was not feasible to conduct the survey across an entire large population due to constraints of time and resources. Therefore, the data collection was limited to 301 respondents selected through simple random sampling. This method ensures that every individual had an equal chance of being chosen, thereby minimizing sampling bias and enhancing the representativeness of the data *[Moorthy, M. N. (1988)]*.

2.2 Correlation Analysis

Correlation analysis is a statistical technique used to measure the strength and direction of a relationship between two variables *[Gupta, & Kapoor (1987)]*. It is commonly applied to identify patterns and trends in data, helping researchers understand how variables influence one another. In the context of this study, correlation analysis was employed to examine the degree of association between various categorical variables collected from a primary dataset.

Since the dataset primarily consisted of categorical variables, some of which had an inherent order (ordinal data), the Spearman's Rank Correlation Coefficient was selected as the most suitable method. This non-parametric technique does not require the assumption of normal distribution and is particularly useful when the relationship between variables is monotonic rather than linear.

Correlation helps explain how two variables are dependent on one another. It quantifies this relationship using a correlation coefficient (ρ or *rho*), which ranges from -1 to +1:

- A positive value indicates that both variables increase or decrease together (positive correlation).
- A negative value indicates that as one variable increases, the other decreases (negative correlation).
- A value of zero indicates no relationship between the two variables (zero correlation).

2.2.1 Types of Correlation

Correlation can be classified into three types:

- Positive Correlation: This occurs when both variables move in the same direction. For example, there is a positive correlation between height and weight as height increases, weight tends to increase.
- Negative Correlation: This occurs when the variables move in opposite directions. For example, there is a negative correlation between price and demand as price increases, demand typically decreases.
- Zero Correlation: This means there is no discernible relationship between the two variables. For example, shoe size and intelligence usually have zero correlation, as changes in one do not affect the other.

The strength of the correlation is measured by the correlation coefficient:

- +1 or -1 indicates a perfect correlation (either positive or negative).
- 0 indicates no correlation.

2.2.2 Ranking Methods for Correlation

There are three commonly used techniques to measure correlation:

- **Spearman's Rank Correlation** – Suitable for ordinal or non-linear data.
- **Kendall's Tau** – Used for small datasets or tied ranks.
- **Pearson's Correlation Coefficient** – Best for continuous and normally distributed data.

Among these, Spearman's Rank Correlation was chosen for this study due to its compatibility with the ordinal categorical data and its ability to identify monotonic relationships.

2.2.3 Spearman's Rank Correlation

Spearman's Rank Correlation Coefficient is a non-parametric statistical measure that evaluates the strength and direction of the association between two ranked variables. It is especially useful when the data is ordinal in nature, meaning the values represent positions or rankings rather than quantities (e.g., 1st, 2nd, 3rd).

2.2.4 Purpose

This coefficient is used to determine whether there is a monotonic relationship between two variables that is, whether the variables tend to move in the same or opposite direction, regardless of the exact rate of change.

2.2.5 How It Works

To compute Spearman's Rank Correlation:

- The raw values of each variable are replaced with their respective ranks.
- The difference between the ranks of each paired observation is calculated.
- These differences are squared and summed.
- The Spearman coefficient (ρ) is calculated using the formula:

$$\boxed{\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}}$$

Where:

- **ρ** = Spearman's Rank Correlation Coefficient
- d_i = Difference between the ranks of paired variables
- **n** = Number of observations

Interpretation

- **ρ = +1** → Perfect positive correlation (ranks increase together)
- **ρ = -1** → Perfect negative correlation (one rank increases, the other decreases)
- **ρ = 0** → No correlation between ranks

2.3 Chi-Square Test

The Chi-Square (χ^2) test is a statistical method used to assess the association or fit between observed and expected frequencies in categorical data. It is widely used in hypothesis testing to determine whether there is a significant relationship between variables or whether the distribution of sample data fits a specific pattern *[Gupta, & Kapoor (1987)]*. There are two main types of Chi-Square tests: the Chi-Square Test of Goodness of Fit and the Chi-Square Test of Independence.

- **Chi-Square Test of Goodness of Fit:** This test is used to determine whether the distribution of a single categorical variable matches a predefined or expected distribution. It compares the observed frequency of each category to the expected frequency under a specific hypothesis.
- **Chi-Square Test of Independence:** This test evaluates whether two categorical variables are independent or associated with each other. It is used when the data consists of counts of observations in different categories.

  Formula:

  The test statistic for the Chi-Square Test of Independence is calculated using the formula:

$$\boxed{\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}}$$

where:

- $\boldsymbol{O_i}$ = observed frequency in the i^{th} category
- $\boldsymbol{E_i}$ = expected frequency in the i^{th} category, calculated under the assumption of independence

CHAPTER 3

ANALYSIS

CHAPTER III
ANALYSIS

3.1 Data Cleaning and Preparation

3.1.1 Using Excel

The dataset used for analysis consisted of **301 survey responses**. Prior to analysis, essential cleaning steps were performed using **Microsoft Excel** to ensure clarity, consistency, and accuracy.

The steps included:

I. **Renaming Columns**: Some column names in the original dataset were lengthy. These were renamed with shorter and more intuitive labels to improve readability and maintain uniformity across the dataset. Examples include:

- "How often do you listen to music while doing something?" → "Duration".
- "What activities do you do while listening to music? (Pick all that apply!)" → "Activity".
- "What kind of music do you enjoy the most? (Pick all that apply!)" → "Genre".
- "How loud do you like your music?" → "Music_Volume".
- "Does music help you concentrate?" → "Help_concentrate".
- "Do you think some types of music make it harder to concentrate?" → "Genre_Distriction".
- "Do you feel more productive ________?" → "Productivity".
- "If you had to focus for 3+ hours, which would you prefer?" → "Long_Duration".
- "Do you listen to music to reduce stress?" → "Stress_relief".
- "Does music affect productivity while performing critical tasks?" → "Critical_productivity".

II. **Correcting Spelling and Capitalization Errors**: Several spelling mistakes and inconsistent capitalization were identified in the responses, particularly in categorical entries. These were manually corrected to maintain uniformity across the dataset. Examples include:

- "Home maker, Home Maker" → "Homemaker".
- "UNEMPLOYED" → "Unemployed".
- "Gym" → "Physical tasks".

These corrections were applied across some columns, including occupation, genre, and activities.

III. **Reclassifying Activities**: Responses such as “sleeping,” “bathing,” “nothing,” and other similar personal routines were grouped under a broader and more structured category titled **“Leisure Activities”**. This helped reduce fragmentation in responses and allowed for clearer interpretation during analysis.

These focused cleaning steps helped refine the dataset and prepare it for correlation and pattern analysis in the subsequent stages of the study.

3.1.2 Using Python

The unrefined dataset contained with multi-valued entries in the Genre and Activity columns. To prepare it for analysis, we performed several key steps using Python:

- Loaded and examined the dataset for structure, missing values, and duplicates.
- Converted multi-valued entries into lists and normalized them into individual rows.
- Removed redundant and missing entries after transformation.
- Grouped records to count unique values per respondent.
- Applied logic to simplify preferences based on defined criteria.
- Finalized the cleaned data with consistent records and added new features for analysis.

A. **Preprocessing Data using Pandas Library** *[Pandas Documentation, September 20,2024, Version – 2.2.3]*

i. Importing Necessary Library

Essential libraries for data analysis:

```
# Importing necessary libraries
import pandas as pd   # For data manipulation
import matplotlib.pyplot as plt   # For plotting graphs
import seaborn as sns   # For data visualization
import statsmodels.api as sm   # For statistical models
from scipy.stats import spearmanr, chi2_contingency, chi2   # For statistical tests
```

Figure 3.1 Importing the libraries

ii. Loading and Previewing a Dataset

Loads a CSV into a Pandas Data Frame and previews the data.

```
df = pd.read_csv('/content/Unprocessed Data.csv')  # Load CSV into DataFrame
df.head()  # Show first 5 rows
```

Figure 3.2 Load the CSV file

	Res_ID	Age	Occupation	Duration	Activity	Genre	Music_Volume	Help_concentrate	Genre_Distraction	Productivity	Long_Duration	Stress_relief	Critical_productivity
0	U_01	18-25	Student	Rarely	Writing or creating something, Physical tasks	Instrumental,Rock,Electronic/Dance,Songs with ...	Moderate	Yes	No	With music	No music at all	Yes	I don't listen to music
1	U_02	18-25	Student	Sometimes	Physical tasks	Instrumental,Electronic/Dance,Nature sounds	Low/Soft	Makes no difference	Maybe	Without music	Music with breaks in between	Yes	Increases productivity
2	U_03	18-25	Student	Often	Writing or creating something, Physical tasks,...	Instrumental,Songs with lyrics	Low/Soft	Yes	Maybe	With music	Music with breaks in between	Yes	Increases productivity
3	U_04	18-25	Student	Often	Solving problems,Physical tasks	Instrumental,Songs with lyrics,No preference	Moderate	Yes	Yes	With music	Music with breaks in between	Yes	No difference
4	U_05	18-25	Student	Sometimes	Writing or creating something, Physical tasks,...	Instrumental,Songs with lyrics,Nature sounds	Moderate	Yes	Yes	With music	Music with breaks in between	sometimes	No difference

Figure 3.3 Structure of the dataset

iii. DataFrame Shape

```
# Shows DataFrame dimensions (rows, columns)
df.shape
```

```
(301, 13)
```

Figure 3.4 Dataframe dimension

The output (301, 13) means the dataset contains **301 rows** and **13 columns**, giving a quick overview of its size.

iv. Inspecting Column Names

```
# Displays column names of the DataFrame
df.columns
```

```
Index(['Res_ID', 'Age', 'Occupation', 'Duration', 'Activity', 'Genre',
       'Music_Volume', 'Help_concentrate', 'Genre_Distraction', 'Productivity',
       'Long_Duration', 'Stress_relief', 'Critical_productivity'],
      dtype='object')
```

Figure 3.5 Display column name

v. Checking Missing Values

```
df.isnull().sum()  # Count missing values per column
```

	0
Res_ID	0
Age	0
Occupation	0
Duration	0
Activity	0
Genre	0
Music_Volume	0
Help_concentrate	0
Genre_Distraction	0
Productivity	0
Long_Duration	0
Stress_relief	0
Critical_productivity	0

Figure 3.6 Count of missing number

The output shows that all columns have **0 missing values**, indicating a complete and clean dataset no need for imputation or removal.

vi. DataFrame Info Summary

```
# Summary of DataFrame: columns, non-null counts, and data types
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 301 entries, 0 to 300
Data columns (total 13 columns):
 #   Column                 Non-Null Count  Dtype 
---  ------                 --------------  ----- 
 0   Res_ID                 301 non-null    object
 1   Age                    301 non-null    object
 2   Occupation             301 non-null    object
 3   Duration               301 non-null    object
 4   Activity               301 non-null    object
 5   Genre                  301 non-null    object
 6   Music_Volume           301 non-null    object
 7   Help_concentrate       301 non-null    object
 8   Genre_Distraction      301 non-null    object
 9   Productivity           301 non-null    object
 10  Long_Duration          301 non-null    object
 11  Stress_relief          301 non-null    object
 12  Critical_productivity  301 non-null    object
dtypes: object(13)
memory usage: 30.7+ KB
```

Figure 3.7 Summary of dataframe

In this output, all 13 columns have 301 non-null values and are of type object, meaning there are **no missing values** and the data is likely stored as text or mixed types.

vii. Handling Duplicate Rows in a DataFrame: Detection and Removal

```
# Count duplicate rows
df.duplicated().sum()
```

```
np.int64(0)
```

Figure 3.8 Count of duplicate

This case involves checking a DataFrame to see if it contains any duplicate rows. After performing the check, it was found that the DataFrame had no duplicates at all.

viii. Column-wise Data Summary and Insights

This summary provides a column-by-column overview of the dataset, highlighting key value distributions and brief insights for each feature.

a) Age Group Frequency Analysis

```
df['Age'].value_counts()   # Frequency of each age value
```

Age	count
18-25	204
26-35	56
36 above	22
Under 18	19

dtype: int64

Figure 3.9 Count of individuals

The majority of individuals fall in the **18–25 (204)** age group, indicating a younger demographic in the dataset.

b) **Occupation Frequency Analysis**

```
df['Occupation'].value_counts()  # Frequency of each occupation value
```

