Music habits and mental disorders

Agata Belczyk

15 June 2023

1 Introduction

Main objective: To investigate the relationship between music habits, mental disorders and well-being. Additionally, to explore how different music preferences are interrelated.

Research Objectives

- To investigate the association between music listening and mental disorders to understand which music preferences may worsen or improve mental well-being. The study results can contribute to the prevention of mental disorders and provide insights into the scientific basis of music therapy.

- To examine whether music preferences influence artistic creativity. By analyzing this correlation, we can discover whether individuals engaged in music-related activities tend to prefer specific music genres such as classical or rock.

- To do a demographic analysis of the data to explore the relationships between age, openness to news and the number of hours listened to. The goal of such research is to gain a better understanding of music preferences and listening habits based on different factors. This information can be valuable for music creators, streaming platforms, music marketers, and researchers studying music preferences and their impact on diverse demographic groups.

- To investigate which music genres require a greater openness to novelty, with a specific focus on classical music and metal, which are often considered extreme genres. The study aims to understand whether there is a relationship between the characteristics of these genres and the need for exploration and discovery of new musical compositions.

2 Data sources

Data have been obtained from:

- page **kaggle.com**[1]

- a **private survey**[2]. The survey was conducted via Google Forms by the author among students. It was completed by 68 people.

3 Data exploration

3.1 Data contents

3.1.1 Features

The data includes the following features:

- Age: Respondents' age

- Primary streaming service: Which streaming platform do they use for listening?

- Hours per day: How many hours do they listen per day?

- While working: Do they listen while working?

- Instrumentalist: Do they play a musical instrument?

- Composer: Do they compose music?

- Fav genre: What is their favorite genre?

- Exploratory: Do they actively search for new music?

- Foreign languages: Do they listen to music in a language they are not fluent in?

- BPM: What is the frequency (BPM) of their favorite genre?

- Frequency[genre]: How often do they listen to a specific genre – 16 genres, including classical?

- Mental disorder: On a scale of 1-10, how much do they experience mental disorders?

- Music effects: Do they believe that music has an impact on their mental health?

For questions regarding frequency of listening, respondents could choose from options: Never; Rarely; Sometimes; Very frequently. For the question about the effects of music, the options were: No effect, Worsen, Improve. Among the mental disorders mentioned were: Anxiety, Depression, Insomnia, OCD.

---

[1]https://www.kaggle.com/datasets/catherinerasgaitis/mxmh-survey-results
[2]https://forms.gle/5opzcfa1jmFEETjV7

3.1.2 Data presentation

The figure 1). displays the number of answers for the example survey questions. Outliers in this data set have been identified as individuals over 75 years old and have been excluded from the analysis. Also feature "BPM" have been excluded, because of large number of nulls.
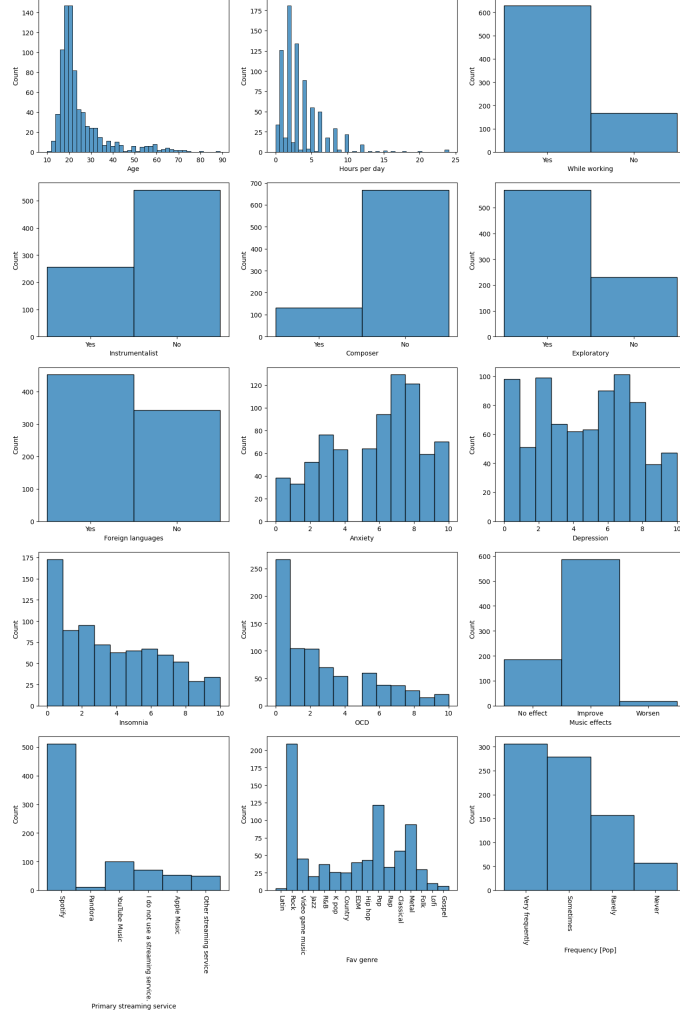

