FINALPART4

June 16, 2025

Names: Zach Bar and Andrew Kang

1 Term Project Part 4: Modeling & Report

[1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

data = pd.read_csv("data/mmxh.csv")
```

2 Question/Problem Formulation

- As a member of Generation Z, I've become increasingly aware of how common mental health struggles are among my peers. Research shows that Gen Z is currently facing more mental health challenges than any other generation. At the same time, we also tend to listen to music more frequently than other generations. This observation led me to wonder whether there might be a meaningful connection between the two.
- As someone from Gen Z who avidly enjoys listening to music, I started thinking about how often our favorite genres play a big role in managing emotions. Whether it's a go-to playlist when we're stressed or music to help us sleep. Music, to me, has always acted as a form of self-care. This got me interested in exploring whether the frequency of listening to music, particularly the genres people personally enjoy the most, has any measurable impact on mental health.

Goal of the Project

- The goal of this project is to use data to predict a person's mental health score based on how often they listen to music and the genre they prefer. My question is "How does the frequency of listening to your favorite genre of music affect mental health in Generation Z?".

My hypothesis is that more frequent listening to one's favorite genre can have a positive effect on mental health, potentially lowering anxiety, depression, OCD, and insomnia levels.

3 Background Research

Generation Z is facing a well-documented mental health crisis. Studies from many different health-care industries show Gen Z reporting having the highest levels of variety of mental health issues, including anxiety, depression, insomnia, etc., compared to any previous generation.

At the same time, Gen Z is also the most musically engaged generation, with many streaming platforms reporting Gen Z users consume more music content than any other age group. Generation Z is also more likely to use music for stress relief, focus, and self-expression.

Many clinical studies suggest the therapeutic benefits of music. Music has a positive effect on neurotransmitters, which highly influence mood, sleep, and mental health. Music also has a lot of positive physical benefits, such as pain reduction and controlling breath rate.

4 Data Acquisition

Dataset The dataset that we use was The Music & Mental Health Survey, which was a survey taken during Fall 2022, which asked numerous questions relating to the frequency of listening to a certain genre, their mental health state, and generic questions focused on musical backgrounds and listening habits. This data was collected via a Google form created by Catherine Rasgaitis, where respondents were not restricted by age or location. The form was posted in various Reddit forums, Discord servers, and other social media platforms. Posters were also used to advertise this form in libraries, parks, and other public locations. This data was then uploaded onto kaggle.com.

Some limitations with this data set were that the data was self-reported, which means topics like self-diagnosed mental health can be subjective. This data also does not contain any other information, such as lifestyle habits, socioeconomic status, or other factors that could influence mental health. Lastly, this dataset was collected in 2022 and has never been updated since.

4.1 Description of Data

The name of the dataset is Music & Mental Health Survey Results, which is accessible to anyone with internet. This dataset aims to identify any correlations between a person's music taste and their mental health. The datasets hope to improve applications of music therapy or to provide more insight on how the mind works. Music therapy is an evidence-based practice that uses music to improve someone's stress, mood, and mental health. Respondents were not restricted by location or age and the survey was posted on multiple popular online platfo details.

Structure The original dataset is a CSV file containing tabular data that has 736 rows and 33 columns. Throughout the datasets there are a lot of qualitative ordinal variables such as frequency. There are also a lot of qualitative nominal variables such as genre. There are a couple of quantitative discrete variables such as age. Lastly there are also a few quantitative variables such as the timestamp.

Granularity This dataset has high granularity, with many individual data points and fine-grained details. The data gives us a pretty detailed look at each respondents with numerous columns ranging from "Timestamp", "Age", "Primary streaming service","Frequency","Foreign Language", mental health scores, etc.

[2]:
```
data.columns
```

[2]:
```
Index(['Timestamp', 'Age', 'Primary streaming service', 'Hours per day',
       'While working', 'Instrumentalist', 'Composer', 'Fav genre',
       'Exploratory', 'Foreign languages', 'BPM', 'Frequency [Classical]',
       'Frequency [Country]', 'Frequency [EDM]', 'Frequency [Folk]',
       'Frequency [Gospel]', 'Frequency [Hip hop]', 'Frequency [Jazz]',
       'Frequency [K pop]', 'Frequency [Latin]', 'Frequency [Lofi]',
       'Frequency [Metal]', 'Frequency [Pop]', 'Frequency [R&B]',
       'Frequency [Rap]', 'Frequency [Rock]', 'Frequency [Video game music]',
       'Anxiety', 'Depression', 'Insomnia', 'OCD', 'Music effects',
       'Permissions'],
      dtype='object')
```

Scope This dataset was small to begin with, so something that can be added is probably more data, especially more data over time, since most respondents responded around August 2022. Also, all mental health scores were self-reported and not professionally diagnosed. There were a lot of factors affecting mental health that were left out as well, which include genetics and environmental factors.

Temporality This dataset is a little outdated and has not been updated since 2022. Although we didn't really use timestamps in our analysis, so we're looking at everything from a single point in time.

Faithfulness There are a couple of limitations from initially looking at the dataset, such as how each data point was inputted by hand, which could cause some human error. Some questions were also left optional and external factors that contribute to mental health were not accounted for.

5 Data Preview

A sneak peek of the first 10 rows of the data

[3]:
```
data
```

[3]:
```
              Timestamp   Age Primary streaming service  Hours per day
0     8/27/2022 19:29:02  18.0                   Spotify            3.0   \
1     8/27/2022 19:57:31  63.0                   Pandora            1.5
2     8/27/2022 21:28:18  18.0                   Spotify            4.0
3     8/27/2022 21:40:40  61.0             YouTube Music            2.5
4     8/27/2022 21:54:47  18.0                   Spotify            4.0
..                   …     …                         …              …
731  10/30/2022 14:37:28  17.0                   Spotify            2.0
732   11/1/2022 22:26:42  18.0                   Spotify            1.0
733   11/3/2022 23:24:38  19.0  Other streaming service            6.0
734   11/4/2022 17:31:47  19.0                   Spotify            5.0
735    11/9/2022 1:55:20  29.0             YouTube Music            2.0
```

```
    While working Instrumentalist Composer         Fav genre Exploratory  \
0             Yes             Yes      Yes             Latin         Yes   
1             Yes              No       No              Rock         Yes   
2              No              No       No  Video game music          No   
3             Yes              No      Yes              Jazz         Yes   
4             Yes              No       No               R&B         Yes   
..            …             …      …               …         …   
731           Yes             Yes       No              Rock         Yes   
732           Yes             Yes       No               Pop         Yes   
733           Yes              No      Yes               Rap         Yes   
734           Yes             Yes       No         Classical          No   
735           Yes              No       No           Hip hop         Yes   

    Foreign languages  …  Frequency [R&B]  Frequency [Rap] Frequency [Rock]  \
0                 Yes  …        Sometimes  Very frequently            Never   
1                  No  …        Sometimes           Rarely  Very frequently   
2                 Yes  …            Never           Rarely           Rarely   
3                 Yes  …        Sometimes            Never            Never   
4                  No  …  Very frequently  Very frequently            Never   
..                …  …              …              …              …   
731               Yes  …            Never           Rarely  Very frequently   
732               Yes  …            Never            Never        Sometimes   
733                No  …        Sometimes        Sometimes           Rarely   
734                No  …            Never            Never            Never   
735               Yes  …  Very frequently  Very frequently  Very frequently   

    Frequency [Video game music] Anxiety Depression Insomnia  OCD  \
0                      Sometimes     3.0        0.0      1.0  0.0   
1                         Rarely     7.0        2.0      2.0  1.0   
2                Very frequently     7.0        7.0     10.0  2.0   
3                          Never     9.0        7.0      3.0  3.0   
4                         Rarely     7.0        2.0      5.0  9.0   
..                           …     …        …      …  …   
731                        Never     7.0        6.0      0.0  9.0   
732                    Sometimes     3.0        2.0      2.0  5.0   
733                       Rarely     2.0        2.0      2.0  2.0   
734                    Sometimes     2.0        3.0      2.0  1.0   
735                       Rarely     2.0        2.0      2.0  5.0   

    Music effects    Permissions  
0             NaN  I understand.  
1             NaN  I understand.  
2       No effect  I understand.  
3         Improve  I understand.  
4         Improve  I understand.  
..            …            …  
```

```
731       Improve  I understand.
732       Improve  I understand.
733       Improve  I understand.
734       Improve  I understand.
735       Improve  I understand.