Occupation	count
Student	183
Office work	62
Self-Employed	18
Work from home	10
Creative/Artistic field	7
Teacher	7
Homemaker	7
Retired	5
Unemployed	2

dtype: int64

Figure 3.10 Count of frequency

Most individuals are students (183), highlighting a primarily academic demographic.

c) Duration of Music Listening

```
df['Duration'].value_counts()  # Frequency of each duration value
```

Duration	count
Often	95
Always	94
Sometimes	89
Rarely	23

dtype: int64

Figure 4.1 Count of how frequently individual listen to music

The responses are quite balanced, but "Often" (95) and "Always" (94) are the most common, suggesting regular music usage.

d) Activity Preference Analysis

```
df['Activity'].value_counts()  # Frequency of each activity value
```

Activity	count
Physical tasks	70
Writing or creating something	35
Writing or creating something, Physical tasks	34
Writing or creating something,Solving problems,Physical tasks	31
Solving problems,Physical tasks	24

Figure 3.12 Count of activity

"Physical tasks" (70) top the list, indicating music is commonly used during physical activities.

e) Music Genre Preference

```
df['Genre'].value_counts()  # Frequency of each genre value
```

Genre	count
Songs with lyrics	86
Instrumental,Songs with lyrics	29
Instrumental	19
No preference	18
Instrumental,Songs with lyrics,Nature sounds	17

Figure 3.13 Count of genre

"Songs with lyrics" (86) top the list, indicating this genre is commonly used.

f) Music Volume Frequency Analysis

```
df['Music_Volume'].value_counts()  # Frequency of each music volume
```

Music_Volume	count
Moderate	163
Low/Soft	102
High	36

dtype: int64

Figure 3.14 Count of music volume

"Moderate" (163) is the most preferred volume, showing a balance between intensity and comfort.

g) Music and Concentration

```
df['Help_concentrate'].value_counts()  # Frequency of each help concentrate value
```

Help_concentrate	count
Yes	220
Makes no difference	48
No	33

dtype: int64

Figure 3.15 Count of concentration with music

A large portion (220) said "Yes", suggesting music positively supports focus.

h) Genre Distraction

```
df['Genre_Distraction'].value_counts()  # Frequency of each genre distraction value
```

Genre_Distraction	count
Yes	170
Maybe	91
No	40

dtype: int64

Figure 3.16 Count of genre distraction

"Yes" (170) is the most frequent, showing many find certain genres distracting.

i) Productivity with Music

```
df['Productivity'].value_counts()  # Frequency of each productivity value
```

Productivity	count
With music	229
Without music	72

dtype: int64

Figure 3.17 Count of productivity with music

Most individuals (229) report being more productive with music, supporting its use as a productivity aid.

j) Music During Long Tasks

```
df['Long_Duration'].value_counts()  # Frequency of each long duration
```

Long_Duration	count
Music with breaks in between	174
Music playing the entire time	77
No music at all	50

dtype: int64

Figure 3.18 Count of music during long task

"Music with breaks in between" (174) is most common, indicating a balanced approach to music during long sessions.

k) Music for Stress Relief

```
df['Stress_relief'].value_counts()  # Frequency of each stress relief
```

Stress_relief	count
Yes	224
sometimes	60
No	17

dtype: int64

Figure 3.19 Count of music for stress relief

The majority (224) answered "Yes", confirming music as a common stress relief method.

l) Critical Productivity Frequency Analysis

```
df['Critical_productivity'].value_counts()  # Frequency of each critical productivity
```

Critical_productivity	count
Increases productivity	157
No difference	69
Decreases productivity	52
I don't listen to music	23

dtype: int64

Figure 3.20 Count of each critical productivity

Most individuals believe that music increases productivity, while fewer report a neutral or negative impact.

ix. Handling Multi-Select Responses in 'Activity' and 'Genre' Columns

Some survey questions allowed multiple responses stored as comma-separated values. To prepare these for analysis:

- **Cleaning:** Text in the Activity and Genre columns was converted to lowercase and stripped of extra spaces for consistency.
- **Splitting:** Responses were split into lists using commas as separators.
- **Exploding:** Each list was expanded into individual rows so that each activity or genre appeared separately, enabling accurate analysis.

This transformation ensured cleaner, more structured data for evaluating activity and genre preferences.

```python
# 1. Clean the 'Activity' column
df['Activity_clean'] = df['Activity'].astype(str).str.lower().str.strip()

# 2. Split by comma
df['activity_list'] = df['Activity_clean'].str.split(r'\s*,\s*')

# 3. Explode into separate rows
df = df.explode('activity_list').reset_index(drop=True)

# 4. Now print to check
print(df.head())
```

```
  Res_ID    Age Occupation   Duration  \
0   U_01  18-25    Student     Rarely
1   U_01  18-25    Student     Rarely
2   U_02  18-25    Student  Sometimes
3   U_03  18-25    Student      Often
4   U_03  18-25    Student      Often

                                        Activity  \
0      Writing or creating something, Physical tasks
1      Writing or creating something, Physical tasks
2                                    Physical tasks
3  Writing or creating something, Physical tasks,...
4  Writing or creating something, Physical tasks,...

                                       Genre Music_Volume  \
0  Instrumental,Rock,Electronic/Dance,Songs with ...     Moderate
1  Instrumental,Rock,Electronic/Dance,Songs with ...     Moderate
2        Instrumental,Electronic/Dance,Nature sounds     Low/Soft
3                      Instrumental,Songs with lyrics     Low/Soft
4                      Instrumental,Songs with lyrics     Low/Soft

     Help_concentrate Genre_Distraction   Productivity  \
0                 Yes                No     With music
1                 Yes                No     With music
2  Makes no difference             Maybe  Without music
3                 Yes             Maybe     With music
4                 Yes             Maybe     With music


               Long_Duration Stress_relief    Critical_productivity  \
0             No music at all           Yes  I don't listen to music
1             No music at all           Yes  I don't listen to music
2  Music with breaks in between           Yes   Increases productivity
3  Music with breaks in between           Yes   Increases productivity
4  Music with breaks in between           Yes   Increases productivity

                                    Activity_clean  \
0      writing or creating something, physical tasks
1      writing or creating something, physical tasks
2                                    physical tasks
3  writing or creating something, physical tasks,...
4  writing or creating something, physical tasks,...

                 activity_list
0  writing or creating something
1                 physical tasks
2                 physical tasks
3  writing or creating something
4                 physical tasks
```

Figure 3.21 Handling multi-valued value in activity column

```python
# 1. Clean the 'Genre' column
df['Genre_clean'] = df['Genre'].astype(str).str.lower().str.strip()

# 2. Split by comma
df['genre_list'] = df['Genre_clean'].str.split(r'\s*,\s*')

# 3. Explode into separate rows
df = df.explode('genre_list').reset_index(drop=True)

# 4. Now print to check
print(df.head())
```

```
  Res_ID    Age Occupation Duration  \
0   U_01  18-25    Student   Rarely
1   U_01  18-25    Student   Rarely
2   U_01  18-25    Student   Rarely
3   U_01  18-25    Student   Rarely
4   U_01  18-25    Student   Rarely

                                         Activity  \
0  Writing or creating something, Physical tasks
1  Writing or creating something, Physical tasks
2  Writing or creating something, Physical tasks
3  Writing or creating something, Physical tasks
4  Writing or creating something, Physical tasks

                                          Genre Music_Volume  \
0  Instrumental,Rock,Electronic/Dance,Songs with ...     Moderate
1  Instrumental,Rock,Electronic/Dance,Songs with ...     Moderate
2  Instrumental,Rock,Electronic/Dance,Songs with ...     Moderate
3  Instrumental,Rock,Electronic/Dance,Songs with ...     Moderate
4  Instrumental,Rock,Electronic/Dance,Songs with ...     Moderate

  Help_concentrate Genre_Distraction Productivity    Long_Duration  \
0              Yes                No   With music  No music at all
1              Yes                No   With music  No music at all
2              Yes                No   With music  No music at all
3              Yes                No   With music  No music at all
4              Yes                No   With music  No music at all

  Stress_relief    Critical_productivity  \
0           Yes  I don't listen to music
1           Yes  I don't listen to music
2           Yes  I don't listen to music
3           Yes  I don't listen to music
4           Yes  I don't listen to music
```

```
                                    Activity_clean  \
0  writing or creating something, physical tasks
1  writing or creating something, physical tasks
2  writing or creating something, physical tasks
3  writing or creating something, physical tasks
4  writing or creating something, physical tasks

                     activity_list  \
0  writing or creating something
1  writing or creating something
2  writing or creating something
3  writing or creating something
4  writing or creating something

                                        Genre_clean         genre_list
0  instrumental,rock,electronic/dance,songs with ...       instrumental
1  instrumental,rock,electronic/dance,songs with ...               rock
2  instrumental,rock,electronic/dance,songs with ...    electronic/dance
3  instrumental,rock,electronic/dance,songs with ...  songs with lyrics
4  instrumental,rock,electronic/dance,songs with ...       nature sounds
```

Figure 3.22 Handling multi-valued value in genre column

```
df.head(15)
```

Res_ID	Age	Occupation	Duration	Activity	Genre	Music_Volume	Help_concentrate	Genre_Distraction	Productivity	Long_Duration	Stress_relief	Critical_productivity	Activity_clean	activity_list	Genre_clean	genre_list
U_01	18-25	Student	Rarely	Writing or creating something, Physical tasks	Instrumental,Rock,Electronic/Dance,Songs with ...	Moderate	Yes	No	With music	No music at all	Yes	I don't listen to music	writing or creating something, physical tasks	writing or creating something	instrumental,rock,electronic/dance,songs with ...	instrumental
U_01	18-25	Student	Rarely	Writing or creating something, Physical tasks	Instrumental,Rock,Electronic/Dance,Songs with ...	Moderate	Yes	No	With music	No music at all	Yes	I don't listen to music	writing or creating something, physical tasks	writing or creating something	instrumental,rock,electronic/dance,songs with ...	rock
U_01	18-25	Student	Rarely	Writing or creating something, Physical tasks	Instrumental,Rock,Electronic/Dance,Songs with ...	Moderate	Yes	No	With music	No music at all	Yes	I don't listen to music	writing or creating something, physical tasks	writing or creating something	instrumental,rock,electronic/dance,songs with ...	electronic/dance
U_01	18-25	Student	Rarely	Writing or creating something, Physical tasks	Instrumental,Rock,Electronic/Dance,Songs with ...	Moderate	Yes	No	With music	No music at all	Yes	I don't listen to music	writing or creating something, physical tasks	writing or creating something	instrumental,rock,electronic/dance,songs with ...	songs with lyrics
U_01	18-25	Student	Rarely	Writing or creating something, Physical tasks	Instrumental,Rock,Electronic/Dance,Songs with ...	Moderate	Yes	No	With music	No music at all	Yes	I don't listen to music	writing or creating something, physical tasks	writing or creating something	instrumental,rock,electronic/dance,songs with ...	nature sounds

Figure 3.23 Dataframe after handling multi-valued column's

x. DataFrame Shape

```
df.shape  # DataFrame dimensions (rows, columns)
```

```
(1222, 17)
```

Figure 3.24 Dataframe Shape after handling multi-valued column's

xi. Checking Missing Values

```
df.isnull().sum()  # Count of missing values in each column
```

	0
Res_ID	0
Age	0
Occupation	0
Duration	0
Activity	0
Genre	0
Music_Volume	0
Help_concentrate	0
Genre_Distraction	0
Productivity	0
Long_Duration	0
Stress_relief	0
Critical_productivity	0
Activity_clean	0
activity_list	0
Genre_clean	0
genre_list	0

dtype: int64

Figure 3.25 Checking missing value

A check was conducted for missing values across all columns. The result confirmed that no missing data is present in any of the columns, indicating a complete and clean dataset.

xii. Handling Duplicate Rows in a DataFrame: Detection and Removal

```
df.duplicated().sum()  # Number of duplicate rows in the DataFrame
```

```
np.int64(2)
```

```
df = df.drop_duplicates().reset_index(drop=True)  # Remove duplicate rows and reset the index
```

```
df.duplicated().sum()  # Number of duplicate rows in the DataFrame
```

```
np.int64(0)
```

Figure 3.26 Handling duplicate

An initial scan of the dataset revealed 2 duplicate entries. These duplicates were removed, and the DataFrame index was reset to maintain consistency.