Figure 1: Subplots showing number of answers of example survey questions

3.2 Artistic activity

The survey consisted of questions about artistic activity: if respondent plays an instrument and if they are a composer. We check if there is any connection
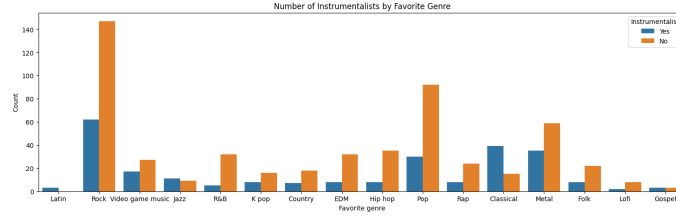
between favorite music genre and being an artist.

Figure 2: Number of instrumentalists per favourite genre

In the figure 2). we can observe that groups which stand out from others are 'Jazz','Classical'. They have more Instrumentalists than Non-Instrumentalists. In group 'Gospel' these numbers are equal, and in group 'Latin' there are only Instrumentalists, but it shouldn't be took into account, because of small 'Count' value.
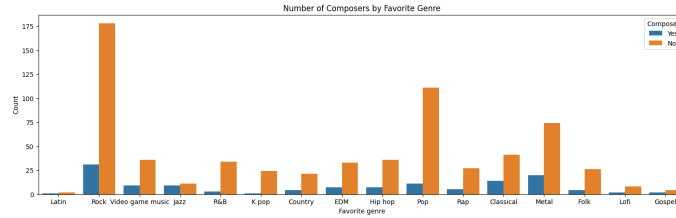
Figure 3: Number of composers per favourite genre

In 3). in all groups there is more non-composers, than composers. We can observe that Jazz bars are approximately similar in size within a group. This characteristic reflects the essence of Jazz music, which is improvisation. Bars in Latin and Gospels are also similar, but they won't be took into account, because of small 'Count' value.

To sum up the most artistic genre turned out to be Jazz in both features. In Classical music group is a lot of instrumentalist, but most of them are non-composers.

3.3 Exploratory vs favourite genre

Now we check which favourite genres are characterized by being less open by listeners.

