

# **MUSIC & MENTAL HEALTH SURVEY RESULTS**

**KELOMPOK 4**

# DATASET

Dataset "**Music xMental Health Survey Results**" mengeksplorasi hubungan antara preferensi musik dan kesehatan mental seseorang. Musik dikenal dapat mempengaruhi suasana hati, mengurangi stres, dan membantu mengatasi gangguan mental seperti kecemasan dan depresi.

Dataset ini memiliki 736 entri dari berbagai responden dengan 33 fitur (kolom), mencakup informasi umum responden, preferensi terhadap genre musik, serta kondisi kesehatan mental seperti Anxiety, Depression, Insomnia, dan OCD yang dinilai dalam skala 0-10.

Link Dataset:

<https://www.kaggle.com/datasets/catherinerasgaitis/mxmh-survey-results/data>

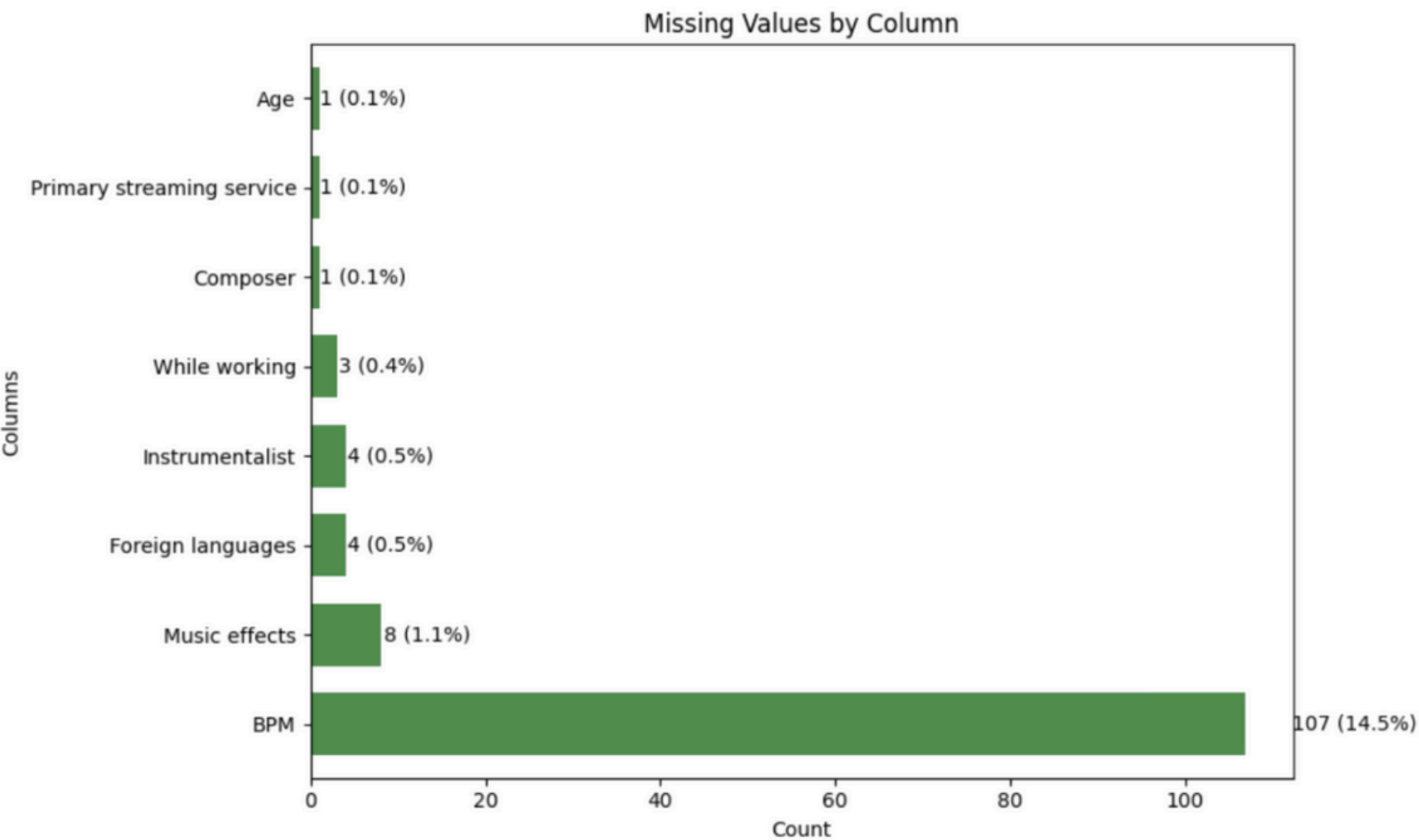
# DATA LOADING

Dataset yang digunakan yaitu "**mxmh\_survey\_results.csv**". Proses memuat dataset ini dilakukan dengan memanfaatkan library kagglehub untuk mendownload dataset ini langsung dari platform kaggle dan membacanya dengan bantuan library pandas dari python.