A follow-up check confirmed that no duplicate rows remained, ensuring the dataset was clean and ready for analysis.

xiii. Unique Activity and Genre Counts per Respondent

```
# Count unique activities and genres per Res_ID
counts = df.groupby('Res_ID').agg({
    'activity_list': pd.Series.nunique,
    'genre_list': pd.Series.nunique
}).reset_index()

# Rename columns for clarity
counts.columns = ['Res_ID', 'Unique_Activity_Count', 'Unique_Genre_Count']

# Show result
counts.head(10)
```

	Res_ID	Unique_Activity_Count	Unique_Genre_Count
0	U_01	2	5
1	U_02	1	3
2	U_03	3	2
3	U_04	2	3
4	U_05	3	3
5	U_06	1	2
6	U_07	2	1
7	U_08	2	3
8	U_09	2	2
9	U_10	4	5

Figure 3.27 Count the unique activity & genre

To understand the diversity of music and activities among individuals, we calculated the number of unique activities and unique genres associated with each respondent.

This provided a personalized profile of engagement, revealing how varied each person's music and activity preferences were.

xiv. Categorizing Genre Preferences

To streamline genre analysis, genre entries were standardized by converting them to lowercase and removing extra spaces. Then, each respondent's unique genre list was examined:

- If a person had five or more genres or selected "no preference", their preference was categorized as "All".
- Otherwise, their actual genre selections were retained.

The refined preferences were then merged back with the main dataset to support further personalized analysis.

```
df = df.drop(columns=['Genre_Preference'], errors='ignore')  # Drop Genre_Preference column if it exists
df = df.merge(genre_summary, on='Res_ID', how='left')  # Merge genre summary with the original DataFrame
```

```
# Ensure genre_list is clean
df['genre_list'] = df['genre_list'].str.strip().str.lower()

# Group by Res_ID and get unique genres
genre_summary = df.groupby('Res_ID')['genre_list'].unique().reset_index()

# Apply logic: All if >=5 g Loading... ontains 'no preference'
def get_genre_preference(genres):
    if len(genres) >= 5 or 'no preference' in genres:
        return ['All']
    else:
        return genres

# Apply and explode
genre_summary['Genre_Preference'] = genre_summary['genre_list'].apply(get_genre_preference)
genre_summary = genre_summary.explode('Genre_Preference').reset_index(drop=True)

# Final output
genre_summary = genre_summary[['Res_ID', 'Genre_Preference']]
```

Figure 3.28 Categorizing Genre Preferences

xv. Data Refinement: Column Pruning

To streamline the dataset for analysis, several columns were removed. These included raw or intermediary versions of activity and genre data, as well as distraction-related genre information. Only the finalized columns such as refined activity lists and categorized genre preferences were retained, ensuring clarity and reducing redundancy in the dataset.

```
# Drop unnecessary columns
df = df.drop(columns=[
    'Activity',
    'Genre',
    'Activity_clean',
    'Genre_clean',
    'genre_list',
    'Genre_Distraction'])

df.head()  # Display the first 5 rows of the DataFrame
```

	Res_ID	Age	Occupation	Duration	Music_Volume	Help_concentrate	Productivity	Long_Duration	Stress_relief	Critical_productivity	activity_list	Genre_Preference
0	U_01	18-25	Student	Rarely	Moderate	Yes	With music	No music at all	Yes	I don't listen to music	writing or creating something	All
1	U_01	18-25	Student	Rarely	Moderate	Yes	With music	No music at all	Yes	I don't listen to music	writing or creating something	All
2	U_01	18-25	Student	Rarely	Moderate	Yes	With music	No music at all	Yes	I don't listen to music	writing or creating something	All
3	U_01	18-25	Student	Rarely	Moderate	Yes	With music	No music at all	Yes	I don't listen to music	writing or creating something	All
4	U_01	18-25	Student	Rarely	Moderate	Yes	With music	No music at all	Yes	I don't listen to music	writing or creating something	All

Figure 3.29 Column pruning

xvi. Duplicate Removal

```
df.shape  # Display DataFrame dimensions (rows, columns)
```
```
(2656, 12)
```
```
df.duplicated().sum()  # Count duplicate rows in the DataFrame
```
```
np.int64(1630)
```
```
# Remove duplicates and reset the index
df = df.drop_duplicates().reset_index(drop=True)
```
```
df.duplicated().sum()  # Count duplicate rows in the DataFrame
```
```
np.int64(0)
```

Figure 3.30 Handling duplicate

Initially, the dataset contained 2,656 rows and 12 columns, with 1,630 identified as exact duplicates. To ensure data integrity, these duplicates were removed, and the index was reset. This process reduced the dataset to only unique entries, eliminating redundancy and ensuring accurate analysis. Now we got our processed dataset

3.2 Descriptive Analysis

- Descriptive analysis is the process of summarizing and visualizing data to understand its key characteristics. It provides insights into the distribution, and relationships within the dataset *[Chart Visualization, Pandas Documentation, September 20,2024, Version – 2.2.3]*.

i. Age Distribution

A pie chart was used to show the age distribution of respondents, highlighting the most represented age groups for a quick view of demographic spread

```
# Group by Age and count unique Res_IDs
age_counts = df.groupby('Age')['Res_ID'].nunique().sort_values(ascending=False)
total = age_counts.sum()

# Create pie chart for Age distribution
plt.figure(figsize=(8, 5))
wedges, texts, autotexts = plt.pie(age_counts, colors=['#6BD7D0', '#FBCCCA', '#EEEFF3', '#F0E3D3'],
                                   autopct=lambda pct: f"{int(round(pct * total / 100.0))}\n({pct:.1f}%)",
                                   startangle=90, textprops={'color': 'black', 'fontsize': 8, 'style': 'italic'})
for w in wedges: w.set_edgecolor('black')
plt.legend(age_counts.index, bbox_to_anchor=(1, 0.5), handleheight=2)
plt.title('Distribution of Age', weight='bold')
plt.tight_layout()
plt.show()
```

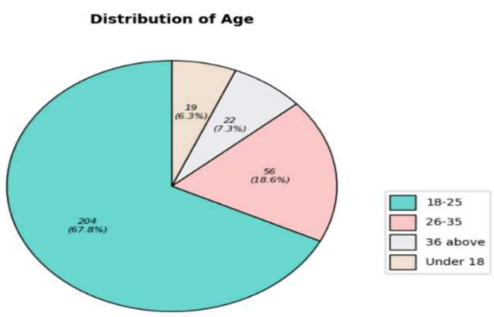

Figure 3.31 Group by age and count unique ID

The majority of participants (67.8%) are between 18-25 years, making it the most represented group. Smaller proportions fall into the 26-35 (18.6%), 36 above (7.3%), and Under 18 (6.3%) categories. This skew suggests a youth-dominated sample, which may influence how music affects productivity across age groups.

ii. Occupation Distribution

A horizontal bar chart shows occupation distribution by counting unique participant IDs. It reveals that most respondents are students, highlighting their dominance in the dataset.

```
# Count the number of responses for each occupation
Occupation_counts = df.groupby('Occupation')['Res_ID'].nunique().sort_values(ascending=False)
total = Occupation_counts.sum()  # Total number of responses

plt.figure(figsize=(10, 6))
plt.barh(Occupation_counts.index, Occupation_counts, color='seagreen')  # Use a green shade for bars

# Extend x-axis to make room for labels on the right
plt.xlim(0, Occupation_counts.max() * 1.2)

# Add value + percentage labels next to each bar
for i, v in enumerate(Occupation_counts):
    label = f'{v} ({v/total:.1%})'  # Example: "35 (23.3%)"
    plt.text(v + 5, i, label, va='center', ha='left', fontsize=11)

# Set chart title and axis labels
plt.title("Occupation", fontsize=14, weight='bold')
plt.xlabel("")  # You can remove this line if you prefer no x-axis label
plt.yticks(fontsize=12)
plt.grid(False)

# Adjust layout to prevent overlap
plt.tight_layout()

# Display the chart
plt.show()
```

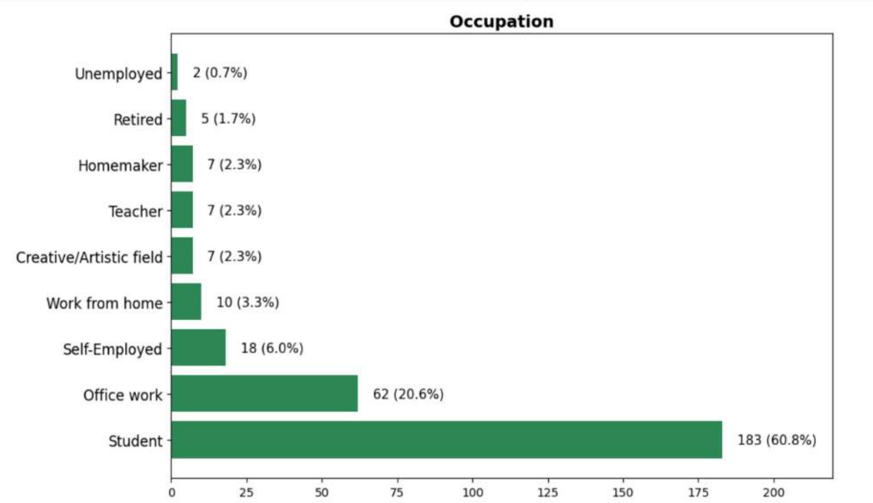

Figure 3.32 Count the no. of responses for each occupation

This bar chart shows the different occupations of survey participants. Most of them are students (60.8%), followed by people in office jobs (20.6%) and self-employed individuals (6%). A few participants work from home, are teachers, homemakers, retired, or unemployed. This suggests that the majority of responses came from a younger, student-based group.

iii. Listening Habits Distribution

A pie chart shows how often respondents listen to music, based on unique IDs from the Duration field. It highlights music's role in daily routines and its impact on focus and productivity.

```
# Group by Duration and count unique Res_IDs
value_counts = df.groupby('Duration')['Res_ID'].nunique().sort_values(ascending=False)
total = value_counts.sum()

# Create pie chart for Listening Habits distribution
plt.figure(figsize=(7, 5))
wedges, texts, autotexts = plt.pie(value_counts, colors=['#F56353', '#FE9430', '#FFE95C', '#11BD98'],
                                   autopct=lambda pct: f"{int(round(pct * total / 100.0))}\n({pct:.1f}%)",
                                   startangle=90, textprops={'color': 'black', 'fontsize': 8, 'style': 'italic', 'weight': 'bold'})
for w in wedges: w.set_edgecolor('black')
plt.legend(value_counts.index, bbox_to_anchor=(1, 0.5), handleheight=2)
plt.title('Listening Habits Distribution', weight='bold')
plt.tight_layout()
plt.show()
```

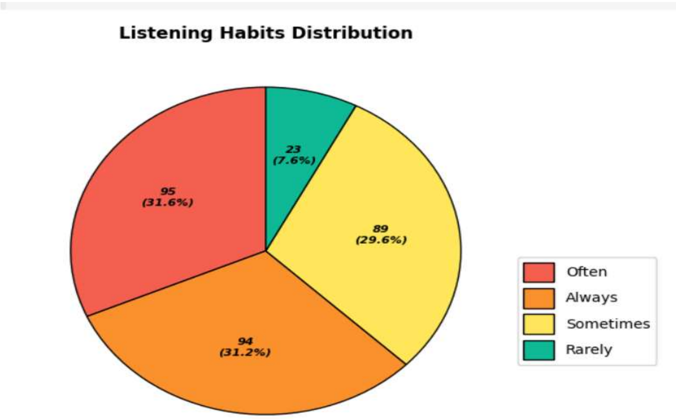

Figure 3.33 Listing habits

Most participants listen to music either often (31.6%) or always (31.2%), showing high regularity in listening habits. Around 29.6% listen sometimes, while only 7.6% reported rarely listening. This suggests that music is a regular part of life for the majority, which may influence how it affects their task performance and mental engagement.

iv. Music Volume Preference

A pie chart shows preferred music volume levels, based on unique respondent IDs from the Music_Volume field. It reveals trends in volume preferences and their possible effects on concentration and productivity.

```python
# Group by Music_Volume and count unique Res_IDs
value_counts = df.groupby('Music_Volume')['Res_ID'].nunique().sort_values(ascending=False)
total = value_counts.sum()

# Create pie chart for Music Volume Preference distribution
plt.figure(figsize=(6, 4))
wedges, texts, autotexts = plt.pie(value_counts, colors=['tomato', 'cornflowerblue', 'gold'],
                                   autopct=lambda pct: f"{int(round(pct * total / 100.0))}\n({pct:.1f}%)",
                                   startangle=90, textprops={'color': 'black', 'fontsize': 8, 'style': 'italic', 'weight': 'bold'})
for w in wedges: w.set_edgecolor('black')
plt.legend(value_counts.index, bbox_to_anchor=(1, 0.5), handleheight=2)
plt.title('Music Volume Preference', weight='bold')
plt.tight_layout()
plt.show()
```

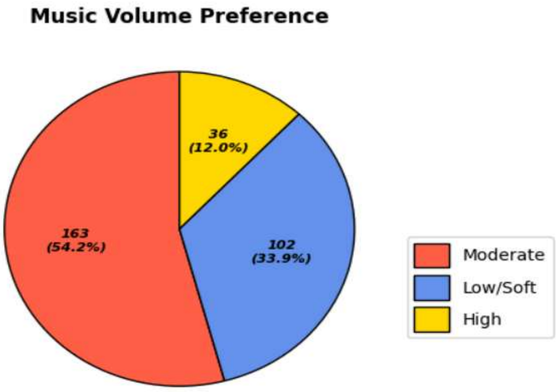

Figure 3.34 Volume preference

Over half the respondents prefer moderate volume (54.2%), making it the most common choice. About 33.9% enjoy low/soft music, while 12.0% prefer high volume. These preferences suggest that most individuals seek a balanced sound level, potentially to maintain focus without distraction, while fewer Favor extreme ends of the volume spectrum.

v. Does Music Help You Concentrate?

