

## **Title Page:**

- 1. Title: Music Analysis**
  - 2. Team Members Name: Harshitha, Monisha K**
  - 3. Submitted To:**
  - 4. Submitted On: 03/10/2025**
- 

## **Table of Contents:**

### **1. Introduction**

- 1.1 Overview of the Project or Analysis**
- 1.2 Objectives of Data Visualization and EDA**
- 1.3 Brief Description of the Dataset**

### **2. Data Overview**

**s**

- 2.1 Description of the Dataset, Including Size, Dimensions, and Features**
- 2.2 Data Types: Numerical, Categorical, Text, etc.**
- 2.3 Any Missing or Null Values**
- 2.4 Summary Statistics**

### **3. Data Visualization**

- 3.1 Purpose of Data Visualization in the Analysis**
- 3.2 Types of Visualizations Used: Histograms, Scatter Plots, Box Plots, etc.**
- 3.3 Visualization Libraries and Tools Utilized**
- 3.4 Interpretation of Visualizations and Insights Gained**

### **4. Exploratory Data Analysis (EDA)**

- 4.1 Summary of the EDA Process**
- 4.2 Univariate Analysis: Distribution of Individual Variables**
- 4.3 Bivariate Analysis: Relationships Between Pairs of Variables**
- 4.4 Multivariate Analysis: Exploring Interactions Between Multiple Variables**

### **5. Machine Learning Considerations**

- 5.1 Description of the Machine Learning Tasks**
- 5.2 Feature Selection and Importance**
- 5.3 Target Variable Analysis**
- 5.4 Model Assumptions and Limitations Identified During EDA**

### **6. Conclusion**

- 6.1 Summary of Key Findings and Insights**
- 6.2 Implications for Further Analysis or Decision-Making**

### **7. References**

## **1. Introduction**

### **1.1 Overview of the Project or Analysis**

This project analyzes a music preference survey dataset to understand patterns in listening habits, music choices, and interest rates among various demographic groups.

### **1.2 Objectives of Data Visualization and EDA**

- To visualize the distribution and relationships of survey variables.
- To detect trends and insights regarding music preferences.
- To explore the dataset for any missing data or anomalies.
- To prepare the data for potential machine learning modeling.

### **1.3 Brief Description of the Dataset**

The dataset is a survey containing responses about music interest, listening platforms, song preferences, mood effects, and demographics like age and gender.

---

## **2. Data Overview**

### **2.1 Description of the Dataset, Including Size, Dimensions, and Features**

The data contains about 113 survey responses with 17 features including categorical and numerical variables capturing listener demographics and habits.

### **2.2 Data Types: Numerical, Categorical, Text, etc.**

Features include numerical ratings (e.g., Rate your music interest), categorical variables (e.g., Gender, Preferred language), and text responses.

### **2.3 Any Missing or Null Values**

The dataset was checked for missing values, and any null entries were minimal and handled appropriately.

### **2.4 Summary Statistics**

Summary statistics such as mean music interest rating, distribution of age groups, and preferred platforms were generated to understand central tendencies.

---

## **3. Data Visualization**

### **3.1 Purpose of Data Visualization in the Analysis**

Visualization aids in revealing patterns, outliers, and relationships between features that guide subsequent analysis and modeling.