[736 rows x 33 columns]
```

5.0.1 Data Cleaning

I chose to work with a subset of my data that contained relevant features for my project. Starting with choosing the subset of my data that focused on Gen Z, which were ages 10-25 at the time.

[4]:
```
genz2022=data[(data["Age"]<=25)&(data["Age"]>=10)]
genz2022
```

[4]:
```
               Timestamp   Age Primary streaming service  Hours per day
0      8/27/2022 19:29:02  18.0                   Spotify            3.0   \
2      8/27/2022 21:28:18  18.0                   Spotify            4.0
4      8/27/2022 21:54:47  18.0                   Spotify            4.0
5      8/27/2022 21:56:50  18.0                   Spotify            5.0
6      8/27/2022 22:00:29  18.0             YouTube Music            3.0
..                    …     …                         …             …
730   10/30/2022 13:15:26  21.0                   Spotify            2.0
731   10/30/2022 14:37:28  17.0                   Spotify            2.0
732    11/1/2022 22:26:42  18.0                   Spotify            1.0
733    11/3/2022 23:24:38  19.0   Other streaming service            6.0
734    11/4/2022 17:31:47  19.0                   Spotify            5.0

    While working Instrumentalist Composer         Fav genre Exploratory
0             Yes             Yes      Yes             Latin         Yes   \
2              No              No       No  Video game music          No
4             Yes              No       No               R&B         Yes
5             Yes             Yes      Yes              Jazz         Yes
6             Yes             Yes       No  Video game music         Yes
..              …               …        …                 …           …
730           Yes              No       No               R&B         Yes
731           Yes             Yes       No              Rock         Yes
732           Yes             Yes       No               Pop         Yes
733           Yes              No      Yes               Rap         Yes
734           Yes             Yes       No         Classical          No

    Foreign languages  …  Frequency [R&B]  Frequency [Rap] Frequency [Rock]
0                 Yes  …        Sometimes  Very frequently            Never   \
2                 Yes  …            Never           Rarely           Rarely
4                  No  …  Very frequently  Very frequently            Never
5                 Yes  …  Very frequently  Very frequently  Very frequently
```

```
6                Yes  ...        Rarely          Never            Never
..               ...  ...           ...            ...              ...
730              Yes  ...  Very frequently     Sometimes        Sometimes
731              Yes  ...         Never           Rarely  Very frequently
732              Yes  ...         Never            Never        Sometimes
733               No  ...     Sometimes        Sometimes           Rarely
734               No  ...         Never            Never            Never

    Frequency [Video game music] Anxiety Depression Insomnia  OCD
0                      Sometimes     3.0        0.0      1.0  0.0  \
2                Very frequently     7.0        7.0     10.0  2.0
4                         Rarely     7.0        2.0      5.0  9.0
5                          Never     8.0        8.0      7.0  7.0
6                      Sometimes     4.0        8.0      6.0  0.0
..                           ...     ...        ...      ...  ...
730                    Sometimes     7.0        6.0      4.0  6.0
731                        Never     7.0        6.0      0.0  9.0
732                    Sometimes     3.0        2.0      2.0  5.0
733                       Rarely     2.0        2.0      2.0  2.0
734                    Sometimes     2.0        3.0      2.0  1.0

    Music effects    Permissions
0             NaN  I understand.
2       No effect  I understand.
4         Improve  I understand.
5         Improve  I understand.
6         Improve  I understand.
..            ...            ...
730       Improve  I understand.
731       Improve  I understand.
732       Improve  I understand.
733       Improve  I understand.
734       Improve  I understand.

[509 rows x 33 columns]
```

[5]:
```
df=genz2022.drop(columns=['Age','Timestamp','Primary streaming service','While working','Instrumentalist','Composer','BPM','Foreign languages','Permissions'])
df
```

[5]:
```
     Hours per day         Fav genre Exploratory Frequency [Classical]
0              3.0             Latin         Yes                Rarely  \
2              4.0  Video game music          No                 Never
4              4.0               R&B         Yes                 Never
5              5.0              Jazz         Yes                Rarely
6              3.0  Video game music         Yes             Sometimes
```

```
..          …               …        …                  …
730         2.0           R&B       Yes         Sometimes
731         2.0          Rock       Yes   Very frequently
732         1.0           Pop       Yes            Rarely
733         6.0           Rap       Yes            Rarely
734         5.0     Classical        No   Very frequently

    Frequency [Country]  Frequency [EDM] Frequency [Folk] Frequency [Gospel]
0                 Never           Rarely            Never              Never  \
2                 Never  Very frequently            Never              Never
4                 Never           Rarely            Never             Rarely
5             Sometimes            Never            Never              Never
6                 Never           Rarely        Sometimes             Rarely
..                  …                …                …                  …
730               Never        Sometimes           Rarely              Never
731              Rarely            Never        Sometimes              Never
732              Rarely            Never            Never              Never
733           Sometimes        Sometimes           Rarely             Rarely
734               Never            Never            Never              Never

    Frequency [Hip hop] Frequency [Jazz]  …  Frequency [Pop]
0             Sometimes            Never  …  Very frequently  \
2                Rarely           Rarely  …           Rarely
4       Very frequently            Never  …        Sometimes
5             Sometimes  Very frequently  …  Very frequently
6                Rarely        Sometimes  …           Rarely
..                  …                …  …                …
730               Never        Sometimes  …        Sometimes
731           Sometimes           Rarely  …  Very frequently
732               Never           Rarely  …  Very frequently
733     Very frequently           Rarely  …        Sometimes
734               Never           Rarely  …            Never

     Frequency [R&B]  Frequency [Rap] Frequency [Rock]
0          Sometimes  Very frequently            Never  \
2              Never           Rarely           Rarely
4    Very frequently  Very frequently            Never
5    Very frequently  Very frequently  Very frequently
6             Rarely            Never            Never
..               …                …                …
730  Very frequently        Sometimes        Sometimes
731            Never           Rarely  Very frequently
732            Never            Never        Sometimes
733        Sometimes        Sometimes           Rarely
734            Never            Never            Never

    Frequency [Video game music] Anxiety Depression Insomnia  OCD
```

```
0                    Sometimes     3.0        0.0       1.0  0.0  \
2              Very frequently     7.0        7.0      10.0  2.0
4                       Rarely     7.0        2.0       5.0  9.0
5                        Never     8.0        8.0       7.0  7.0
6                    Sometimes     4.0        8.0       6.0  0.0
..                         …    …          …       …   …
730                  Sometimes     7.0        6.0       4.0  6.0
731                      Never     7.0        6.0       0.0  9.0
732                  Sometimes     3.0        2.0       2.0  5.0
733                     Rarely     2.0        2.0       2.0  2.0
734                  Sometimes     2.0        3.0       2.0  1.0

     Music effects
0              NaN
2        No effect
4          Improve
5          Improve
6          Improve
..             …
730        Improve
731        Improve
732        Improve
733        Improve
734        Improve

[509 rows x 24 columns]
```

In the dataset above, I dropped the columns that weren't really relevant to my project

[6]:
```
projectdf=df.reset_index()
projdf=projectdf.drop(columns=['index'])
projdf
```

[6]:
```
     Hours per day         Fav genre Exploratory Frequency [Classical]
0              3.0             Latin         Yes                Rarely  \
1              4.0  Video game music          No                 Never
2              4.0               R&B         Yes                 Never
3              5.0              Jazz         Yes                Rarely
4              3.0  Video game music         Yes             Sometimes
..             …               …        …                   …
504            2.0               R&B         Yes             Sometimes
505            2.0              Rock         Yes       Very frequently
506            1.0               Pop         Yes                Rarely
507            6.0               Rap         Yes                Rarely
508            5.0         Classical          No       Very frequently

    Frequency [Country]  Frequency [EDM] Frequency [Folk] Frequency [Gospel]
```

```
0                Never          Rarely            Never              Never  \
1                Never  Very frequently           Never              Never
2                Never          Rarely            Never             Rarely
3            Sometimes           Never            Never              Never
4                Never          Rarely        Sometimes             Rarely
..                 …               …                …                  …
504              Never       Sometimes           Rarely              Never
505             Rarely           Never        Sometimes              Never
506             Rarely           Never            Never              Never
507          Sometimes       Sometimes           Rarely             Rarely
508              Never           Never            Never              Never

    Frequency [Hip hop] Frequency [Jazz]  …  Frequency [Pop]
0             Sometimes            Never  …  Very frequently  \
1                Rarely           Rarely  …           Rarely
2       Very frequently            Never  …        Sometimes
3             Sometimes  Very frequently  …  Very frequently
4                Rarely        Sometimes  …           Rarely
..                  …               …  …              …
504               Never        Sometimes  …        Sometimes
505           Sometimes           Rarely  …  Very frequently
506               Never           Rarely  …  Very frequently
507     Very frequently           Rarely  …        Sometimes
508               Never           Rarely  …            Never

     Frequency [R&B]  Frequency [Rap] Frequency [Rock]
0          Sometimes  Very frequently            Never  \
1              Never           Rarely           Rarely
2    Very frequently  Very frequently            Never
3    Very frequently  Very frequently  Very frequently
4             Rarely            Never            Never
..               …               …                …
504  Very frequently        Sometimes        Sometimes
505            Never           Rarely  Very frequently
506            Never            Never        Sometimes
507        Sometimes        Sometimes           Rarely
508            Never            Never            Never

    Frequency [Video game music] Anxiety Depression Insomnia  OCD
0                      Sometimes     3.0        0.0      1.0  0.0  \
1                Very frequently     7.0        7.0     10.0  2.0
2                         Rarely     7.0        2.0      5.0  9.0
3                          Never     8.0        8.0      7.0  7.0
4                      Sometimes     4.0        8.0      6.0  0.0
..                           …       …          …        …   …
504                    Sometimes     7.0        6.0      4.0  6.0
505                        Never     7.0        6.0      0.0  9.0
```

```
506                    Sometimes     3.0         2.0       2.0  5.0
507                       Rarely     2.0         2.0       2.0  2.0
508                    Sometimes     2.0         3.0       2.0  1.0

     Music effects
0              NaN
1        No effect
2          Improve
3          Improve
4          Improve
..             …
504        Improve
505        Improve
506        Improve
507        Improve
508        Improve

[509 rows x 24 columns]
```

Lastly, I resetted the index of my dataset to maintain order. Now, we can analyze the average ratings of each mental health struggle relative to the participant's favorite genre using a pivot table.

5.1 Potential issues with this data?

In the sample above, some of the data is reported as NaN. I decided to input the values as "No Effect" rather than drop the rows, which helped retain a larger sample size.

[7]:

```
NoNan=projdf["Music effects"].fillna("No effect")
NoNan
projdf["Music effects"]=NoNan
projdf
```

[7]:

```
     Hours per day         Fav genre Exploratory Frequency [Classical]
0              3.0             Latin         Yes                Rarely  \
1              4.0  Video game music          No                 Never
2              4.0               R&B         Yes                 Never
3              5.0              Jazz         Yes                Rarely
4              3.0  Video game music         Yes             Sometimes
..             …               …         …                   …
504            2.0               R&B         Yes             Sometimes
505            2.0              Rock         Yes       Very frequently
506            1.0               Pop         Yes                Rarely
507            6.0               Rap         Yes                Rarely
508            5.0         Classical          No       Very frequently

    Frequency [Country]  Frequency [EDM] Frequency [Folk] Frequency [Gospel]
0                 Never           Rarely            Never              Never  \
```

```
1                 Never  Very frequently            Never              Never
2                 Never           Rarely            Never             Rarely
3             Sometimes            Never            Never              Never
4                 Never           Rarely        Sometimes             Rarely
..                    …                …                …                  …
504               Never        Sometimes           Rarely              Never
505              Rarely            Never        Sometimes              Never
506              Rarely            Never            Never              Never
507           Sometimes        Sometimes           Rarely             Rarely
508               Never            Never            Never              Never

    Frequency [Hip hop] Frequency [Jazz]  …  Frequency [Pop]
0             Sometimes            Never  …  Very frequently  \
1                Rarely           Rarely  …           Rarely
2       Very frequently            Never  …        Sometimes
3             Sometimes  Very frequently  …  Very frequently
4                Rarely        Sometimes  …           Rarely
..                    …                …  …                …
504               Never        Sometimes  …        Sometimes
505           Sometimes           Rarely  …  Very frequently
506               Never           Rarely  …  Very frequently
507     Very frequently           Rarely  …        Sometimes
508               Never           Rarely  …            Never

     Frequency [R&B]  Frequency [Rap] Frequency [Rock]
0          Sometimes  Very frequently            Never  \
1              Never           Rarely           Rarely
2    Very frequently  Very frequently            Never
3    Very frequently  Very frequently  Very frequently
4             Rarely            Never            Never
..                 …                …                …
504  Very frequently        Sometimes        Sometimes
505            Never           Rarely  Very frequently
506            Never            Never        Sometimes
507        Sometimes        Sometimes           Rarely
508            Never            Never            Never

    Frequency [Video game music] Anxiety Depression Insomnia  OCD
0                      Sometimes     3.0        0.0      1.0  0.0  \
1                Very frequently     7.0        7.0     10.0  2.0
2                         Rarely     7.0        2.0      5.0  9.0
3                          Never     8.0        8.0      7.0  7.0
4                      Sometimes     4.0        8.0      6.0  0.0
..                             …       …          …        …    …
504                    Sometimes     7.0        6.0      4.0  6.0
505                        Never     7.0        6.0      0.0  9.0
506                    Sometimes     3.0        2.0      2.0  5.0
```

```
507                          Rarely     2.0         2.0       2.0  2.0
508                       Sometimes     2.0         3.0       2.0  1.0

     Music effects
0        No effect
1        No effect
2          Improve
3          Improve
4          Improve
..             …
504        Improve
505        Improve
506        Improve
507        Improve
508        Improve

[509 rows x 24 columns]
```

5.2 Exploratory Data Analysis

5.2.1 1. How is your data organized and what does it contain?

- The dataset shows responses for a survey where participants self-report their favorite music genre and rate their mental health struggles on numerical scales for:
  - Anxiety
  - Depression
  - Insomnia
  - OCD
- Columns of interest:
  - `Fav genre` (categorical)
  - `Frequency` (ordinal ratings)
  - `Anxiety`, `Depression`, `Insomnia`, `OCD` (numerical, likely ordinal ratings, e.g., 1–10)
- Data Granularity:
  - Each row represents one individual's ratings, making the data granular at the individual level. #### How many respondents do we have?

[8]:
```
print(len(projdf))
```

```
509
```

As a result of our data cleaning process, only self-reported Generation Z respondents were retained

5.2.2 2. How is your data distributed? How are different subsets of your data distributed?

Average Self-Reported Mental Health

[9]:
```
#Dataset of just the Self Reported Mental Health
proj_mental_cols=projdf[["Anxiety","Depression","Insomnia","OCD"]]
```

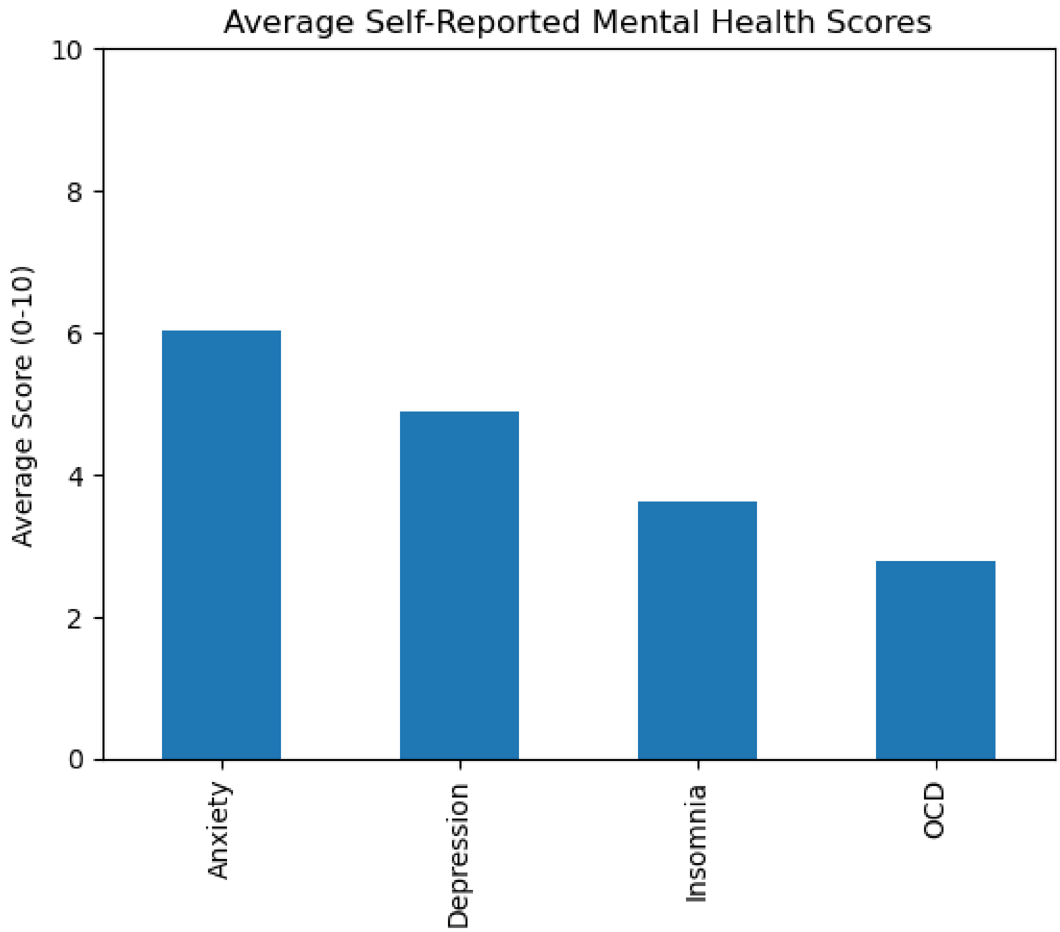
```
proj_mental_cols
```

[9]:
```
     Anxiety  Depression  Insomnia  OCD
0        3.0         0.0       1.0  0.0
1        7.0         7.0      10.0  2.0
2        7.0         2.0       5.0  9.0
3        8.0         8.0       7.0  7.0
4        4.0         8.0       6.0  0.0
..       …           …       …  …
504      7.0         6.0       4.0  6.0
505      7.0         6.0       0.0  9.0
506      3.0         2.0       2.0  5.0
507      2.0         2.0       2.0  2.0
508      2.0         3.0       2.0  1.0

[509 rows x 4 columns]
```

Looking at the first and last few rows, we can see there were a lot of highly reported scores for Mental Health, where 0 represents the lowest severity and 10 represents the highest. We could also see that individuals who reported a high score for a certain mental health condition have also often reported a similar score for other mental health conditions, suggesting a possible correlation among those conditions.

[10]:
```
#Calculate average scores for each condition
avg=proj_mental_cols.mean()

#Bar chart of the average scores
avg.plot(kind='bar')
plt.title("Average Self-Reported Mental Health Scores")
plt.ylabel("Average Score (0-10)")
plt.ylim(0,10)
```

[10]:
```
(0.0, 10.0)
```



From this graph, we can see that although we saw a lot of highly reported mental health scores, the average scores for most of the mental health conditions surveyed are between 3-6. #### Descriptive Statistics for Mental Struggles

- The table below shows the average mental health score for each mental health surveyed against each genre surveyed,

[11]:
```
mental_cols=["Anxiety","Depression","Insomnia","OCD"]
pivot_data1 = data.groupby('Fav genre')[mental_cols].mean()
pivot_data1
```

[11]:

	Anxiety	Depression	Insomnia	OCD
Fav genre				
Classical	4.886792	4.075472	3.792453	2.377358
Country	5.400000	4.320000	2.720000	2.760000
EDM	5.486486	5.243243	3.972973	3.000000
Folk	6.566667	5.066667	3.633333	2.200000
Gospel	4.833333	2.666667	5.333333	0.333333