A pie chart shows how music affects concentration, using self-reported data from the Help_concentrate column and unique participant IDs. It captures overall views on music as a productivity aid.

```python
# Group by Help_concentrate and count unique Res_IDs
value_counts = df.groupby('Help_concentrate')['Res_ID'].nunique().sort_values(ascending=False)
total = value_counts.sum()

# Create pie chart for "Does Music Help You Concentrate?"
plt.figure(figsize=(6, 6))
wedges, texts, autotexts = plt.pie(value_counts, colors=['#FFB6C6', '#FFD6D4', '#D0F1FA'],
                                   autopct=lambda pct: f"{int(round(pct * total / 100.0))}\n({pct:.1f}%)",
                                   startangle=90, textprops={'color': 'black', 'fontsize': 10, 'style': 'italic', 'weight': 'bold'})
for w in wedges: w.set_edgecolor('black')

plt.legend(value_counts.index, bbox_to_anchor=(1, 0.5), handleheight=2)
plt.title("Does Music Help You Concentrate?", fontsize=14, weight='bold')
plt.axis('equal')
plt.tight_layout()
plt.show()
```

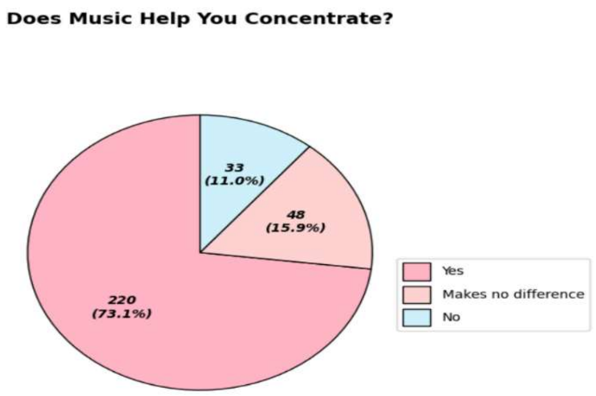

Figure 3.35 How music affect concentration

A significant majority of participants (73.1%) believe that music helps them concentrate. Meanwhile, 15.9% feel it makes no difference, and only 11.0% reported that music does not help their focus. This data clearly highlights the positive association between music and concentration among the surveyed individuals.

vi. Productivity

A pie chart was used to show how music affects productivity based on self-reported data. Unique participant IDs were counted for accuracy.

```
# Group by Productivity and count unique Res_IDs
pro_counts = df.groupby('Productivity')['Res_ID'].nunique().sort_values(ascending=False)
total = pro_counts.sum()

# Create pie chart for "Productivity With or Without Music?"
plt.figure(figsize=(6, 6))
wedges, texts, autotexts = plt.pie(pro_counts, colors=['#88d8b0', 'lightgrey'],
                                   autopct=lambda pct: f"{int(round(pct * total / 100.0))}\n({pct:.1f}%)",
                                   textprops={'color': 'black', 'fontsize': 10, 'style': 'italic', 'weight': 'bold'})
for w in wedges: w.set_edgecolor('black')
plt.title("Productivity With or Without Music?", fontsize=14, weight='bold')
plt.legend(pro_counts.index, bbox_to_anchor=(1, 0.5), handleheight=2)
plt.axis('equal')
plt.show()
```

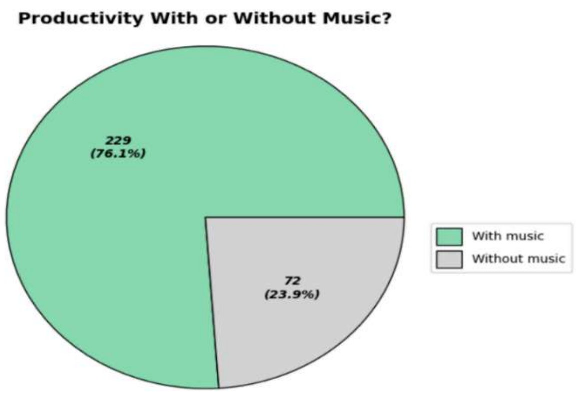

Figure 3.35 Music effect on productivity

76.1% of people (229 respondents) feel more productive with music. Only 23.9% (72 respondents) prefer working without music. A large majority find music enhances their productivity

vii. Focus Duration Strategies

A pie chart was generated to explore how participants use music to maintain focus over longer periods. Unique responses were grouped by their preferred method, ensuring accuracy by counting only distinct participant IDs.

```
# Step 1: Count unique Res_IDs per Long_Duration category
duration_counts = df.groupby('Long_Duration')['Res_ID'].nunique().sort_values(ascending=False)
total = duration_counts.sum()

# Step 2: Re-fetch value counts (if needed for plotting proportions)
duration_counts = df['Long_Duration'].value_counts()

# Step 3: Clean up category names
duration_counts.index = duration_counts.index.str.replace('Music with breaks in between', 'With Breaks')
duration_counts.index = duration_counts.index.str.replace('Music playing the entire time', 'Entire Time')
duration_counts.index = duration_counts.index.str.replace('No music at all', 'No Music')

# Step 4: Plot pie chart
plt.figure(figsize=(6, 4))
wedges, _, _ = plt.pie(
    duration_counts,
    autopct=lambda pct: f"{int(round(pct * total / 100.0))}\n({pct:.1f}%)",
    startangle=140,
    colors=plt.cm.Oranges(range(60, 60 + len(duration_counts)*40, 40)))
# Step 5: Add legend
plt.legend(
    wedges,
    duration_counts.index,
    loc="center left",
    bbox_to_anchor=(1, 0.5),
    handleheight=2)
# Step 6: Final touches
plt.title('How to Use Music to Stay Focused Longer', weight='bold')
plt.axis('equal')  # Equal aspect ratio ensures pie is circular
plt.tight_layout()
plt.show()
```

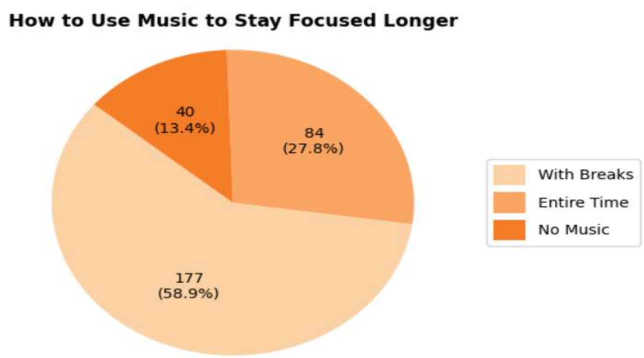

Figure 3.37 Focus duration strategies

The majority (58.9%) reported using music with breaks to stay focused, while 27.8% listened to music the entire time. A smaller portion (13.4%) preferred no music at all. These insights highlight that strategic use of music, especially incorporating breaks, is a popular method for sustaining focus.

viii. Stress Reduction

This pie chart visualizes participants' self-reported experiences with music as a tool for stress relief, using distinct respondent IDs for accurate representation.

```
# Count the number of responses for each category in 'Stress_relief' column
stress_counts = df.groupby('Stress_relief')['Res_ID'].nunique().sort_values(ascending=False)
total = stress_counts.sum()

plt.figure(figsize=(6, 4))
plt.pie(
    stress_counts,
    colors=['#86CFBF', '#C6F5E9', '#F7FCE6'],
    autopct=lambda pct: f"{int(round(pct * total / 100.0))}\n({pct:.1f}%)",
    startangle=90,
    textprops={
        'color': 'black',
        'fontsize': 8,
        'style': 'italic'
    }
)
# Add black border to wedges
for w in plt.gca().patches:
    w.set_edgecolor('grey')

plt.legend(stress_counts.index, bbox_to_anchor=(1, 0.5),handleheight=2)
plt.title('Stress Reduction Through Music', weight='bold')
plt.tight_layout()
plt.show()
```

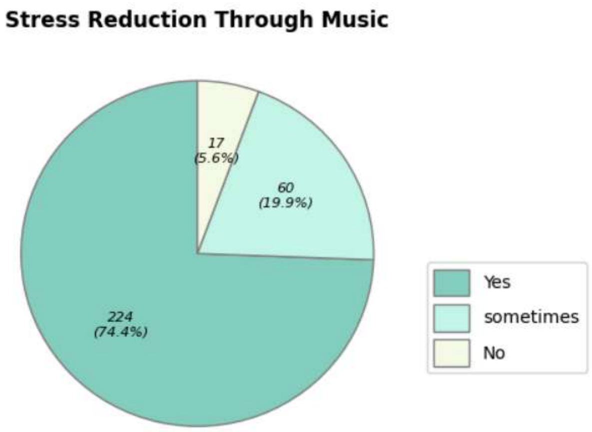

Figure 3.38 Stress relief

A strong majority (74.4%) indicated that music helps reduce their stress, while 19.9% reported it helps sometimes. Only 5.6% said that music does not help them with stress reduction. These findings highlight music’s perceived effectiveness as a stress management aid for most individuals.

ix. Critical Tasks

This pie chart highlights how individuals perceive music’s influence on productivity during critical tasks, based on distinct respondent data.

```python
critical_counts = df.groupby('Critical_productivity')['Res_ID'].nunique().sort_values(ascending=False)
total = critical_counts.sum()

plt.figure(figsize=(6, 4))

# Show percentage on top and count below it
plt.pie(
    critical_counts,
    colors=['#49C8F5','#AFE1FC', '#E2F3FD', '#FFFEF9'],
    autopct=lambda pct: f"{int(round(pct * total / 100.0))}\n({pct:.1f}%)",
    startangle=90,
    textprops={'color': 'black', 'fontsize': 8, 'style': 'italic', 'weight':'bold'}
)

for w in plt.gca().patches:
    w.set_edgecolor('black')

plt.legend(critical_counts.index, bbox_to_anchor=(1, 0.5), handleheight=2)
plt.title("Music's Role in Critical Tasks", weight='bold')
plt.tight_layout()
plt.show()
```

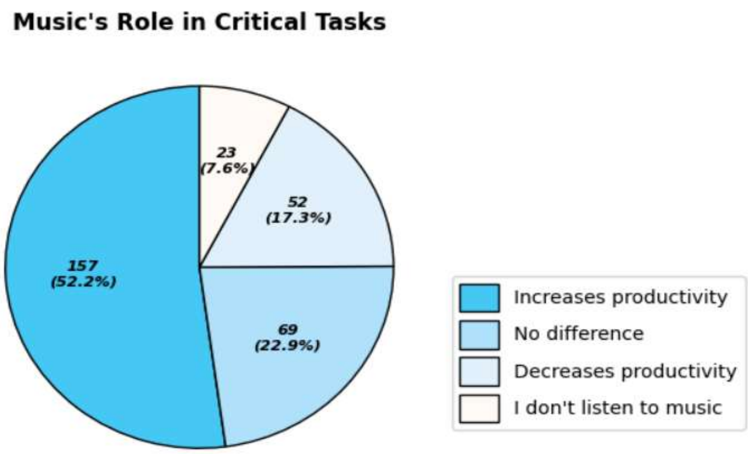

Figure 3.39 music’s influence on productivity during critical tasks

More than half of the participants (52.2%) believe that music increases their productivity during such tasks. Meanwhile, 22.9% feel it decreases productivity, 17.3% said it makes no difference, and only 7.6% don’t listen to music at all. These results show that for many, music plays a positive role even during high-focus activities.