In the figure 4). most groups display, that members of them are exploratory, except Gospel. Small difference between bars in groups occurs in Classical and Video game music. This may indicate the conservation of individuals of these genres.
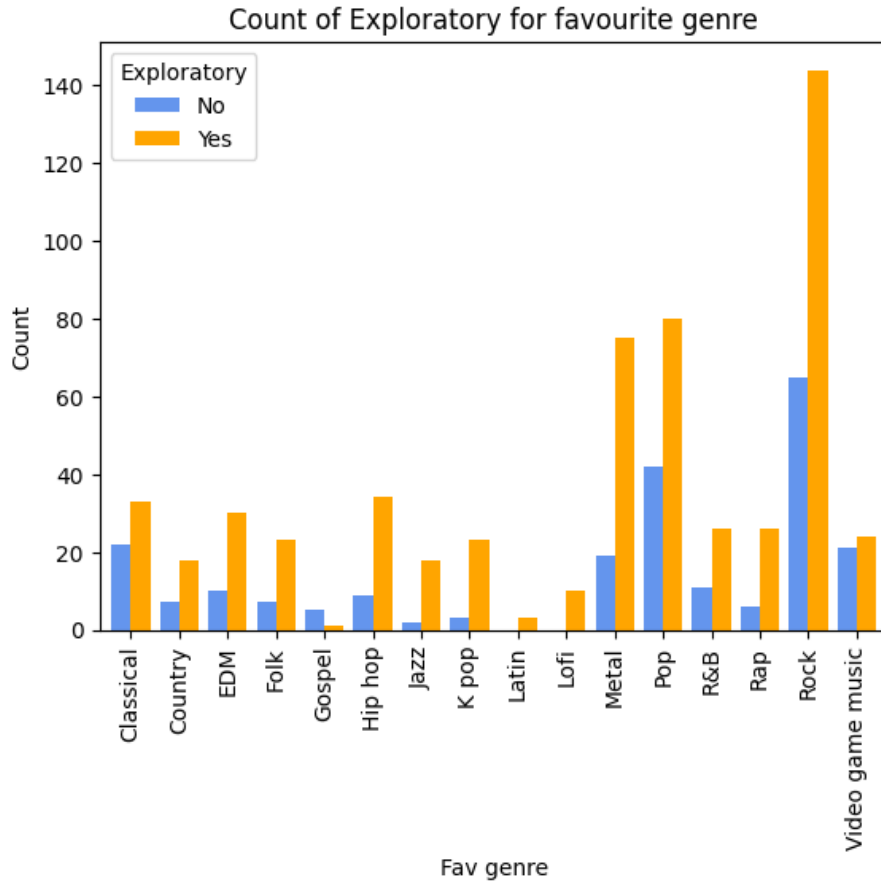
Figure 4: Count of exploratory and non-exploratory of listeners per genre

3.4 Hours of listening and mental illness level

Now we will check if there is a correlation between level of mental illness and hours of daily listening to music. The plot will display the average number of hours for each disease, along with the corresponding confidence intervals.
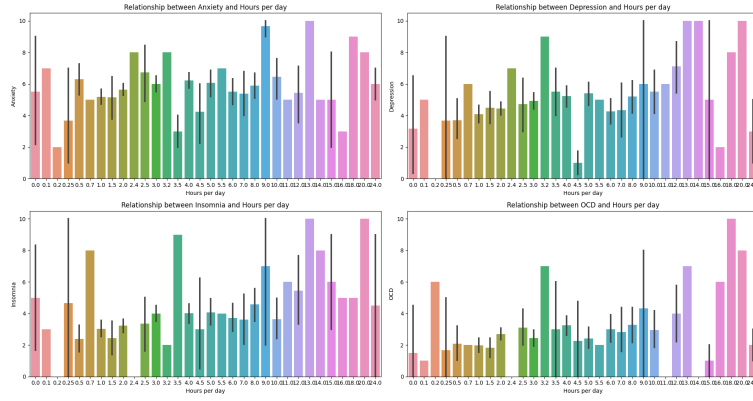
Figure 5: Association between Daily Listening Hours and Levels of Mental Disorders: Anxiety, Depression, Insomnia, OCD

As we can observe, the topmost bars predominate on the right side of plots. We can conclude that individuals with a high level of mental illness tend to listen to more hours of music. However, it is worth noting that the bars on the right side of the Depression and OCD plots are very low. Therefore, listening to many hours of music does not necessarily indicate a high level of mental illness.

The next observation that can be made by comparing sum of the heights of the bars in the charts is that anxiety has the highest number of people affected, while OCD has the lowest.