```
Hip hop            6.200000   5.800000  3.428571  2.714286
Jazz               5.900000   4.500000  3.850000  2.800000
K pop              6.230769   4.423077  3.461538  2.538462
Latin              4.333333   3.000000  3.333333  1.666667
Lofi               6.100000   6.600000  5.600000  3.400000
Metal              5.761364   5.068182  4.556818  2.397727
Pop                6.074561   4.486842  3.368421  2.855263
R&B                5.171429   3.828571  2.885714  2.742857
Rap                5.090909   4.000000  2.272727  3.181818
Rock               6.122340   5.236702  3.880319  2.678191
Video game music   5.886364   4.477273  4.000000  2.386364
```

[12]:
```python
mental_cols = ['Anxiety', 'Depression', 'Insomnia', 'OCD']
data[mental_cols].describe()
```

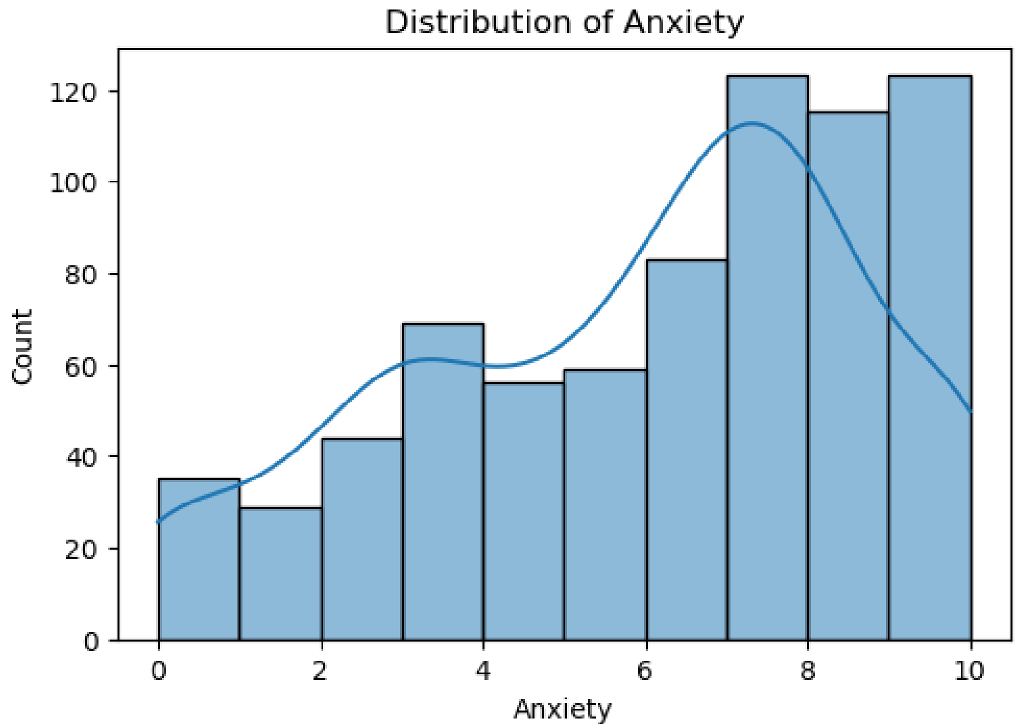
[12]:

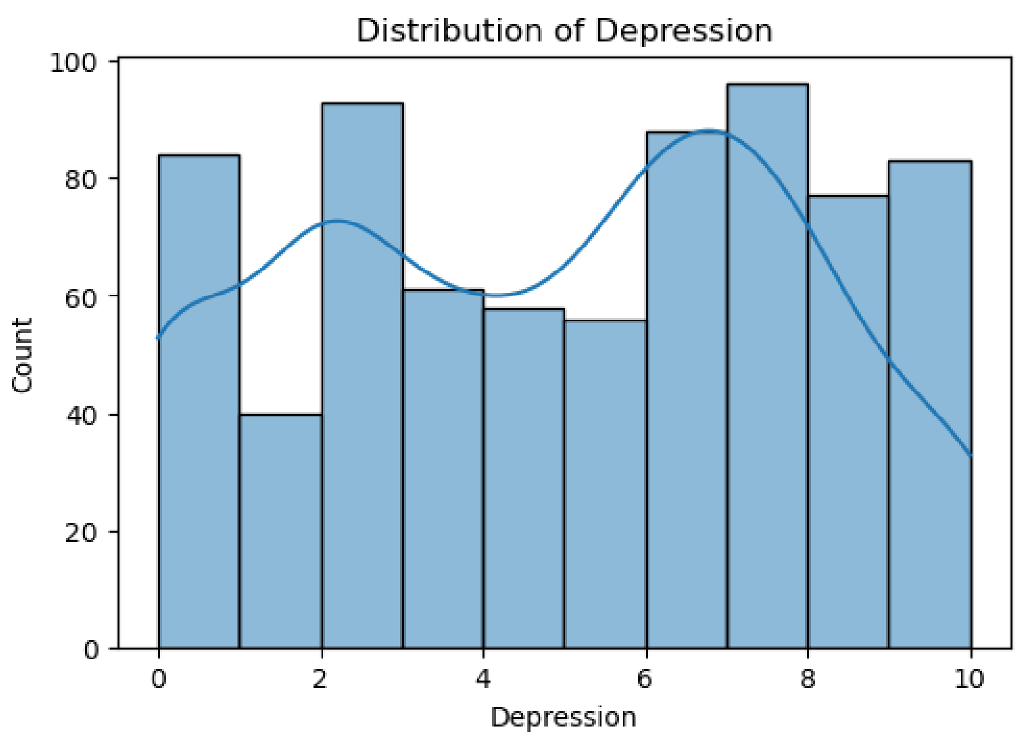
	Anxiety	Depression	Insomnia	OCD
count	736.000000	736.000000	736.000000	736.000000
mean	5.837636	4.796196	3.738451	2.637228
std	2.793054	3.028870	3.088689	2.842017
min	0.000000	0.000000	0.000000	0.000000
25%	4.000000	2.000000	1.000000	0.000000
50%	6.000000	5.000000	3.000000	2.000000
75%	8.000000	7.000000	6.000000	5.000000
max	10.000000	10.000000	10.000000	10.000000

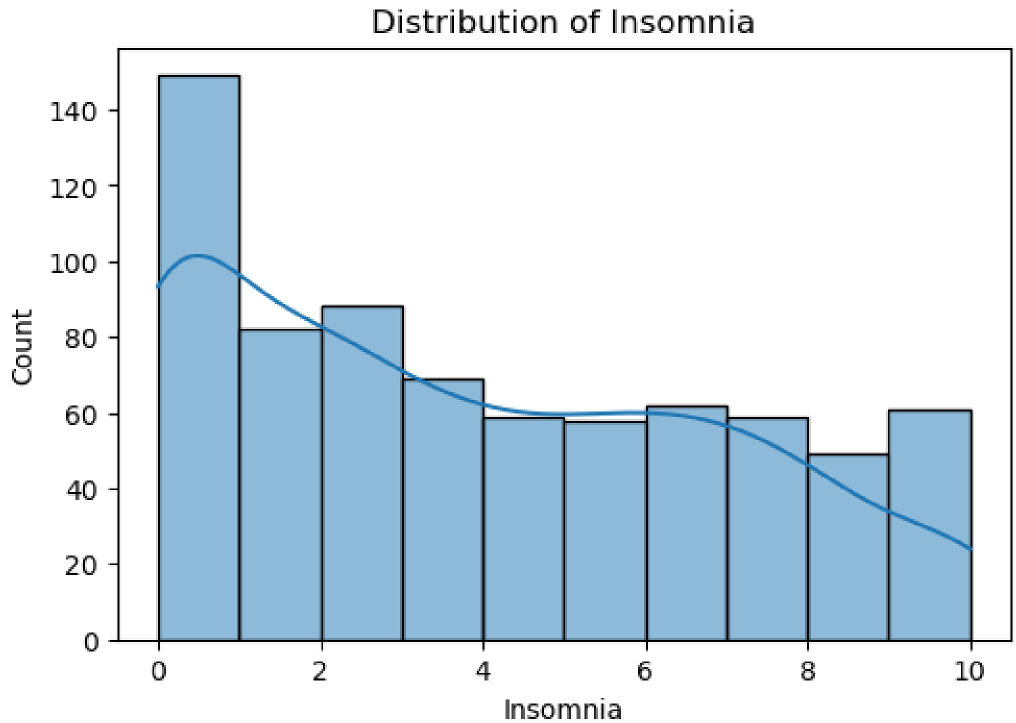
- Through these findings, we could conclude that generally we will find higher rates of anxiety over all these other mental struggles. On average anxiety rated at more than 5, and with the 25th percentile (75% of those that took the survey) rating it a 4 or higher.
- When analyzing different levels of each mental struggle and how it may be attributed to the music the participant enjoys, we have to do so by comparing it to the standard mean, not to other mental struggle columns

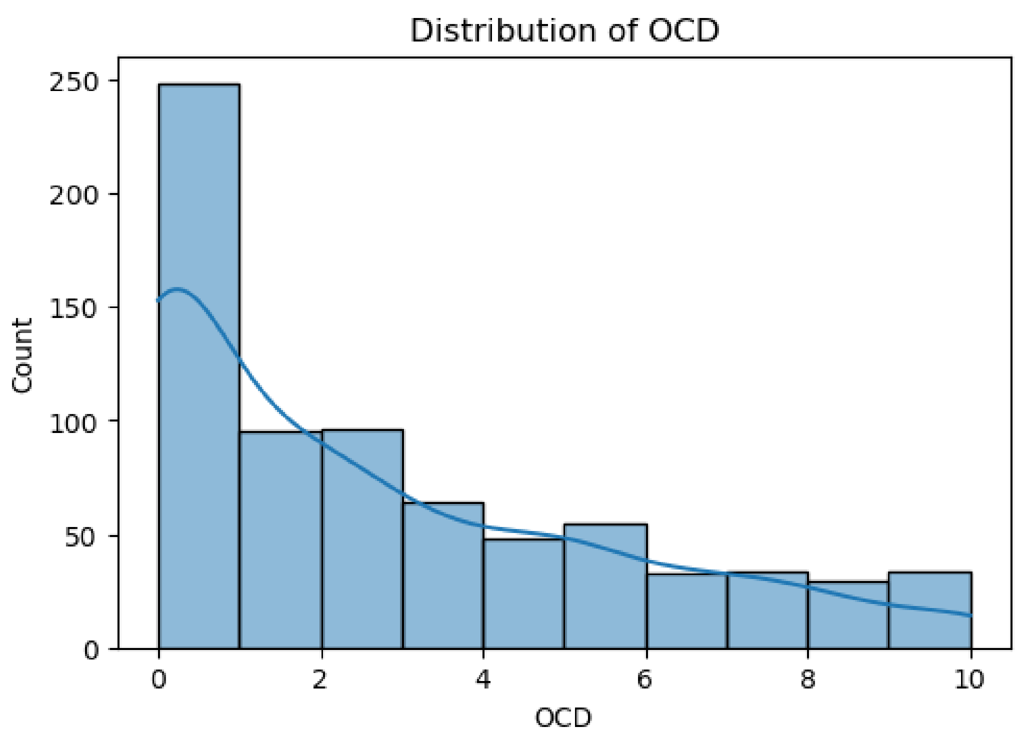
Distribution Plots

[13]:
```python
# Plot histograms for each mental struggle
for col in mental_cols:
    plt.figure(figsize=(6,4))
    sns.histplot(data[col], bins=10, kde=True)
    plt.title(f'Distribution of {col}')
    plt.show()
```


Distribution of Anxiety
Count
Anxiety
0
20
40
60
80
100
120
0
2
4
6
8
10


Distribution of Depression
Count
Depression
0
20
40
60
80
100
0
2
4
6
8
10


Distribution of Insomnia
Count
140
120
100
80
60
40
20
0
0
2
4
6
8
10
Insomnia


Distribution of OCD
Count
250
200
150
100
50
0
0
2
4
6
8
10
OCD

- We can see by using distribution plots for each of the mental struggles that none of them have normal distribution
- Interestingly, the peaks for OCD and Insomnia are at 0, while for depression the peak is 8 and anxiety is 8 and 10. This shows a skewness in the dataset towards abnormal rates of depression and anxiety amongst participants at large. We will get to see how much music genres have an effect on in the outliers of this dataset.
- There is more variability amongst participants in rating their levels of depression and insomnia, while OCD and anxiety both show a skew, each to opposing ends.

What proportions of the respondents are exploratory with their music?

[14]:
```
projdf['Exploratory'].value_counts()
```

[14]:
```
Exploratory
Yes    378
No     131
Name: count, dtype: int64
```

[15]:
```
print(378/len(projdf)*100)
```