x. Activity Preferences

This horizontal bar chart explores the preferred contexts in which individuals choose to engage with music.

```
activity_counts = df.groupby('activity_list')['Res_ID'].nunique().sort_values(ascending=False)
total = activity_counts.sum()

plt.figure(figsize=(10, 6))
plt.barh(activity_counts.index, activity_counts, color='lightcoral')  # Light red color

max_count = activity_counts.max()
plt.xlim(0, max_count * 1.2)  # Adds whitespace on the right for text

# Add labels to each bar showing the count and percentage
for i, v in enumerate(activity_counts):
    label = f'{v} ({v/total:.1%})'  # Format as "count (percent)"
    plt.text(v + 5, i, label, va='center', ha='left', fontsize=11)  # Position text right of the bar

plt.title("Activity Preferences", fontsize=14, weight = 'bold')

# Remove x-axis label
plt.xlabel("")

# Set font size of y-axis (activity names)
plt.yticks(fontsize=12)

# Adjust layout so everything fits nicely
plt.tight_layout()
plt.show()
```

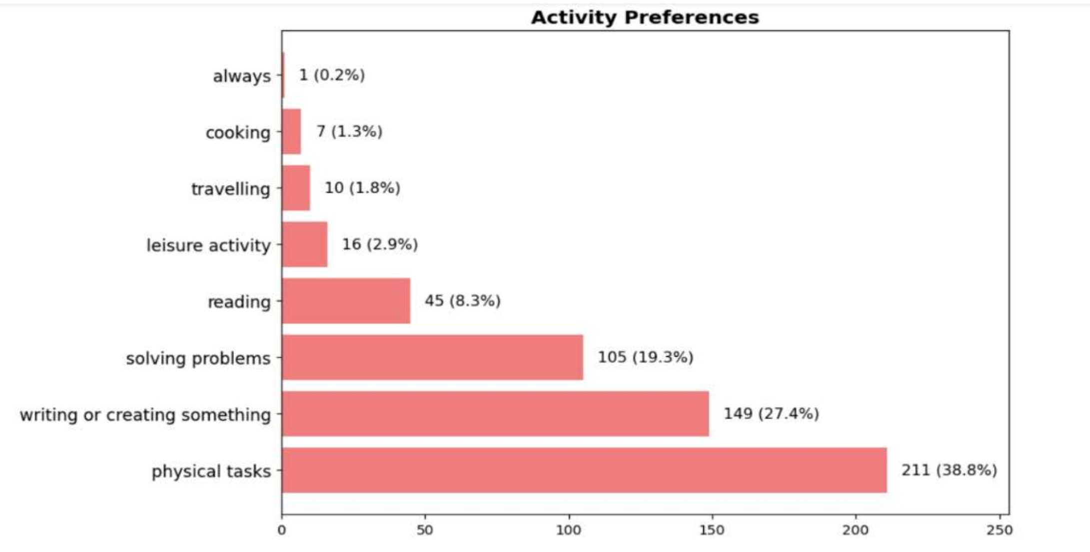

Figure 3.40 Activity preference of an individual

The most common activity paired with music is physical tasks, with 38.8% of respondents indicating it as their top choice. This is followed by writing or creating something (27.4%) and solving problems (19.3%). Other preferences like reading (8.3%), leisure activities (2.9%), and traveling (1.8%) had smaller followings. Very few people listen to music always (0.2%) or while cooking (1.3%). These results suggest that music is most often associated with dynamic or creative tasks rather than passive ones.

xi. Genre Preferences

```
Genre_pre_counts = df.groupby('Genre_Preference')['Res_ID'].nunique().sort_values(ascending=False)

# Calculate the total number of responses
total = Genre_pre_counts.sum()

plt.figure(figsize=(10, 6))

# Create a horizontal bar plot (barh)
bars = plt.barh(Genre_pre_counts.index, Genre_pre_counts, color='#00bfff')  # Light blue color

# Add count and percentage labels inside each bar
for i, (genre, Genre_pre_counts) in enumerate(Genre_pre_counts.items()):
    pct = Genre_pre_counts / total  # Calculate percentage
    label = f'{Genre_pre_counts} ({pct:.1%})'  # Create label text
    text_color = 'black'  # Set text color

    # Position the label near the end of each bar (adjusted by -5 for padding)
    plt.text(Genre_pre_counts - 5, i, label, va='center', ha='right', fontsize=10, color=text_color)


plt.title("Genre Preferences", fontsize=14, weight= 'bold')

# Remove x-axis label
plt.xlabel("")

plt.yticks(fontsize=12)

# Automatically adjust layout to fit all elements
plt.tight_layout()
plt.show()
```

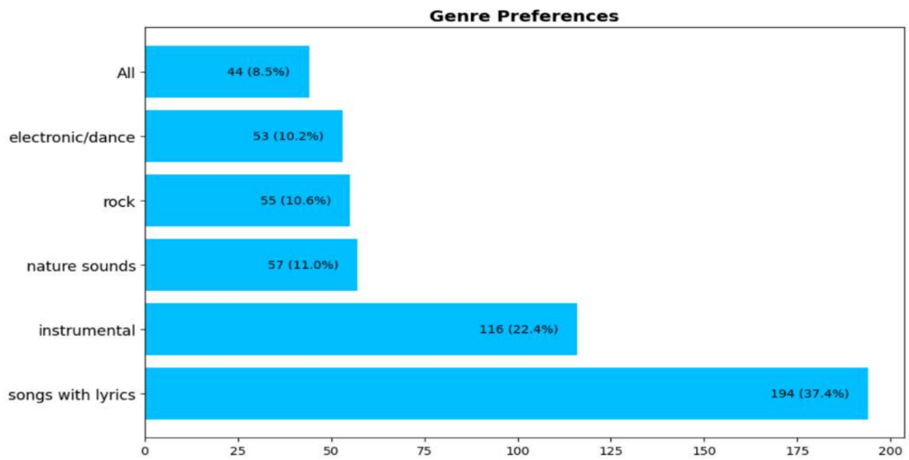

Figure 3.41 Genre preference

- Songs with lyrics lead the chart significantly, chosen by 37.4% of respondents.
- Instrumental music follows at 22.4%, indicating strong support for wordless melodies often favoured during focus or relaxation.
- Nature sounds (11.0%), rock (10.6%), and electronic/dance (10.2%) are almost equally preferred.
- A smaller segment opted for All genres (8.5%).

This shows a notable tilt toward lyrical content, but also a solid appreciation for instrumental and ambient options depending on mood or activity.

3.3 Correlation Analysis

Correlation Analysis helps us understand the relationship between two variables whether they move together or not.

- To see if one variable affects another (e.g., Age vs. Music_Volume).
- This script computes Spearman's rank correlation between variables and visualizes the results using a heatmap.

i. **Encoding Categorical Survey Data into Numeric Format for Analysis**

- This script encodes categorical survey responses into numeric values using predefined mappings, enabling easier analysis and machine learning on the transformed DataFrame (df_encoded) *[Seaborn Heatmap – A comprehensive guide, March 29, 2025]*.

```
# Age range in increasing order
age_map = {'Under 18': 1,'18-25': 2,'26-35': 3,'36 above': 4}

# Duration ranking from least to most
duration_map = {'Rarely': 1,'Sometimes': 2,'Often': 3,'Always': 4}

# Music_Volume ranking from low to high
Music_Volume_map = {'Low/Soft': 1,'Moderate': 2,'High': 3}

# Help_concentrate ordinal map
help_map = {'No': -1,'Makes no difference': 0,'Yes': 1}

# Long_Duration ordinal map
long_duration_map = {'No music at all': -1,
                     'Music with breaks in between': 0,
                     'Music playing the entire time': 1}

# Stress_relief binary map
stress_map = {'No': -1,'sometimes': 0,'Yes': 1}

# Critical productivity ranking
critical_prod_map = {'Decreases productivity' : -1,
                     "I don't listen to music": -2 ,
                     'No difference' : 0,
                     'Increases productivity': 1}

# Binary mapping for Productivity
productivity_map = {'Without music': -1,'With music': 1}

# Apply mappings
df_encoded = df.copy()
df_encoded['Age'] = df_encoded['Age'].map(age_map)
df_encoded['Duration'] = df_encoded['Duration'].map(duration_map)
df_encoded['Music_Volume'] = df_encoded['Music_Volume'].map(Music_Volume_map)
df_encoded['Help_concentrate'] = df_encoded['Help_concentrate'].map(help_map)
df_encoded['Long_Duration'] = df_encoded['Long_Duration'].map(long_duration_map)
df_encoded['Stress_relief'] = df_encoded['Stress_relief'].map(stress_map)
df_encoded['Critical_productivity'] = df_encoded['Critical_productivity'].map(critical_prod_map)
df_encoded['Productivity'] = df_encoded['Productivity'].map(productivity_map)
```

```
df_encoded.head()
```

	Res_ID	Age	Occupation	Duration	Music_Volume	Help_concentrate	Productivity	Long_Duration	Stress_relief	Critical_productivity	activity_list	Genre_Preference
0	U_01	2	Student	1	2	1	1	-1	1	-2	writing or creating something	All
1	U_01	2	Student	1	2	1	1	-1	1	-2	physical tasks	All
2	U_02	2	Student	2	1	0	-1	0	1	1	physical tasks	instrumental
3	U_02	2	Student	2	1	0	-1	0	1	1	physical tasks	electronic/dance
4	U_02	2	Student	2	1	0	-1	0	1	1	physical tasks	nature sounds

Figure 3.42 Encoded data

a) Exploring Variable Relationships with Spearman's Correlation

To better understand patterns in our survey data, we'll explore potential relationships between key variables. By using Spearman's rank correlation, we can assess how ordinal factors influence each other.

b) Correlation Analysis Utilities

This block includes two key functions:

- **Statistical Significance Check**: Evaluates the p-value to determine if the correlation is statistically significant.
- **Strength & Direction Interpreter**: Analyses Spearman's ρ (rho) to describe how strong and in which direction variables are related.

```
# Function for finding the Statistical significance
def p_value(p_val,rho_value):
  print(f"Spearman's ρ (rho): {rho_value:.2f}")
  print(f"P-value: {p_val:.4f}")
  if p_val < 0.05:
    print("✅ The correlation is statistically significant (p < 0.05)")
  else:
    print("❌ The correlation is NOT statistically significant (p ≥ 0.05)")

  # --- Technical Note for p ≈ 0 ---
  if pval < 0.0001:
      print("\n" + "-"*40)
      print("TECHNICAL NOTE: p-value approximation".center(40))
      print("-"*40)
      print(
    f"""1. **Interpretation**
      - The reported p-value ({pval:.4f}) indicates the true value is smaller than 0.0001.
      - This provides strong evidence against the null hypothesis.

    2. **Common Causes**
      - Large sample sizes (301 observations)
      - Highly consistent monotonic relationships

    3. **Recommendations**
      - Verify effect size (ρ = {rho:.2f}) for practical significance
      - Inspect the scatterplot for nonlinear patterns
      - Consider subgroup analyses if categorical variables exist"""
        )
```

```
# Function for find the Strength & direction (using original rho)
def interpret_correlation(rho_value):
  abs_rho = abs(rho_value)
  if abs_rho < 0.2:
      strength = "very weak"
  elif abs_rho < 0.4:
      strength = "weak"
  elif abs_rho < 0.6:
      strength = "moderate"
  elif abs_rho < 0.8:
      strength = "strong"
  else:
      strength = "very strong"
  direction = "positive" if rho_value > 0 else "negative"
  return strength, direction
```

Figure 3.43 music's influence on productivity during critical tasks

c) Is there a relationship between Age and Music Listening Volume?