3.5 Level of anxiety and music effects

Respondents could answer if music improves, worsens or has no effect on their well-being. We check if music effect is connected with anxiety level.
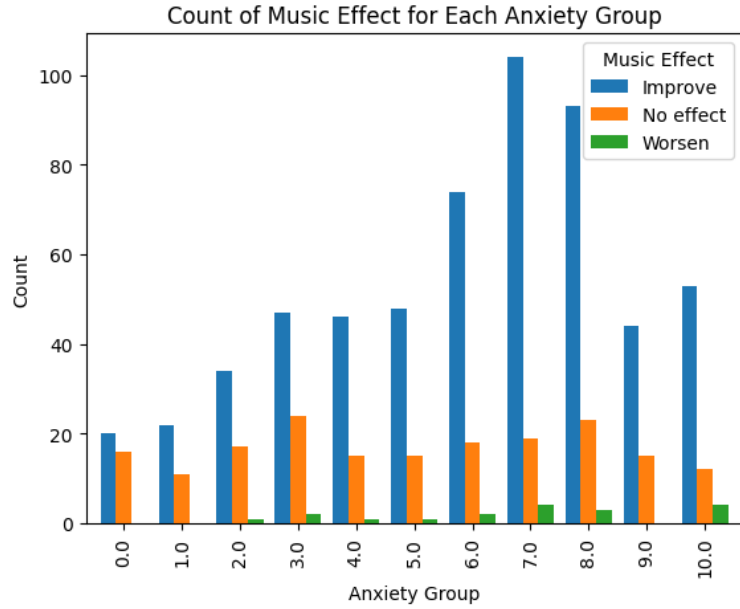
Figure 6: Correlation between music effects and level of anxiety

It can be observed that there are no 'Worsen' bars on the left side. However, starting from group 2, almost all bars have a 'Worsen' component. In group 10, the height of the 'Worsen' bar is comparable to the height of the 'Worsen' bar of 7. However, the 'Improve' bar is almost twice smaller in group 10 than in 7. Therefore, we can conclude that individuals with the highest level of anxiety may experience the most negative impact from music.

3.6 Level of anxiety vs other features

In the table there are levels of anxiety and average values of other numerical features.

	Age	Hours per day	Depression	Insomnia	OCD
Anxiety					
0.0	28.35	3.32	1.81	1.86	1.03
1.0	31.45	3.14	1.42	2.12	0.97
2.0	27.87	3.68	2.21	2.58	1.31
3.0	25.83	3.12	3.05	2.18	1.88
4.0	24.92	3.14	3.78	3.05	1.65
5.0	24.25	3.76	4.64	4.45	2.59
6.0	23.85	3.12	4.41	3.46	1.90
7.0	22.91	3.58	5.36	3.81	3.29
7.5	16.00	4.00	5.00	7.00	8.50
8.0	23.32	3.85	6.36	4.62	3.22
9.0	24.81	3.84	6.61	4.83	4.19
10.0	23.11	3.80	6.86	4.69	4.29

Figure 7: Level of anxiety with mean values of other numerical features

For anxiety, the age group with the highest prevalence of mental illness tends to be younger, specifically those under 24 years old. On the other hand, individuals around 30 years old appear to have relatively better mental well-being in terms of anxiety.

Furthermore, there is a positive correlation between the severity of anxiety and the average level of other diseases. In other words, as anxiety levels increase, so does the average level of other illnesses.

3.7 Age groups vs other features

We can also do clustering for different age groups and show mean values of numerical features and most recent for categorical features:

We can observe that Rock is not overrated, as it emerged as the most favorite genre across all age groups. Classical music is most frequently listened to by individuals in the age group of 36 and above. On the other hand, genres like Hip Hop and Rap are more popular among younger people compared to older individuals.

Furthermore, individuals between the ages of 20-30 tend to have the highest levels of mental health issues. However, there is a decreasing trend in the prevalence of mental disorders as age increases.