```
74.26326129666012
```

- About 74% of respondents are exploratory with their music, which means 74% of respondents actively explore new genres/artists.

5.2.3 3. Do you have all the relevant data you need?

- From my analysis of the dataset, there are definitely some missing genres and far too many missing mental struggles. Its safe to say that of the genres and mental health struggles contained in the dataset majority of people fall under.

[16]:
```
genre_counts = data['Fav genre'].value_counts()
genre_counts
```

[16]:
```
Fav genre
Rock                188
Pop                 114
Metal                88
Classical            53
Video game music     44
EDM                  37
R&B                  35
Hip hop              35
Folk                 30
K pop                26
Country              25
Rap                  22
Jazz                 20
Lofi                 10
```

```
Gospel                  6
Latin                   3
Name: count, dtype: int64
```

- If we take note of the distribution of genres and number of participants that voted it as their favorite, we see a heavy skew towards rock genres, with Rock and Metal. These two genres sum up to over a third of total participants.

[17]:
```
np.sum(genre_counts)
```

[17]:
```
736
```

- It is also important to note that some genres have a very low number of participants and cannot effectively suggest an association between listening to the music and having that mental struggle. These genres will be most likely to not be consistent with standard rates and can skew due to outliers.

5.2.4 4.What are the biases, anomalies, or other issues with the data?

Biases with the data:

- Since all the responses are self-reported, there's a risk of self-reporting bias. Individuals might not always accurately report their music habits or mental health.
  - All mental health scores were self-diagnosed and not clinically diagnosed, which can limit the accuracy and reliability of conclusions drawn from these values
- Although we are focusing on Gen Z, the original survey itself probably does not fully represent a proper sample of that population, much less a proper sample of the generation.

Anomalies with the data:

- Some genres are clearly much more popular than others, such as Rock, Pop, and Metal, while others like Latin and Gospel have very few listeners. This imbalance could skew the analysis

Other issues with the data - This dataset lacks information such as geographic location of the respondent, cultural background, socioeconomic status, etc., which are all factors that can heavily influence both music preferences and mental health, which makes it harder to understand the full picture.

5.2.5 5. How does the data need to be transformed to enable effective analysis?

In this dataset, to prepare the dataset for effective analysis, a few transformations are needed. - Columns that record how often someone listens to each genre using words such as "Never", "Rarely", "Sometimes", and "Very frequently" have a clear order. - We used ordinal encoding to convert them into numbers (1 to 4) - This allows models to interpret frequency levels meaningfully

[18]:
```
projdff=projdf.replace(["Rarely","Never","Sometimes","Very␣
↪frequently"],[1,0,2,3])
projdff.head(5)
```

[18]:
```
   Hours per day        Fav genre Exploratory  Frequency [Classical]
0            3.0            Latin         Yes                      1  \
1            4.0  Video game music          No                      0
2            4.0              R&B         Yes                      0
3            5.0             Jazz         Yes                      1
4            3.0  Video game music         Yes                      2

   Frequency [Country]  Frequency [EDM]  Frequency [Folk]  Frequency [Gospel]
0                    0                1                 0                   0  \
1                    0                3                 0                   0
2                    0                1                 0                   1
3                    2                0                 0                   0
4                    0                1                 2                   1

   Frequency [Hip hop]  Frequency [Jazz]  …  Frequency [Pop]
0                    2                 0  …                3  \
1                    1                 1  …                1
2                    3                 0  …                2
3                    2                 3  …                3
4                    1                 2  …                1

   Frequency [R&B]  Frequency [Rap]  Frequency [Rock]
0                2                3                 0  \
1                0                1                 1
2                3                3                 0
3                3                3                 3
4                1                0                 0

   Frequency [Video game music]  Anxiety  Depression  Insomnia  OCD
0                             2      3.0         0.0       1.0  0.0  \
1                             3      7.0         7.0      10.0  2.0
2                             1      7.0         2.0       5.0  9.0
3                             0      8.0         8.0       7.0  7.0
4                             2      4.0         8.0       6.0  0.0

   Music effects
0      No effect
1      No effect
2        Improve
3        Improve
4        Improve

[5 rows x 24 columns]
```

5.2.6 6. Structure, granularity, scope, temporality, and faithfulness. Did your understanding change?

Let's look at the dataset

[19]:
```
projdff
```

[19]:
```
     Hours per day        Fav genre Exploratory  Frequency [Classical]
0              3.0            Latin         Yes                      1  \
1              4.0  Video game music          No                      0
2              4.0              R&B         Yes                      0
3              5.0             Jazz         Yes                      1
4              3.0  Video game music         Yes                      2
..             …                …         …                    …
504            2.0              R&B         Yes                      2
505            2.0             Rock         Yes                      3
506            1.0              Pop         Yes                      1
507            6.0              Rap         Yes                      1
508            5.0        Classical          No                      3

     Frequency [Country]  Frequency [EDM]  Frequency [Folk]
0                      0                1                 0  \
1                      0                3                 0
2                      0                1                 0
3                      2                0                 0
4                      0                1                 2
..                   …              …               …
504                    0                2                 1
505                    1                0                 2
506                    1                0                 0
507                    2                2                 1
508                    0                0                 0

     Frequency [Gospel]  Frequency [Hip hop]  Frequency [Jazz]  …
0                     0                    2                 0  …  \
1                     0                    1                 1  …
2                     1                    3                 0  …
3                     0                    2                 3  …
4                     1                    1                 2  …
..                  …                  …               …  …
504                   0                    0                 2  …
505                   0                    2                 1  …
506                   0                    0                 1  …
507                   1                    3                 1  …
508                   0                    0                 1  …

     Frequency [Pop]  Frequency [R&B]  Frequency [Rap]  Frequency [Rock]
0                  3                2                3                 0  \
```

```
1                    1                 0                 1                  1
2                    2                 3                 3                  0
3                    3                 3                 3                  3
4                    1                 1                 0                  0
..                 ...               ...               ...                ...
504                  2                 3                 2                  2
505                  3                 0                 1                  3
506                  3                 0                 0                  2
507                  2                 2                 2                  1
508                  0                 0                 0                  0