We're exploring if there's a correlation between Age and Music Volume**.** This helps us understand whether listening volume preferences vary across different age groups.

```
# STEP 1: Clean the data
longsession_data = df_encoded[['Age', 'Music_Volume']].dropna()

# STEP 2: Calculate Spearman's correlation
rho, pval = spearmanr(longsession_data['Age'], longsession_data['Music_Volume'])

# Step 3: Output correlation and p-value
p_value(pval,rho)

# STEP 4: Interpret correlation strength
strength, direction = interpret_correlation(rho)
print(f"📌 Interpretation: There is a {strength} {direction} correlation between Age and Music Volumn.")
```

```
Spearman's ρ (rho): -0.02
P-value: 0.5168
❌ The correlation is NOT statistically significant (p ≥ 0.05)
📌 Interpretation: There is a very weak negative correlation between Age and Music Volumn.
```

Figure 3.44 Relationship between Age and Music Listening Volume

A Spearman's rank correlation was conducted to evaluate the relationship between participants' age and how frequently they listen to music. The analysis yielded a correlation coefficient of $\rho = -0.02$ with a p-value of 0.5168. This indicates a very weak negative correlation that is not statistically significant ($p > 0.05$).

Therefore, we cannot conclude that there is a meaningful relationship between age and music listening Music_Volume in this dataset.

d) **Is there a correlation between how helpful music is for concentration and self-reported productivity?**

We're checking if there's a correlation between how helpful music is for concentration and self-reported productivity.

```
# STEP 1: Clean the data
longsession_data = df_encoded[['Help_concentrate', 'Productivity']].dropna()

# STEP 2: Calculate Spearman's correlation
rho, pval = spearmanr(longsession_data['Help_concentrate'], longsession_data['Productivity'])

# Step 3: Output correlation and p-value
p_value(pval,rho)

# STEP 4: Interpret correlation strength
strength, direction = interpret_correlation(rho)
print(f"📌 Interpretation: There is a {strength} {direction} correlation between Help concentrate and productivity.")
```

```
Spearman's ρ (rho): 0.56
P-value: 0.0000
✅ The correlation is statistically significant (p < 0.05)

----------------------------------------
 TECHNICAL NOTE: p-value approximation
----------------------------------------
1. **Interpretation**
      - The reported p-value (0.0000) indicates the true value is smaller than 0.0001.
      - This provides strong evidence against the null hypothesis.

    2. **Common Causes**
      - Large sample sizes (301 observations)
      - Highly consistent monotonic relationships

    3. **Recommendations**
      - Verify effect size (ρ = 0.56) for practical significance
      - Inspect the scatterplot for nonlinear patterns
      - Consider subgroup analyses if categorical variables exist
📌 Interpretation: There is a moderate positive correlation between Help concentrate and productivity.
```

Figure 3.45 how helpful music is for concentration and self-reported productivity

A Spearman's rank-order correlation was conducted to examine the relationship between how helpful individuals perceive music to be for concentration and their self-reported productivity. The analysis revealed a moderate positive correlation, ρ = 0.56, with a p-value of < 0.0001, indicating a statistically significant association.

This suggests that individuals who find music helpful for concentration are more likely to report higher productivity when listening to music.

e) Is there a correlation between using music to reduce stress and feeling more productive?

We're exploring the correlation between using music for stress relief and feeling more productive.

```
# Step 1: Filter and drop NaN
stress_prod_data = df_encoded[['Stress_relief', 'Productivity']].dropna()

# Step 2: Spearman correlation
rho, pval = spearmanr(stress_prod_data['Stress_relief'], stress_prod_data['Productivity'])

# Step 3: Output correlation and p-value
p_value(pval,rho)

# STEP 4: Interpret correlation strength
strength, direction = interpret_correlation(rho)

print(f"📌 Interpretation: There is a {strength} {direction} correlation between stress relief through music and productivity.")
```

```
Spearman's ρ (rho): 0.20
P-value: 0.0000
✅ The correlation is statistically significant (p < 0.05)

----------------------------------------
 TECHNICAL NOTE: p-value approximation
----------------------------------------
1. **Interpretation**
      - The reported p-value (0.0000) indicates the true value is smaller than 0.0001.
      - This provides strong evidence against the null hypothesis.

   2. **Common Causes**
      - Large sample sizes (301 observations)
      - Highly consistent monotonic relationships

   3. **Recommendations**
      - Verify effect size (ρ = 0.20) for practical significance
      - Inspect the scatterplot for nonlinear patterns
      - Consider subgroup analyses if categorical variables exist
📌 Interpretation: There is a weak positive correlation between stress relief through music and productivity.
```

Figure 3.46 Music to reduce stress and feeling more productive

A Spearman's rank-order correlation was conducted to examine the relationship between using music to relieve stress and perceived productivity. The results revealed a weak positive correlation, ρ = 0.20, with a p-value of < 0.0001, indicating that the relationship is statistically significant.

This suggests that individuals who use music as a means to reduce stress may also tend to report slightly higher productivity levels.

f) Is there a relationship between preferred volume level and perceived helpfulness of music for concentration?

We're examining if the preferred music volume and the perceived helpfulness of music for concentration are related. [Music_Volume + help_concentrate as numeric/ordinal].

```
# Step 1: Filter relevant columns and drop NaNs
volume_help_data = df_encoded[['Music_Volume', 'Help_concentrate']].dropna()

# Step 2: Run Spearman correlation
rho, pval = spearmanr(volume_help_data['Music_Volume'], volume_help_data['Help_concentrate'])

# Step 3: Output correlation and p-value
p_value(pval,rho)

# STEP 4: Interpret correlation strength
strength, direction = interpret_correlation(rho)

print(f"📌 Interpretation: There is a {strength} {direction} correlation between volume level and help concentration.")
```

```
Spearman's ρ (rho): -0.08
P-value: 0.0122
✅ The correlation is statistically significant (p < 0.05)
📌 Interpretation: There is a very weak negative correlation between volume level and help concentration.
```

Figure 3.47 preferred volume level and perceived helpfulness of music for concentration

A Spearman's rank-order correlation was performed to assess the relationship between individuals' preferred volume level and how helpful they find music for concentration. The analysis yielded a very weak negative correlation, $\rho = -0.08$, with a p-value of 0.0122, indicating the result is statistically significant.

Although significant, the weak negative relationship suggests that as volume preference slightly increases, the perceived helpfulness of music for concentration may slightly decrease but the effect is minimal.

g) Does music used in long sessions correlate with self-rated focus levels?

We're analyzing whether the genre or style of music used during long listening sessions is correlated with self-rated focus levels.

By encoding long-duration music types and comparing them with productivity ratings, we can uncover potential patterns.

```
# STEP 1: Clean the data
longsession_data = df_encoded[['Long_Duration', 'Productivity']].dropna()

# STEP 2: Calculate Spearman's correlation
rho, pval = spearmanr(longsession_data['Long_Duration'], longsession_data['Productivity'])

# Step 3: Output correlation and p-value
p_value(pval,rho)

# STEP 4: Interpret correlation strength
strength, direction = interpret_correlation(rho)

print(f"📌 Interpretation: There is a {strength} {direction} correlation between long duration and productivity.")
```

```
Spearman's ρ (rho): 0.43
P-value: 0.0000
✅ The correlation is statistically significant (p < 0.05)

----------------------------------------
 TECHNICAL NOTE: p-value approximation
----------------------------------------
1. **Interpretation**
      - The reported p-value (0.0000) indicates the true value is smaller than 0.0001.
      - This provides strong evidence against the null hypothesis.

    2. **Common Causes**
      - Large sample sizes (301 observations)
      - Highly consistent monotonic relationships

    3. **Recommendations**
      - Verify effect size (ρ = 0.43) for practical significance
      - Inspect the scatterplot for nonlinear patterns
      - Consider subgroup analyses if categorical variables exist
📌 Interpretation: There is a moderate positive correlation between long duration and productivity.
```

Figure 3.49 Type of music used in long sessions correlate with self-rated focus levels

A Spearman's rank-order correlation was conducted to explore the relationship between the type of music used during long sessions and self-reported productivity. The analysis showed a moderate positive correlation, ρ = 0.43, with a p-value of < 0.0001, indicating the result is statistically significant.

This suggests that individuals who use music during longer work or study sessions tend to feel more productive, although the effect is not very strong.

h) Is there a relationship between age and how often you listen to music?

We're exploring if age is related to how often individuals listen to music (measured by duration in hours or frequency).

This analysis will help understand if music listening habits vary across different age groups.

```
# Step 1: Select and clean the data
age_duration_data = df_encoded[['Age', 'Duration']].dropna()

# Step 2: Run Spearman correlation
rho, pval = spearmanr(age_duration_data['Age'], age_duration_data['Duration'])

# Step 3: Output correlation and p-value
p_value(pval,rho)

# STEP 4: Interpret correlation strength
strength, direction = interpret_correlation(rho)
print(f"📌 Interpretation: There is a {strength} {direction} correlation between age and duration of music listening.")
```

```
Spearman's ρ (rho): -0.12
P-value: 0.0002
✅ The correlation is statistically significant (p < 0.05)
📌 Interpretation: There is a very weak negative correlation between age and duration of music listening.
```

Figure 3.50 How often individuals listen to music

A Spearman's rank-order correlation was conducted to examine the relationship between age and how often individuals listen to music. The results revealed a very weak negative correlation, ρ = -0.12, with a p-value of 0.0002, indicating the correlation is statistically significant.

This suggests that as age slightly increases, the duration of music listening tends to slightly decrease though the relationship is minimal.

i) Spearman Correlation Matrix with Heatmap Visualization

A Spearman correlation matrix reveals how encoded survey variables relate to each other. The heatmap highlights the strength and direction of these relationships visually.

i. Correlation Matrix

The correlation matrix displays Spearman coefficients between numeric survey variables, showing the strength and direction of relationships between factors like age, music habits, and productivity.

```
# Select only numeric (int/float) columns
numeric_df = df_encoded.select_dtypes(include=["int64", "float64"])
# Calculate the correlation matrix
correlation_matrix = numeric_df.corr()
correlation_matrix
```

	Age	Duration	Music_Volume	Help_concentrate	Productivity	Long_Duration	Stress_relief	Critical_productivity
Age	1.000000	-0.154563	-0.050644	-0.104010	-0.053668	-0.100284	-0.151697	-0.159845
Duration	-0.154563	1.000000	-0.001143	0.245640	0.365649	0.339630	0.389416	0.347725
Music_Volume	-0.050644	-0.001143	1.000000	-0.046403	0.079892	0.008881	0.032981	-0.054837
Help_concentrate	-0.104010	0.245640	-0.046403	1.000000	0.511408	0.314374	0.183315	0.370943
Productivity	-0.053668	0.365649	0.079892	0.511408	1.000000	0.441982	0.207939	0.420814
Long_Duration	-0.100284	0.339630	0.008881	0.314374	0.441982	1.000000	0.139117	0.334002
Stress_relief	-0.151697	0.389416	0.032981	0.183315	0.207939	0.139117	1.000000	0.218332
Critical_productivity	-0.159845	0.347725	-0.054837	0.370943	0.420814	0.334002	0.218332	1.000000

Figure 3.51 Correlation Matrix

ii. **Heatmap Visualization**

This heatmap visualizes Spearman correlation coefficients between encoded survey variables. The colour scale indicates the strength and direction of relationships from strong negative (blue) to strong positive (red).

```
# Plot heatmap
plt.figure(figsize=(8, 7))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f", linewidths=0.5)
plt.title("Correlation Between Columns", weight='bold',fontsize='14')
plt.show()
plt.tight_layout()
plt.show()
```