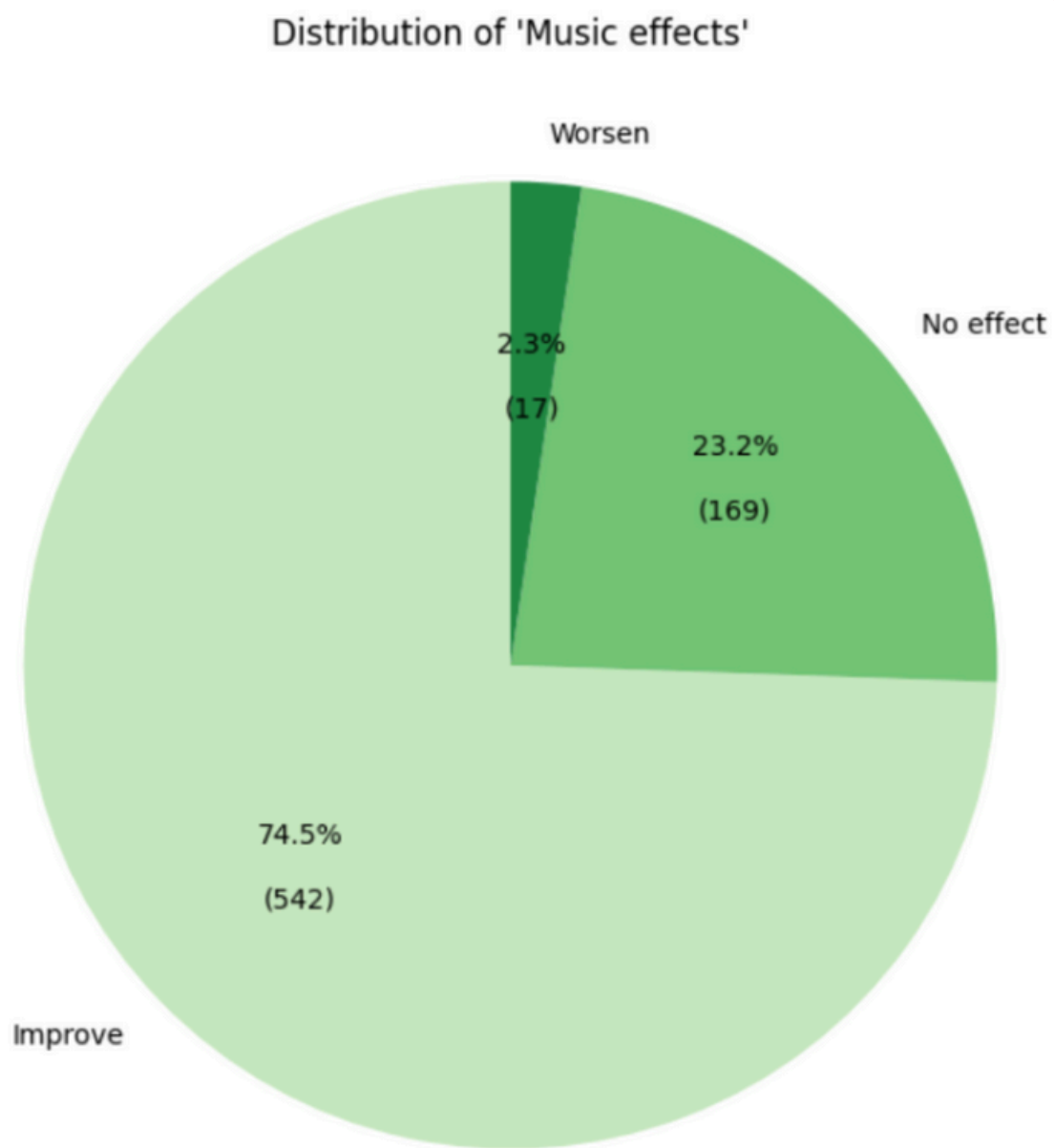
```
1 path = kagglehub.dataset_download("catherinerasgaitis/mxmh-survey-results")
2
3 data = pd.read_csv(path + "/mxmh_survey_results.csv")
4 data.head(10)
```