     Frequency [Video game music]  Anxiety  Depression  Insomnia  OCD
0                               2      3.0         0.0       1.0  0.0  \
1                               3      7.0         7.0      10.0  2.0
2                               1      7.0         2.0       5.0  9.0
3                               0      8.0         8.0       7.0  7.0
4                               2      4.0         8.0       6.0  0.0
..                            ...      ...         ...       ...  ...
504                             2      7.0         6.0       4.0  6.0
505                             0      7.0         6.0       0.0  9.0
506                             2      3.0         2.0       2.0  5.0
507                             1      2.0         2.0       2.0  2.0
508                             2      2.0         3.0       2.0  1.0

     Music effects
0        No effect
1        No effect
2          Improve
3          Improve
4          Improve
..             ...
504        Improve
505        Improve
506        Improve
507        Improve
508        Improve

[509 rows x 24 columns]
```

Structure - After cleaning the data, my understanding of it definitely deepened. This dataset feels cleaner and organized, which makes it easier to work with. Structurally, the dataset is now in good shape, with 509 rows and 24 columns. Each row represents one Gen Z respondent, and their columns include their music habits, favorite genres, and self-reported mental health scores.

[20]:
```
projdff.shape
#shape of dataframe
```

[20]: (509, 24)

Granularity

- In terms of granularity, it has a high granularity, with many individual data points and fine-grained details. The data gives us a pretty detailed look at each respondent with numerous columns ranging from "Hours per day", "Fav genre", mental health scores, listening frequency of genres, etc.

[21]:
```
#Each of the 24 columns
projdff.columns
```

[21]:
```
Index(['Hours per day', 'Fav genre', 'Exploratory', 'Frequency [Classical]',
       'Frequency [Country]', 'Frequency [EDM]', 'Frequency [Folk]',
       'Frequency [Gospel]', 'Frequency [Hip hop]', 'Frequency [Jazz]',
       'Frequency [K pop]', 'Frequency [Latin]', 'Frequency [Lofi]',
       'Frequency [Metal]', 'Frequency [Pop]', 'Frequency [R&B]',
       'Frequency [Rap]', 'Frequency [Rock]', 'Frequency [Video game music]',
       'Anxiety', 'Depression', 'Insomnia', 'OCD', 'Music effects'],
      dtype='object')
```

Scope

- By filtering the dataset to include on Gen Z, we were able to focus our analysis on the generation we're most interested in. That said, it also means we can't generalize any of our findings beyond that group. There are also general limitations to the data, such as genetic or environmental factors that contribute to mental health and music habits were not accounted for.

Temporality

- As for temporality, this data is a little outdated and hasn't been updated since 2022. Even though the original data includes timestamps, we didn't really use them in our analysis. So we're looking at everything from a single point in time.

Faithfulness

- When I first chose this dataset, I thought it seemed like a great dataset for exploring the relationship between music and mental health, but as I continued working with it, I started to notice some limitations, such as genre imbalance, missing values, self-diagnosed mental health scores, etc. So while the data is still useful, there are a lot of factors that question the data's faithfulness, which requires careful interpretation and some caution when drawing conclusions.

6 Modeling

For our model, we chose to use a Linear Regression Model to explore whether the frequency of listening to different music genres has any relationship with self-reported mental health scores. Before building the model, we checked for multicollinearity using VIF. A few genres showed slightly higher VIF values, but none were high enough to seriously interfere with the results.

We ran 4 models, one for each mental health scores that were surveyed. We randomly shuffled the data and manually split it into 80/20 training-test-split. Once the data was split, we used Linear Regression and fitted each model, along with providing a plot for each model. Afterwards, we calculated the RMSE for each model, which turned out to be pretty high for most models. Which means that the model isn't very accurate at predicting mental health scores just based on frequency. This could be due to how our dataset was very small and limited to begin with.

[22]:
```
from sklearn.linear_model import LinearRegression
from statsmodels.stats.outliers_influence import variance_inflation_factor
```

[23]:
```
genre_frequency_cols= [col for col in projdff.columns if col.
  ↪startswith("Frequency [")]
wdata=projdff[genre_frequency_cols + ["Depression"]]

#Split 80/20

shuffdata=wdata.sample(frac=1, random_state=1).reset_index(drop=True)
shuffdata
split_index=int(0.8*len(shuffdata))
training=shuffdata[:split_index]
testing=shuffdata[split_index:]

x_training=training[genre_frequency_cols]
y_training=training["Depression"]

X = projdff[genre_frequency_cols]

# Add a constant (required for statsmodels VIF calculation)
X = X.copy()
X['Intercept'] = 1

# Compute VIF
vif_data = pd.DataFrame()
vif_data['Feature'] = X.columns
vif_data['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.
  ↪shape[1])]
vif_data = vif_data[vif_data['Feature'] != 'Intercept']
print(vif_data)
```

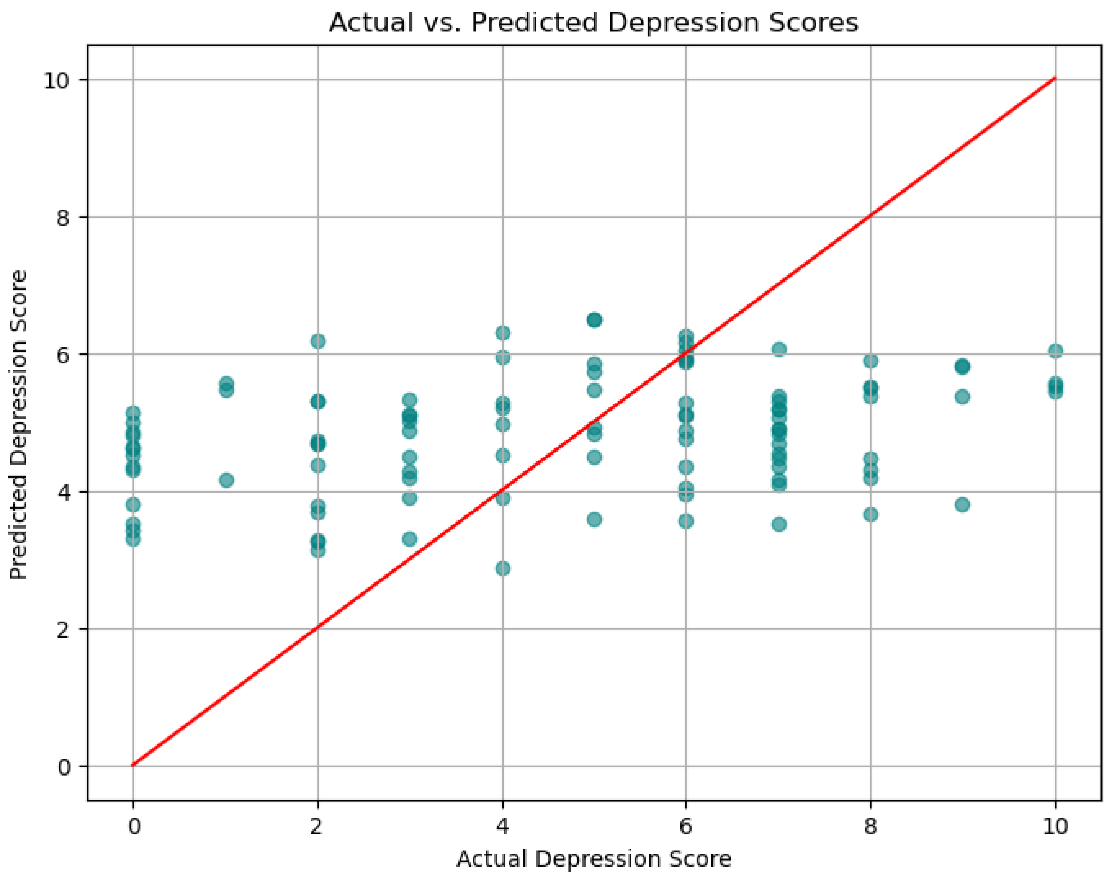
```
                         Feature       VIF
0          Frequency [Classical]  1.246554
1            Frequency [Country]  1.280215
2                Frequency [EDM]  1.202593
3               Frequency [Folk]  1.392621
4             Frequency [Gospel]  1.222284
5            Frequency [Hip hop]  2.811656
6               Frequency [Jazz]  1.360351
7              Frequency [K pop]  1.379643
```

```
8              Frequency [Latin]  1.219998
9               Frequency [Lofi]  1.321791
10             Frequency [Metal]  1.617686
11               Frequency [Pop]  1.312737
12               Frequency [R&B]  1.886568
13               Frequency [Rap]  2.814344
14              Frequency [Rock]  1.754133
15  Frequency [Video game music]  1.283054
```

[24]:
```
x_testing=testing[genre_frequency_cols]
Actual=testing["Depression"]