AgeGroup	Age	Hours per day	Anxiety	Depression	Insomnia	OCD
10-20	17.57	3.79	5.81	4.46	3.57	2.66
21-25	22.31	3.39	6.03	5.15	3.38	2.88
26-30	27.65	3.69	6.17	5.47	4.01	2.73
31-35	32.62	3.11	5.92	5.23	4.08	2.73
36-40	37.67	2.74	5.29	4.50	4.88	2.33
41-60	50.63	2.86	4.12	3.29	3.88	1.65
61-75	66.63	3.01	4.79	3.47	3.05	1.21

	Primary streaming service	While working	Instrumentalist	Composer	Fav genre	Exploratory	Foreign languages	Frequency [Classical]	Frequency [Country]	Frequency [EDM]	Frequency [Folk]	Frequency [Gospel]	Frequency [Hip hop]
10-20	Spotify	Yes	No	No	Rock	Yes	Yes	Rarely	Never	Never	Never	Never	Sometimes
21-25	Spotify	Yes	No	No	Rock	Yes	Yes	Rarely	Never	Never	Never	Never	Rarely
26-30	Spotify	Yes	No	No	Rock	Yes	Yes	Rarely	Never	Never	Rarely	Never	Sometimes
31-35	Spotify	Yes	No	No	Rock	Yes	Yes	Rarely	Never	Never	Sometimes	Never	Sometimes
36-40	Spotify	Yes	No	No	Rock	Yes	Yes	Sometimes	Never	Never	Sometimes	Never	Rarely
41-60	Spotify	Yes	No	No	Rock	Yes	No	Sometimes	Never	Never	Never	Never	Never
61-75	I do not use a streaming service.	Yes	No	No	Rock	No	No	Sometimes	Never	Never	Rarely	Never	Never

	Frequency [Jazz]	Frequency [K pop]	Frequency [Latin]	Frequency [Lofi]	Frequency [Metal]	Frequency [Pop]	Frequency [R&B]	Frequency [Rap]	Frequency [Rock]	Frequency [Video game music]	Music effects
10-20	Never	Never	Never	Never	Never	Very frequently	Never	Rarely	Very frequently	Rarely	Improve
21-25	Never	Never	Never	Never	Never	Very frequently	Never	Sometimes	Very frequently	Sometimes	Improve
26-30	Rarely	Never	Never	Never	Rarely	Sometimes	Never	Rarely	Very frequently	Rarely	Improve
31-35	Rarely	Never	Never	Never	Very frequently	Sometimes	Never	Rarely	Very frequently	Never	Improve
36-40	Rarely	Never	Never	Never	Very frequently	Sometimes	Rarely	Rarely	Sometimes	Never	Improve
41-60	Rarely	Never	Never	Never	Never	Sometimes	Sometimes	Never	Sometimes	Never	Improve
61-75	Never	Never	Sometimes	Never	Never	Sometimes	Rarely	Never	Sometimes	Never	Improve

Figure 8: Mean values and most recent categorical answers in age groups

4 Modelling

Now we will examine the possibility of predicting one feature based on others, as well as determining which features have the greatest impact on that particular feature. We will use **Decision Tree Classifier** to count impact of features and Logistic Regression and confusion matrix, because the features we try to predict are categorical.

Predicted features: Instrumentalist, Age group, Favourite genre

4.1 Predicting Instrumentalist

	Feature	Importance
0	Composer_Yes	0.403299
1	Age	0.134710
2	Frequency [Classical]_Very frequently	0.104783
3	Frequency [EDM]_Never	0.033557
4	Fav genre_Folk	0.030307
5	OCD	0.022310
6	Depression	0.019331
7	While working_No	0.017100
8	Frequency [Country]_Sometimes	0.013979
9	Primary streaming service_Spotify	0.009593

```
              precision    recall  f1-score   support

         No       0.74      0.92      0.82       111
        Yes       0.59      0.27      0.37        48

   accuracy                           0.72       159
  macro avg       0.67      0.59      0.60       159
weighted avg      0.70      0.72      0.69       159

Regression accuracy: 0.72
```