# DATA UNDERSTANDING

## MISSING VALUES



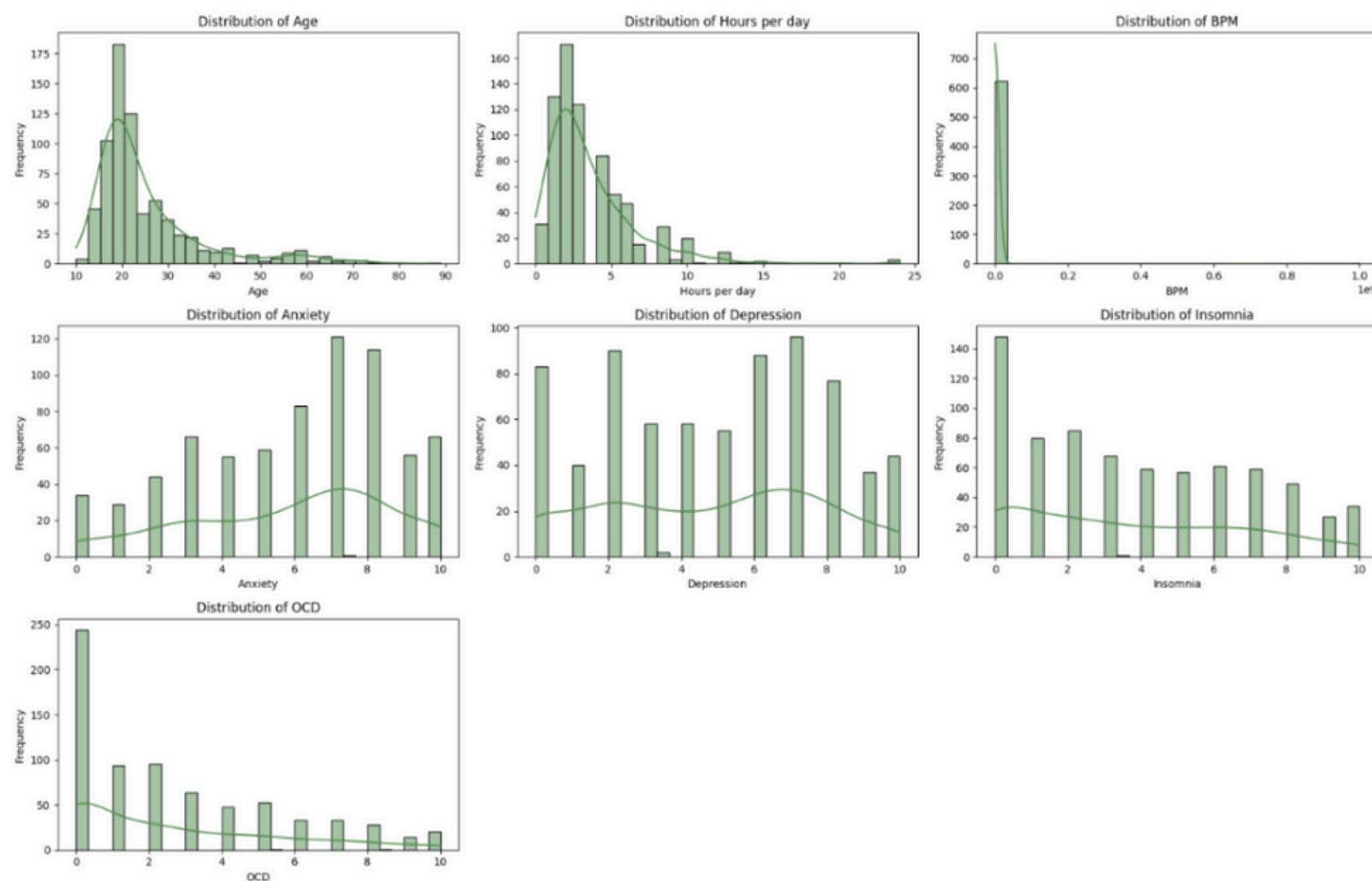
## DISTRIBUSI KELAS



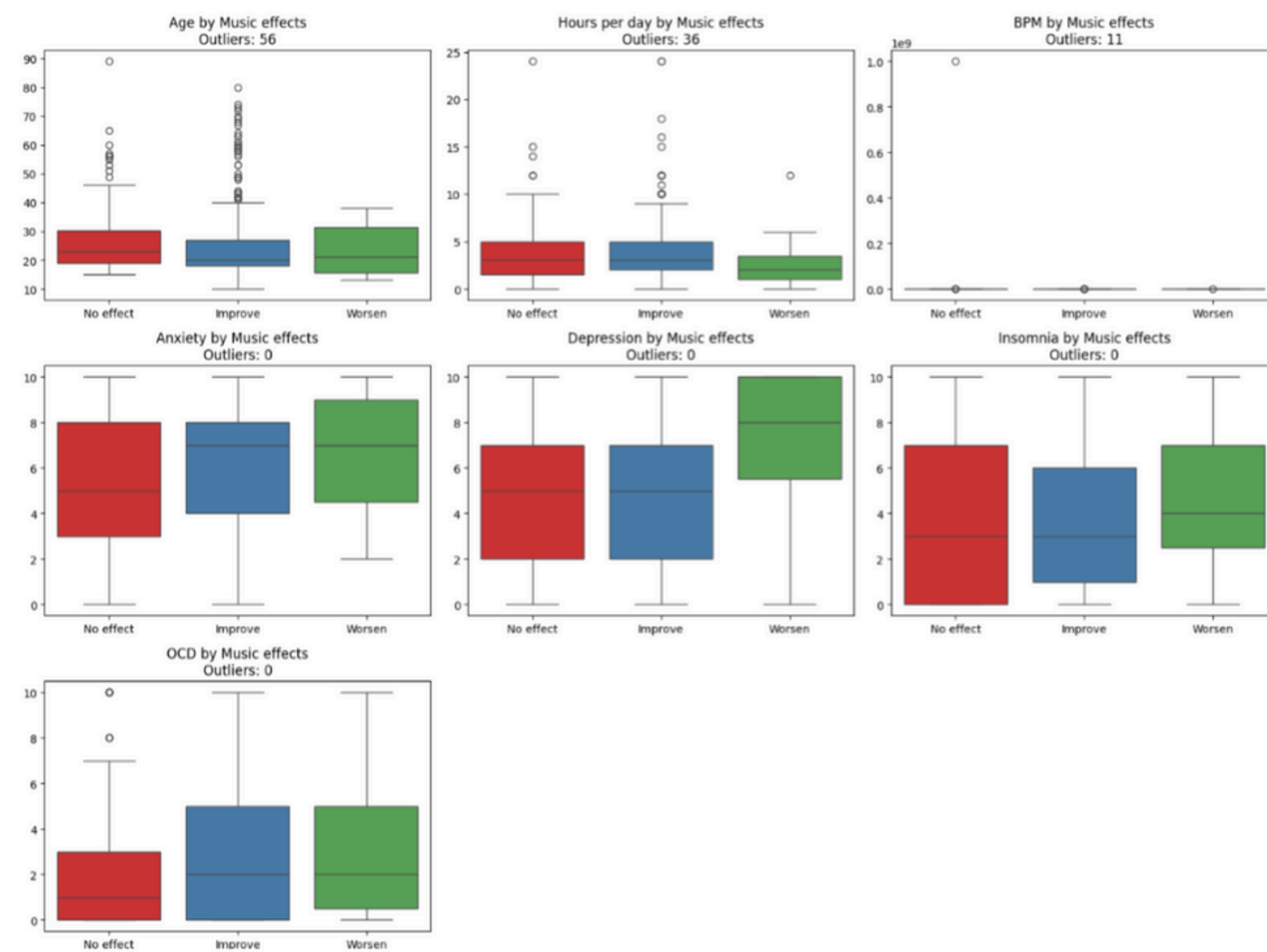
# DATA UNDERSTANDING

## NUMERICAL FEATURE DISTRIBUTION

Each Columns Distribution by Histogram



## OUTLIERS





## Missing Values Handling

- **Fitur Age**

**MEAN IMPUTATION** berdasarkan rata-rata usia responden dengan genre musik favorit yang sama.

- **Fitur BPM**

**MODE IMPUTATION** berdasarkan genre musik favorit sesuai penjelasan dataset, setelah mengatasi anomali

- **Fitur Foreign languages, While working, Instrumentalist, dan Primary streaming service**

**MODE IMPUTATION** yaitu mengganti missing values dengan nilai yang paling banyak muncul

- **Fitur Target (Music effects)**

Missing values dihapus, karena fitur ini merupakan target utama dalam analisis.

02

## Encoding Categorical Features

**Label Encoding**, efisien dalam menghemat memori dan cocok untuk model yang dapat menangani variabel kategorikal secara numerik.

03

## Feature Selection atau Extraction

Timestamp dihapus karena hanya metadata waktu submit, sementara Permissions dihapus karena memiliki variansi nol (hanya satu nilai unik).