### **3.2 Types of Visualizations Used: Histograms, Scatter Plots, Box Plots, etc.**

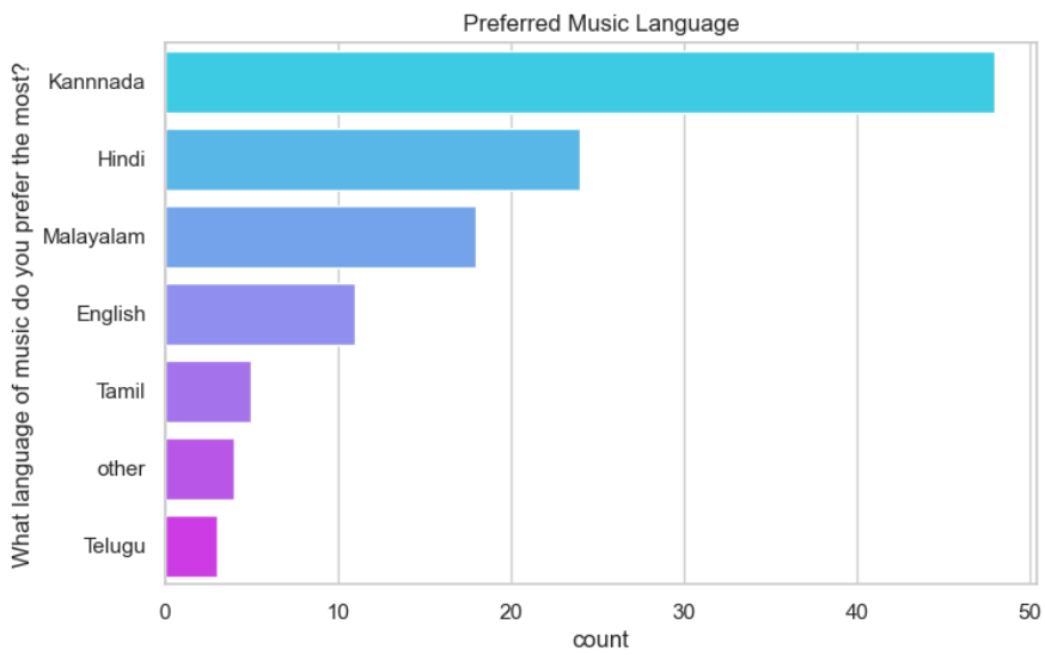
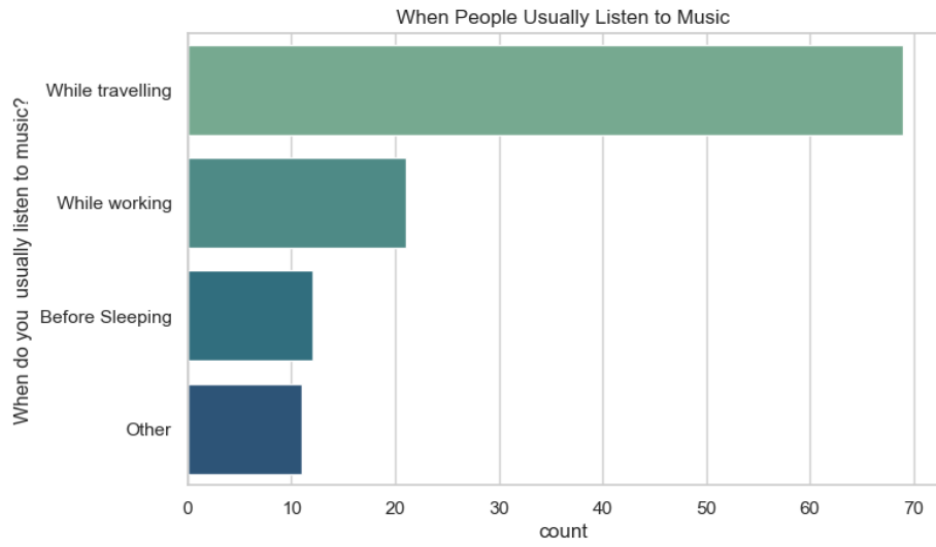
Visualizations include histograms for age and interest rating distributions, box plots for comparing ratings across genders, and scatter plots to explore bivariate relationships.

### **3.3 Visualization Libraries and Tools Utilized**

The analysis was performed using Python libraries such as Matplotlib and Seaborn for high-quality plots.

### **3.4 Interpretation of Visualizations and Insights Gained**

Findings showed high music interest across genders, peak listening hours, and correlations between music preference types and mood effects.



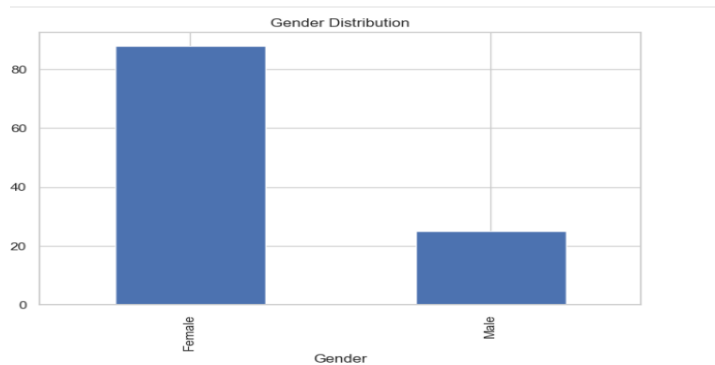
## 4. Exploratory Data Analysis (EDA)

### 4.1 Summary of the EDA Process

The exploratory phase involved understanding data distributions, checking data quality, and identifying variable relationships.