Figure 9: Results of model predicting if instrumentalist

The Decision Tree Classifier classified properly 72% of test set. The model has relatively good precision for the "No" class but poor precision for the "Yes" class:

For the "No" class: - Precision: 0.74, indicating that the model correctly classified 74% of the samples as "No" out of all the samples classified as "No". - Recall: 0.92, meaning that the model correctly identified 92% of all "No" samples relative to all true "No" samples. - F1-score: 0.82, which represents

the harmonic average of precision and recall. A higher F1-score suggests a good balance between accuracy and recall for the "No" class.

For the "Yes" class: - Precision: 0.59, indicating that the model correctly classified 59% of the samples as "Yes" out of all the samples classified as "Yes". - Recall: 0.27, meaning that the model correctly identified only 27% of all "Yes" samples relative to all true "Yes" samples. - F1-score: 0.37, which indicates a lower performance of the model in classifying the "Yes" samples.

The three most crucial features determining whether someone is an instrumentalist are: Composer Yes 0.403299 Age 0.134710 Frequency [Classical] Very frequently 0.104783

4.2 Predicting Age group

	Feature	Importance
0	Frequency [Folk]_Never	0.146794
1	Depression	0.074304
2	Exploratory_No	0.072042
3	Fav genre_Rock	0.055876
4	Primary streaming service_I do not use a strea...	0.054129
5	Primary streaming service_Pandora	0.041130
6	Fav genre_Pop	0.039430
7	Hours per day	0.038556
8	Frequency [Country]_Rarely	0.019630
9	Frequency [Classical]_Rarely	0.018845

	precision	recall	f1-score	support
10-20	0.48	0.58	0.53	72
21-25	0.23	0.43	0.30	35
26-30	0.00	0.00	0.00	22
31-35	1.00	0.00	0.00	13
36-40	1.00	0.00	0.00	3
41-60	0.33	0.14	0.20	14
61-75	1.00	0.00	0.00	1
accuracy			0.37	160
macro avg	0.58	0.16	0.15	160
weighted avg	0.40	0.37	0.32	160

Regression accuracy: 0.44

Figure 10: Results of predicting age group by the Decision Tree Classifier

The model exhibits limited capability in predicting age groups based on the available data. It performed relatively better in predicting individuals belonging to the 10-20 age group compared to other age groups. The three most crucial features determining whether someone is an age group are: - Frequency [Folk]_Never 0.146794 - Depression 0.074304 - Exploratory_No 0.072042