#Linear Regression
model=LinearRegression()
model.fit (x_training,y_training)

predict=model.predict(x_testing)
error=Actual-predict
RMSE=round(np.sqrt(np.mean(error**2)),2)
RMSE
```

[24]:
```
2.72
```

[25]:
```
plt.figure(figsize=(8, 6))
plt.scatter(Actual, predict, color='teal', alpha=0.6)
plt.plot([Actual.min(), Actual.max()], [Actual.min(), Actual.max()],
 ↪color='red', linestyle='-')
plt.xlabel("Actual Depression Score")
plt.ylabel("Predicted Depression Score")
plt.title("Actual vs. Predicted Depression Scores")
plt.grid(True)
plt.show()
```

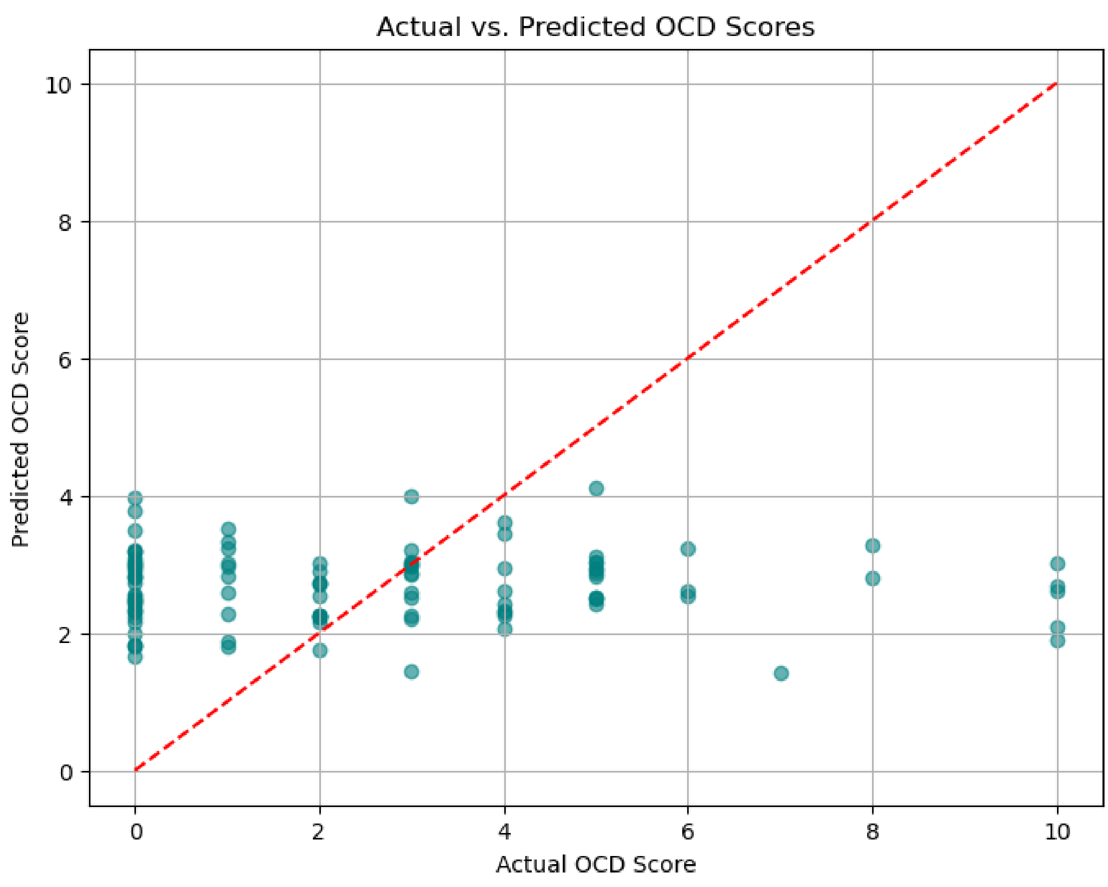



```
[26]: w2data=projdff[genre_frequency_cols + ["OCD"]]

#Split 80/20

shuff2data=w2data.sample(frac=1, random_state=1).reset_index(drop=True)
shuff2data
split_index2=int(0.8*len(shuff2data))
training2=shuff2data[:split_index2]
testing2=shuff2data[split_index2:]

x_training2=training2[genre_frequency_cols]
y_training2=training2["OCD"]
x_testing2=testing2[genre_frequency_cols]
Actual2=testing2["OCD"]

#Linear Regression
model=LinearRegression()
model.fit (x_training2,y_training2)
```

```
predict2=model.predict(x_testing2)
error2=Actual2-predict2
RMSE2=round(np.sqrt(np.mean(error2**2)),2)
RMSE2
```

[26]: 2.74

[27]:
```
plt.figure(figsize=(8, 6))
plt.scatter(Actual2, predict2, color='teal', alpha=0.6)
plt.plot([Actual2.min(), Actual2.max()], [Actual2.min(), Actual2.max()],␣
↪color='red', linestyle='--')
plt.xlabel("Actual OCD Score")
plt.ylabel("Predicted OCD Score")
plt.title("Actual vs. Predicted OCD Scores")
plt.grid(True)
plt.show()
```

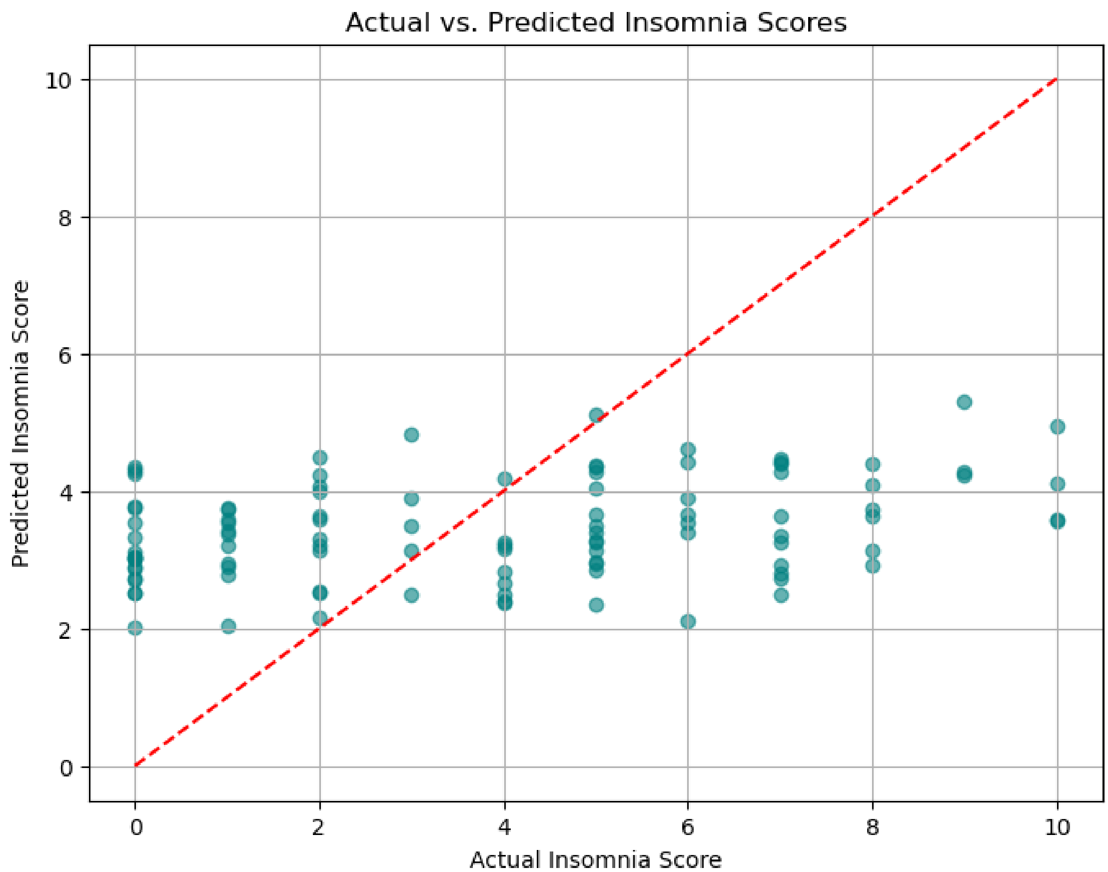


[28]:
```
w3data=projdff[genre_frequency_cols + ["Insomnia"]]
```

```
#Split 80/20

shuff3data=w3data.sample(frac=1, random_state=1).reset_index(drop=True)
shuff3data
split_index3=int(0.8*len(shuff3data))
training3=shuff3data[:split_index3]
testing3=shuff3data[split_index3:]

x_training3=training3[genre_frequency_cols]
y_training3=training3["Insomnia"]
x_testing3=testing3[genre_frequency_cols]
Actual3=testing3["Insomnia"]