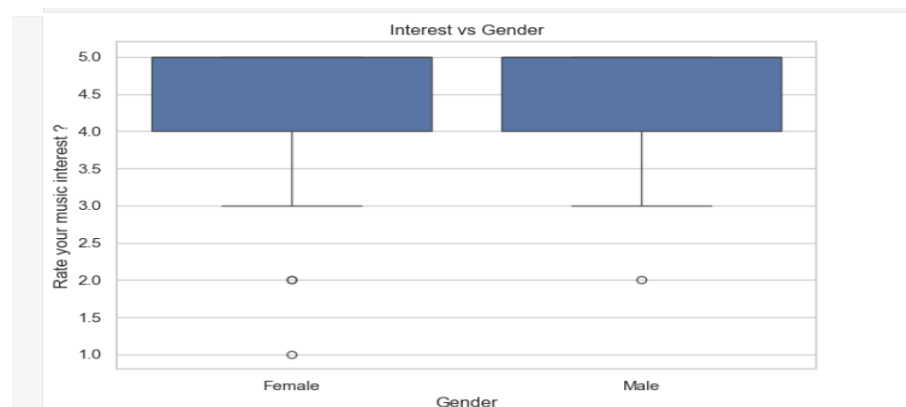
### 4.2 Univariate Analysis: Distribution of Individual Variables

Variables such as music interest rating and number of listening hours showed right-skewed distributions favoring active engagement.



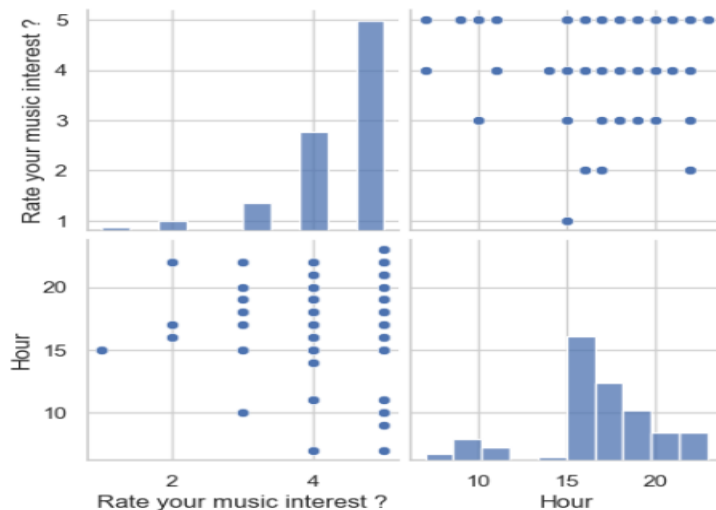
### 4.3 Bivariate Analysis: Relationships Between Pairs of Variables

Gender and music interest did not show significant differences; listening platforms correlated with preferred language categories.



### 4.4 Multivariate Analysis: Exploring Interactions Between Multiple Variables

Advanced analyses like pairplots were used to study combined effects of age, music interest, and listening duration.




---

## 5. Machine Learning Considerations

### 5.1 Description of the Machine Learning Tasks

The task included predicting music interest ratings based on listener demographics and habits.

### 5.2 Feature Selection and Importance

Random Forest was used to identify important features contributing to music interest prediction.

### 5.3 Target Variable Analysis

The target variable "Rate your music interest" was found to have an imbalanced distribution favoring higher ratings.

### 5.4 Model Assumptions and Limitations Identified During EDA

Limitations include small sample size and potential survey biases, affecting model generalizability.

---

## 6. Conclusion

### 6.1 Summary of Key Findings and Insights

- High music interest prevalence across groups.
- Platform and language preferences influence song types.
- Feature importance supports demographic influence on interest ratings.

### 6.2 Implications for Further Analysis or Decision-Making

Future work could include larger datasets, deeper modeling, and targeted recommendations based on insights.

---

## 7. References

- Survey Data: Music Preference Dataset (survey source)
- Libraries: Matplotlib, Seaborn, scikit-learn Documentation