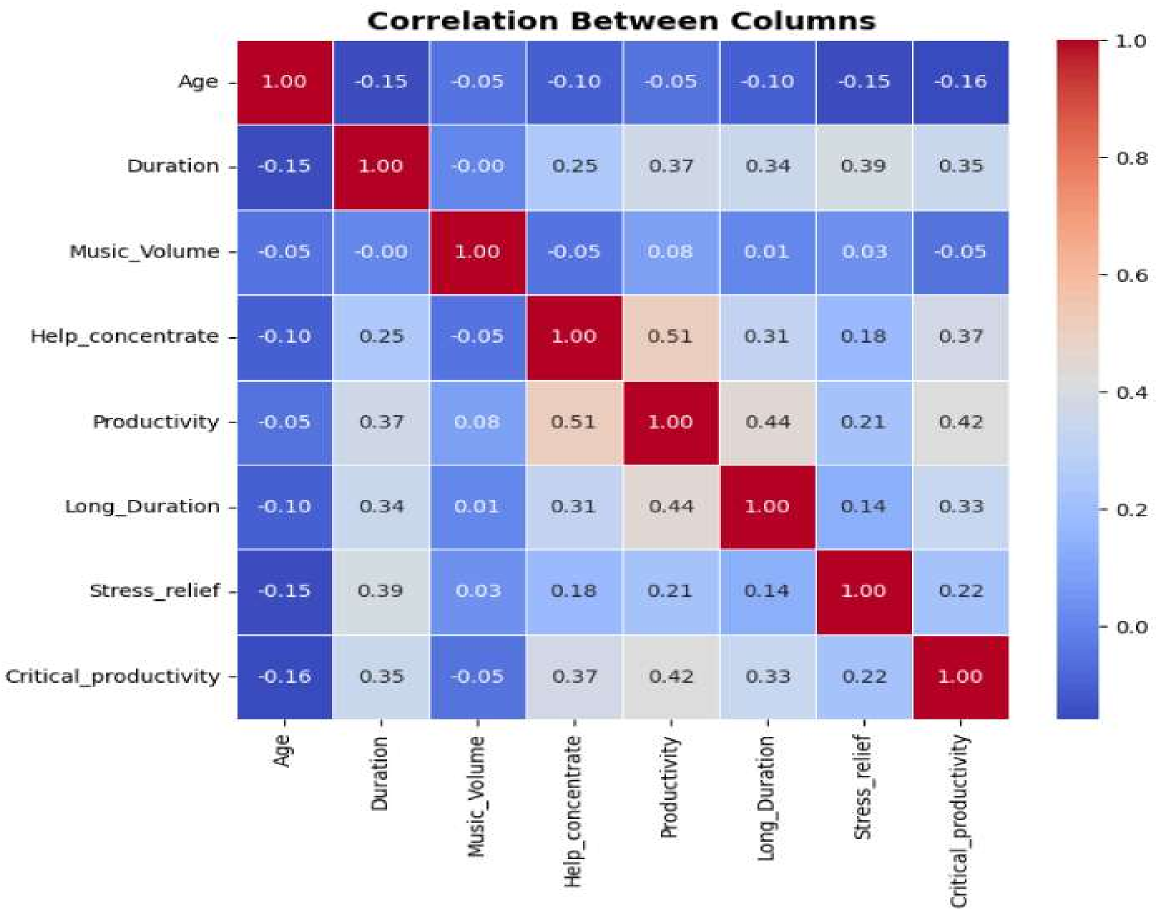

Figure 3.52 Heatmap

- Music Duration is positively linked with productivity, stress relief, and long sessions.
- Help_concentrate shows a strong positive correlation with productivity.
- Age has weak negative correlations with most music-related variables.
- Overall, music especially when it aids concentration tends to enhance productivity and well-being.

3.4 Hypothesis Testing

- **Chi-Square Test**

  This function chi_square_analysis () performs a Chi-Square test between two categorical columns, interprets the results against a critical value, and neatly prints the outcome for better understanding *[Chi Square test in Python, Mysha Rysh, 2 years ago]*.

```
def chi_square_analysis(df, col_x, col_y, title, cmap='YlGnBu'):
    # Step 1: Remove duplicates
    df_unique = df.drop_duplicates(subset='Res_ID')

    # Step 2: Create contingency table
    contingency_table = pd.crosstab(df_unique[col_x], df_unique[col_y])

    # Step 3: Perform Chi-Square Test
    chi2_stat, p_value, dof, expected = chi2_contingency(contingency_table)

    # Step 4: Critical value
    alpha = 0.05
    critical_value = chi2.ppf(1 - alpha, dof)

    # Step 5: Print results
    print(f"--- {title} ---")
    print("Chi-Square Value:", chi2_stat)
    print("Degrees of Freedom:", dof)
    print("Critical Value (alpha = 0.05):", critical_value)

    # Step 6: Interpretation
    if chi2_stat > critical_value:
        print(f"Reject H₀: There is a significant difference between {col_x} and {col_y} are associated.")
    else:
        print(f"Fail to reject H₀: There is No significant link between {col_x} and {col_y}.")
    print("\n")
```

```
# Step 7: Plot heatmap
plt.figure(figsize=(10, 6))
sns.heatmap(contingency_table, annot=True, fmt='d', cmap=cmap, cbar=True)
plt.title(f'Heatmap: {title} (Unique ID-wise)')
plt.xlabel(col_y)
plt.ylabel(col_x)
plt.tight_layout()
plt.show()
```

Figure 3.54 Function to perform Chi-square test

i. Is music helpful for focus depending on the kind of work you do?

H_0: The type of work does not affect whether music helps with focus.

H_1: The type of work affects whether music helps with focus.

```
chi_square_analysis(df, 'Occupation', 'Help_concentrate', 'Occupation vs Help Concentrate','Greens')
```

```
--- Occupation vs Help Concentrate ---
Chi-Square Value: 36.551833403219575
Degrees of Freedom: 16
Critical Value (alpha = 0.05): 26.29622760486423
Reject H₀: There is a significant difference between Occupation and Help_concentrate are associated.
```

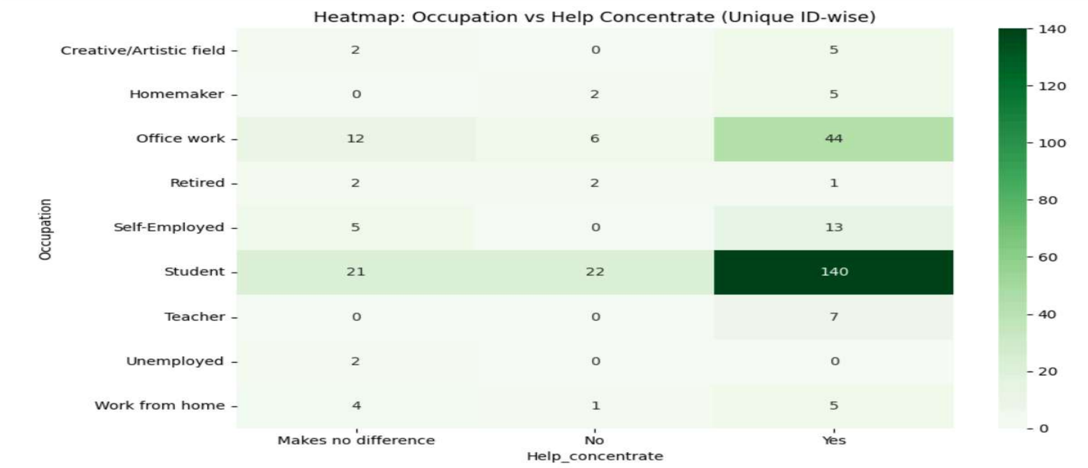

Figure 3.55 Focus depending on the kind of work you do

Conclusion:

There is a significant relationship between Occupation and Help with Concentration, as the Chi-Square value (36.55) exceeds the critical value (26.30).

ii. **How do activities influence the music genre preferences of individuals?**

H_0: No link between activity and genre

H_1: There is a link

```
# Example 2
chi_square_analysis(df, 'activity_list', 'Genre_Preference', 'Activity vs Genre Preference','BuGn')

--- Activity vs Genre Preference ---
Chi-Square Value: 21.51542231177464
Degrees of Freedom: 30
Critical Value (alpha = 0.05): 43.77297182574219
Fail to reject H₀: There is No significant link between activity_list and Genre_Preference.
```

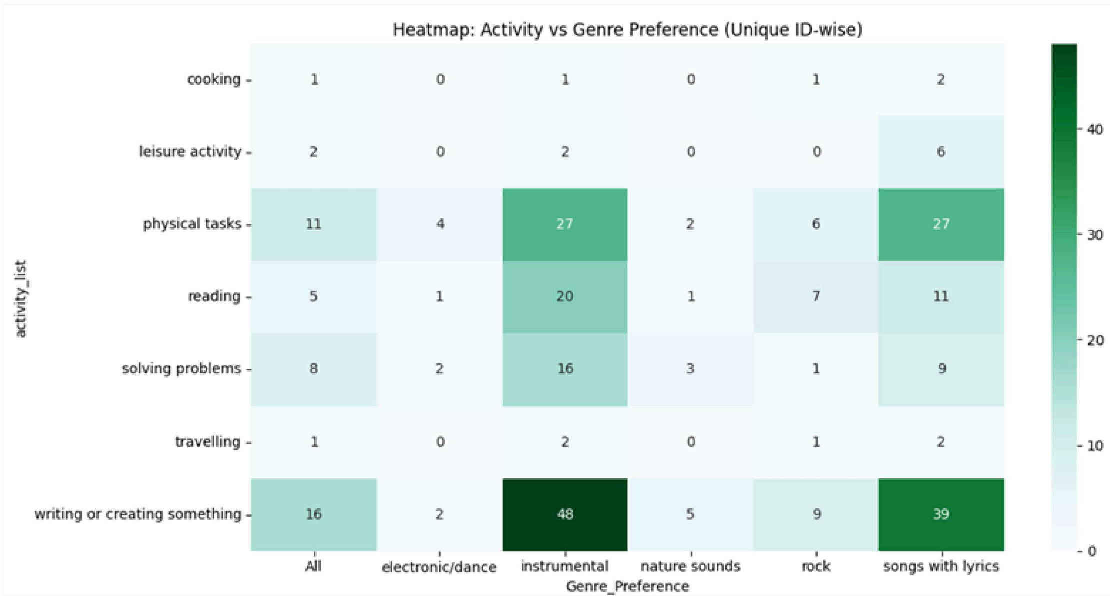

Figure 3.56 Activities influence the music genre preferences of individuals

Conclusion:

There is no significant relationship between Activity and Genre Preference, as the Chi-Square value (21.51) is less than the critical value (43.77).

iii. **How does music used in long session impact a person's productivity?**

H_0: No difference in productivity

H_1: There is a difference

```
chi_square_analysis(df, 'Productivity', 'Long_Duration', 'Productivity vs Long Duration','Oranges')
```

```
--- Productivity vs Long Duration ---
Chi-Square Value: 55.78705451974952
Degrees of Freedom: 2
Critical Value (alpha = 0.05): 5.991464547107979
Reject H₀: There is a significant difference between Productivity and Long_Duration are associated.
```

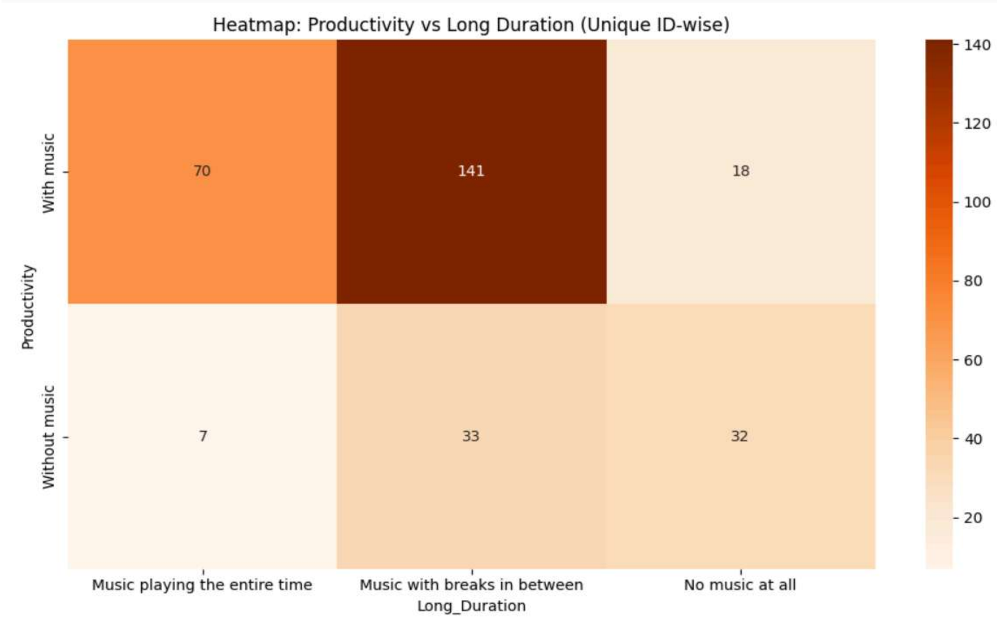

Figure 3.57 Music used in long session impact a person's productivity

Conclusion:

A significant relationship exists between Productivity and Long Duration of music listening, as the Chi-Square value (55.79) exceeds the critical value (5.99).

iv. **How does stress reduction through music listening impact an individual's productivity during work-related tasks?**