4.3 Predicting Favourite genre


```
                        Feature   Importance
0   Frequency [Classical]_Very frequently   0.104352
1      Frequency [EDM]_Very frequently      0.057141
2          Frequency [Hip hop]_Never        0.020785
3   Frequency [Country]_Very frequently     0.017221
4    Frequency [K pop]_Very frequently      0.013632
5             Frequency [Jazz]_Rarely       0.012793
6           Frequency [Gospel]_Never        0.011476
7       Frequency [Hip hop]_Sometimes       0.010931
8     Frequency [Folk]_Very frequently      0.009001
9          Frequency [EDM]_Rarely           0.004725

                 precision    recall  f1-score   support

      Classical       0.69      0.60      0.64        15
        Country       1.00      0.67      0.80         3
            EDM       0.00      0.00      0.00        10
           Folk       1.00      0.00      0.00         4
         Gospel       1.00      0.00      0.00         2
        Hip hop       0.06      0.29      0.11         7
           Jazz       1.00      0.00      0.00         4
          K pop       1.00      0.00      0.00         4
           Lofi       1.00      0.00      0.00         4
          Metal       0.58      0.33      0.42        21
            Pop       0.42      0.92      0.58        26
            R&B       1.00      0.00      0.00         5
            Rap       1.00      0.00      0.00         2
           Rock       0.50      0.40      0.45        47
Video game music       0.25      0.17      0.20         6

       accuracy                           0.40       160
      macro avg       0.70      0.23      0.21       160
   weighted avg       0.54      0.40      0.37       160

Regression accuracy: 0.43
```
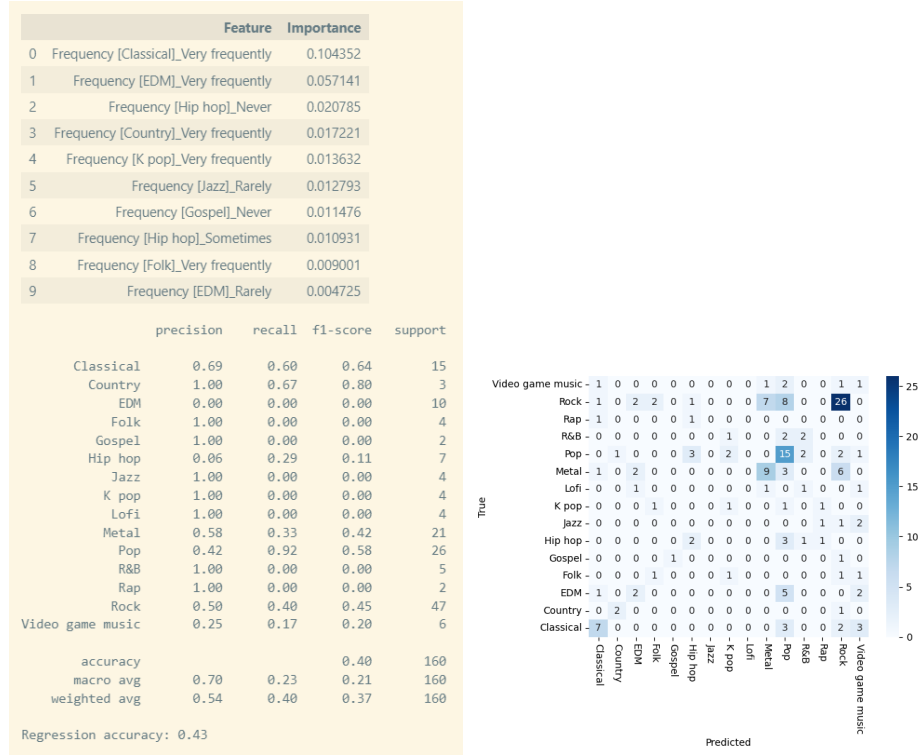
Figure 11: Results of model predicting favourite genre

The model performs poorly, but there is a diagonal outline in the confusion matrix. The model attempted to assign a significant number of samples to the rock genre instead of metal, and vice versa.

13

4.4 Better model for Favourite genre

Below there is a result of trying to find better model for predicting feature favourite genre.

```
Best Hyperparameters:
{'C': 0.05, 'kernel': 'linear'}
Model accuracy: 0.54
                    precision    recall  f1-score   support

         Classical       0.73      0.53      0.62        15
           Country       1.00      0.33      0.50         3
               EDM       0.57      0.40      0.47        10
              Folk       0.40      0.50      0.44         4
            Gospel       0.00      0.00      0.00         2
           Hip hop       0.33      0.57      0.42         7
              Jazz       0.00      0.00      0.00         4
             K pop       0.25      0.25      0.25         4
              Lofi       0.00      0.00      0.00         4
             Metal       0.60      0.57      0.59        21
               Pop       0.52      0.65      0.58        26
               R&B       0.33      0.20      0.25         5
               Rap       0.00      0.00      0.00         2
              Rock       0.66      0.74      0.70        47
  Video game music       0.20      0.33      0.25         6

          accuracy                           0.54       160
         macro avg       0.37      0.34      0.34       160
      weighted avg       0.53      0.54      0.53       160
```

Figure 12: Results of predicting favourite genre by better model

For predicting the favorite genre, a better model was found to be Grid-SearchCV with C=0.05 and a linear kernel.