H_0: No difference in productivity

H_1: There is a difference

```
chi_square_analysis(df, 'Stress_relief', 'Critical_productivity', 'Stress Relief vs Critical Productivity','PuBu')
```

```
--- Stress Relief vs Critical Productivity ---
Chi-Square Value: 22.411779258192283
Degrees of Freedom: 6
Critical Value (alpha = 0.05): 12.591587243743977
Reject H₀: There is a significant difference between Stress_relief and Critical_productivity are associated.
```

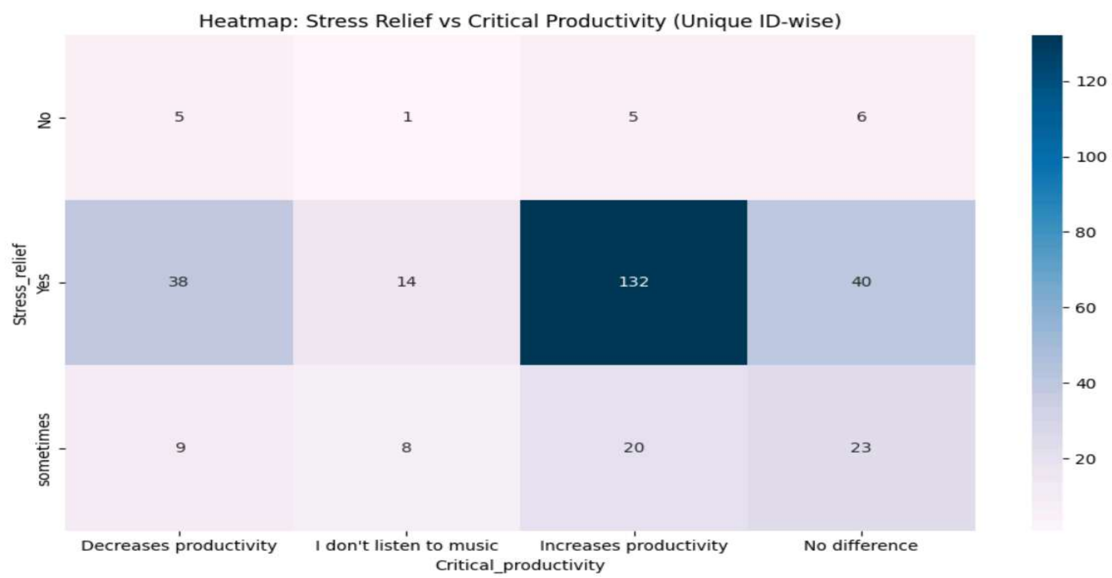

Figure 3.58 Music reduce stress and boost productivity during work

Conclusion:

There is a significant relationship between Stress Relief and Critical Productivity, as the Chi-Square value (22.41) exceeds the critical value (12.59).

v. **Do people who believe music helps them focus tend to listen to it more frequently?**

H_0: No difference in Music_Volume

H_1: There is a difference

```
chi_square_analysis(df, 'Help_concentrate', 'Duration', 'Help Concentrate vs Duration','RdPu')
```

```
--- Help Concentrate vs Duration ---
Chi-Square Value: 25.380091538154
Degrees of Freedom: 6
Critical Value (alpha = 0.05): 12.591587243743977
Reject H₀: There is a significant difference between Help_concentrate and Duration are associated.
```

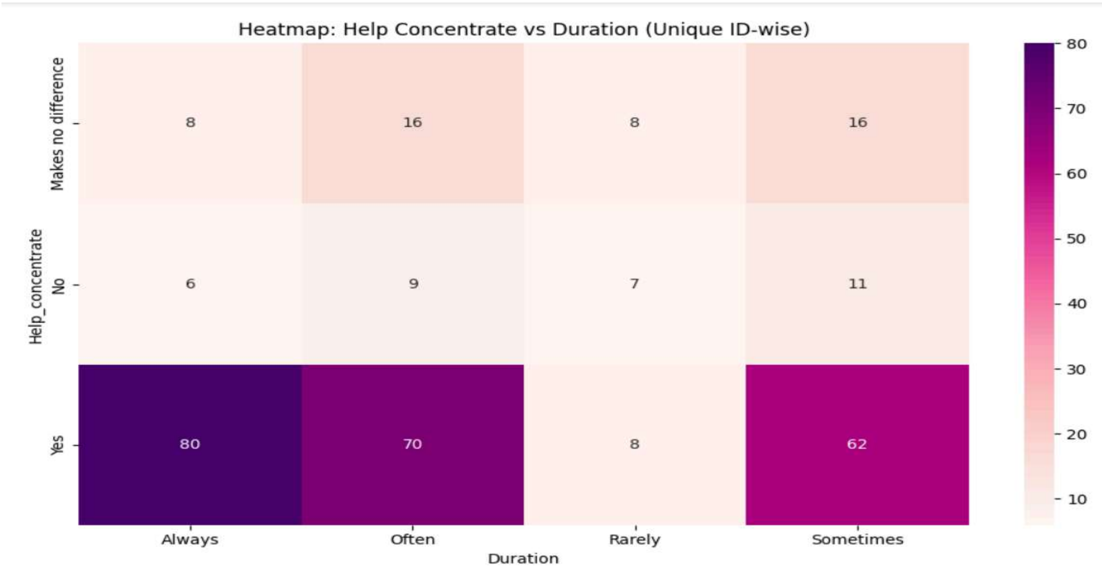

Figure 3.59 Music reduce stress and boost productivity during work

Conclusion:

A significant relationship exists between Help with Concentration and Duration of music listening, as the Chi-Square value (25.38) is greater than the critical value (12.59).

CHAPTER 4

CONCLUSION

CHAPTER IV
CONCLUSION

4.1 Conclusion

This study concludes that music can positively influence productivity, especially when matched to individual preferences and task demands. It improves focus, supports mental effort, reduces stress, and enhances performance—particularly in repetitive or physically engaging activities.

Key findings include:

- Music is helpful during work, with 73.1% reporting improved concentration, especially those in occupations demanding sustained focus.
- Lyrical music can be distracting for reading or writing, while instrumental or ambient music is more effective for deep-focus tasks.
- Longer listening sessions and using music for stress relief significantly boost productivity.
- Music's effectiveness varies by occupation, confirming that job type influences how beneficial music is during tasks.
- Statistical analysis through Chi-Square testing confirmed significant relationships between productivity, listening duration, occupation, and stress relief.

4.2 Suggestions

Based on the findings, the following are recommended:

- Use instrumental or ambient music for reading, writing, or studying.
- Avoid lyrical or fast-paced music during language-intensive or complex tasks.
- Take music breaks during long work sessions; listening in intervals is more effective than continuous playback.
- Allow flexibility in music choices at workplaces or educational settings, as occupation affects how music supports concentration.
- Integrate stress-relieving music into routines, as it helps reduce anxiety and improves performance.
- Customize music environments based on job roles and task types for maximum benefit.

REFERENCE

REFERENCE

- Levitin, D. J. (2006**).** *This Is Your Brain on Music: The Science of a Human Obsession , accessed – February 10, 2025* https://www.amazon.com/This-Your-Brain-Music-Obsession/dp/0452288525

- Lesiuk, T. (2005). The effect of music listening on work performance. *Psychology of Music, 33*(2), 173–191, accessed – February 14, 2025 https://doi.org/10.1177/0305735605050650

- Schellenberg, E. G. (2005). Music and cognitive abilities. *Current Directions in Psychological Science, 14*(6), 317–320, accessed - February 15, 2025 https://doi.org/10.1111/j.0963-7214.2005.00389.x

- Kämpfe, J., Sedlmeier, P., & Renkewitz, F. (2011). *The impact of background music on adult listeners: A meta-analysis.* Psychology of Music, accessed – February 20, 2025 https://doi.org/10.1177/0305735610376261

- Furnham, A., & Strbac, L. (2002). Music is as distracting as noise: The differential distraction of background music and noise on the cognitive test performance of introverts and extraverts. *Ergonomics, 45*(3), 203–217, accessed – February 23, 2025 https://doi.org/10.1080/00140130210121932

- Perham, N., & Currie, H. (2014). *Does listening to preferred music improve reading comprehension performance? Applied Cognitive Psychology, 28*(2), 279–284, accessed – March 2, 2025 https://doi.org/10.1002/acp.2994

- Anderson, C. A., & Fuller, D. R. (2010). The effects of music on task performance. *Journal of Undergraduate Psychological Research, 5*, 42–47.

- Gupta, S. C., & Kapoor, V. K. (1987). *Fundamentals of mathematical statistics*. S. Chand & Sons.

- Moorthy, M. N. (1988). *Sampling theory and method*. Statistical Publishing Society.

- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering, 9*(3), 90–95.

- North, A. C., & Hargreaves, D. J. (2008). *The social and applied psychology of music*. Oxford University Press.

- Hallam, S., Price, J. & Katsarou, G. (2002). The effects of background music on primary school pupils’ task performance. *Educational Studies, 28*(2), 111–122, accessed – March 7, 2025, https://doi.org/10.1080/03055690220124551

- McKinney, W. (2010). Data structures for statistical computing in Python. Proceedings of the 9th Python in Science Conference (pp. 51–56), accessed–March 12, 2025, Retrieved from https://doi.org/10.25080/Majora-92bf1922-00a

- Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology, 84*(6), 1236–1256, accessed – March 17, 2025 https://doi.org/10.1037/0022-3514.84.6.1236

- Microsoft Corporation. (2024). *Microsoft Excel* (Version 16.0). Retrieve from https://office.microsoft.com/excel

- Python Software Foundation. (2024). *Python language reference* (Version 3.11). Retrieved from https://www.python.org

- The Pandas Development Team. (2024). *pandas: Python data analysis library*, accessed – April 8, 2025. Retrieved from https://pandas.pydata.org/

- Seaborn Development Team. (2024). *Seaborn: Statistical data visualization, accessed - April 15, 2025,* https://seaborn.pydata.org/

- Seaborn Heatmap – A comprehensive guide , Last updated – March 29 , 2025, accessed – April 20 , 2025, https://www.geeksforgeeks.org/seaborn-heatmap-a-comprehensive-guide/

- Chart Visualization, Pandas Documentation, Date – September 20,2024, Version – 2.2.3,accessed – April 23, 2025, https://pandas.pydata.org/docs/user_guide/visualization.html#plotting-directly-with-matplotlib

- Hypothesis Testing: Chi Square test in Python , Mysha Rysh, 3 years ago, accessed – May 4, 2025, https://www.kaggle.com/code/mysha1rysh/chi-square-test-in-python/notebook

APPENDIX

APPENDIX

ANALYSIS ON "EXPLORING THE RELATIONSHIP BETWEEN MUSIC, CONCENTRATION, AND TASK PERFORMANCE"

QUESTIONNAIRE

1) **Age**
   a) Under 18
   b) 18–25
   c) 26–35
   d) 36 above
2) **Occupation**
   a) Student
   b) Office work
   c) Work from home
   d) Creative/Artistic field
   e) Self-employed
   f) Other
3) **How often do you listen to music while doing something?**
   a) Always
   b) Often
   c) Sometimes
   d) Rarely
   e) Never
4) **What activities do you do while listening to music? (Pick all that apply)**
   a) Reading
   b) Writing or creating something
   c) Solving problems (math, coding, puzzles, etc.)
   d) Physical tasks (exercise, cleaning, etc.)
   e) Other
5) **What kind of music do you enjoy the most? (Pick all that apply)**
   a) Instrumental (Classical, Lo-Fi, Jazz, etc.)
   b) Rock

c) Electronic/Dance

d) Songs with lyrics (any genre)

e) Nature sounds (rain, ocean waves)

f) No preference

6) **How loud do you like your music?**

a) Low/Soft

b) Moderate

c) High

7) **Does music help you concentrate?**

a) Yes

b) No

c) Makes no difference

8) **Do you think some types of music make it harder to concentrate?**

a) Yes

b) No

c) Maybe

9) **Do you feel more productive ________?**

a) With music

b) Without music

10) **If you had to focus for 3+ hours, which would you prefer?**

a) Music playing the entire time

b) Music with breaks in between

c) No music at all

11) **Do you listen to music to reduce stress?**

a) Yes

b) No

c) Sometimes

12) **Does music affect productivity while performing critical tasks?**

a) Increases productivity

b) No difference

c) Decreases productivity

d) I don't listen to music