#Linear Regression
model=LinearRegression()
model.fit (x_training3,y_training3)

predict3=model.predict(x_testing3)
error3=Actual3-predict3
RMSE3=round(np.sqrt(np.mean(error3**2)),2)
RMSE3
```

[28]: 2.87

[29]:
```
plt.figure(figsize=(8, 6))
plt.scatter(Actual3, predict3, color='teal', alpha=0.6)
plt.plot([Actual3.min(), Actual3.max()], [Actual3.min(), Actual3.max()],␣
↪color='red', linestyle='--')
plt.xlabel("Actual Insomnia Score")
plt.ylabel("Predicted Insomnia Score")
plt.title("Actual vs. Predicted Insomnia Scores")
plt.grid(True)
plt.show()
```

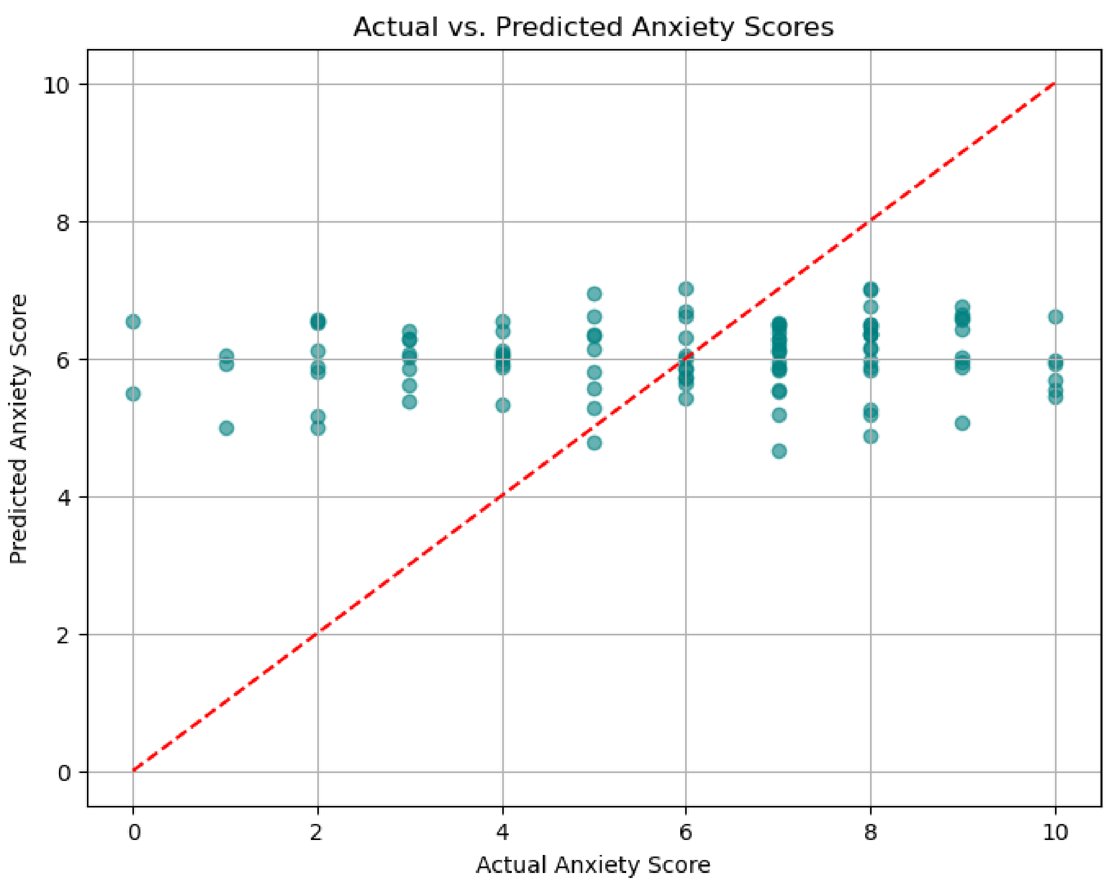


[30]:
```
w4data=projdff[genre_frequency_cols + ["Anxiety"]]

#Split 80/20

shuff4data=w4data.sample(frac=1, random_state=1).reset_index(drop=True)
shuff4data
split_index4=int(0.8*len(shuff4data))
training4=shuff4data[:split_index4]
testing4=shuff4data[split_index4:]

x_training4=training4[genre_frequency_cols]
y_training4=training4["Anxiety"]
x_testing4=testing4[genre_frequency_cols]
Actual4=testing4["Anxiety"]

#Linear Regression
model=LinearRegression()
model.fit (x_training4,y_training4)
```

```
predict4=model.predict(x_testing4)
error4=Actual4-predict4
RMSE4=round(np.sqrt(np.mean(error4**2)),2)
RMSE4
```

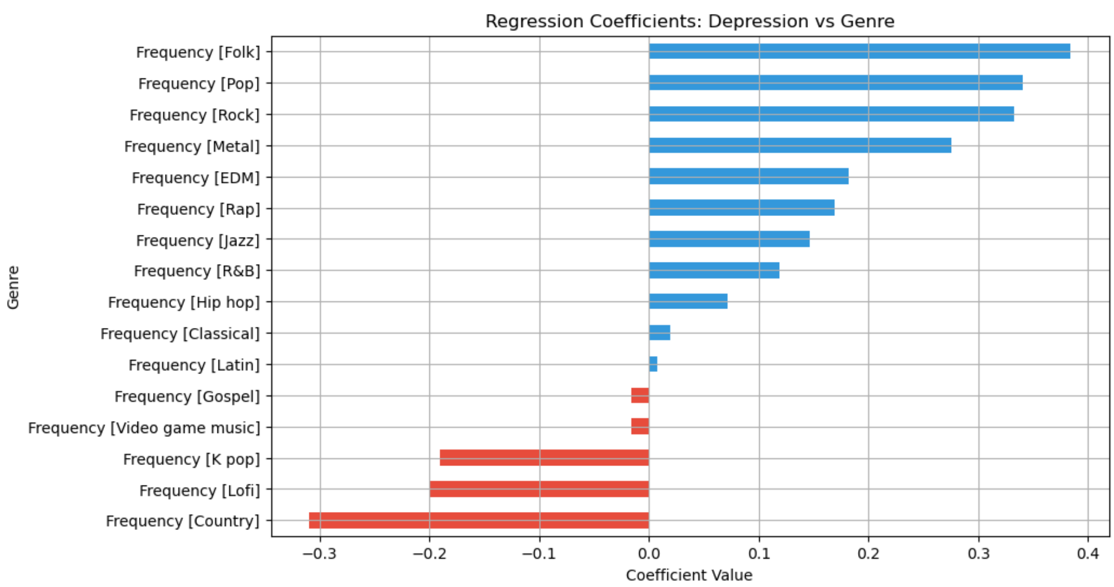
[30]: 2.53

[31]:
```
plt.figure(figsize=(8, 6))
plt.scatter(Actual4, predict4, color='teal', alpha=0.6)
plt.plot([Actual4.min(), Actual4.max()], [Actual4.min(), Actual4.max()],
 ↪color='red', linestyle='--')
plt.xlabel("Actual Anxiety Score")
plt.ylabel("Predicted Anxiety Score")
plt.title("Actual vs. Predicted Anxiety Scores")
plt.grid(True)
plt.show()
```



[32]:
```
freq_map = {"Never": 0, "Rarely": 1, "Sometimes": 2, "Very frequently": 3}
genre_columns = [col for col in df.columns if col.startswith("Frequency [")]
```

```
df_encoded = df[genre_columns + ["Depression"]].replace(freq_map).dropna()

# Regression
X = df_encoded[genre_columns]
y = df_encoded["Depression"]
model = LinearRegression()
model.fit(X, y)

# Plot coefficients
coefficients = pd.Series(model.coef_, index=X.columns).sort_values()
coefficients.plot(kind="barh", figsize=(10, 6), color=['#e74c3c' if c < 0 else␣
↪'#3498db' for c in coefficients])
plt.title("Regression Coefficients: Depression vs Genre")
plt.xlabel("Coefficient Value")
plt.ylabel("Genre")
plt.grid(True)
plt.show()
print(pd.Series(model.coef_, index=genre_frequency_cols))
```

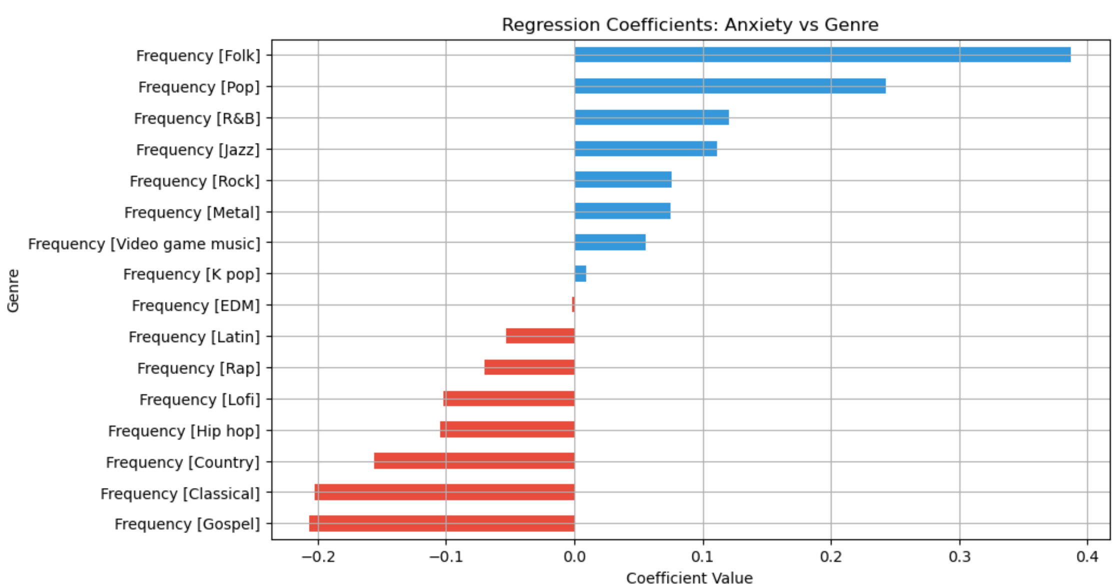



```
Frequency [Classical]          0.019109
Frequency [Country]           -0.309952
Frequency [EDM]                0.181888
Frequency [Folk]               0.384251
Frequency [Gospel]            -0.016262
Frequency [Hip hop]            0.071919
Frequency [Jazz]               0.146778
Frequency [K pop]             -0.190759
Frequency [Latin]              0.008051
```

```
Frequency [Lofi]                -0.200333
Frequency [Metal]                0.275752
Frequency [Pop]                  0.340778
Frequency [R&B]                  0.118955
Frequency [Rap]                  0.168806
Frequency [Rock]                 0.333113
Frequency [Video game music]    -0.016525
dtype: float64
```

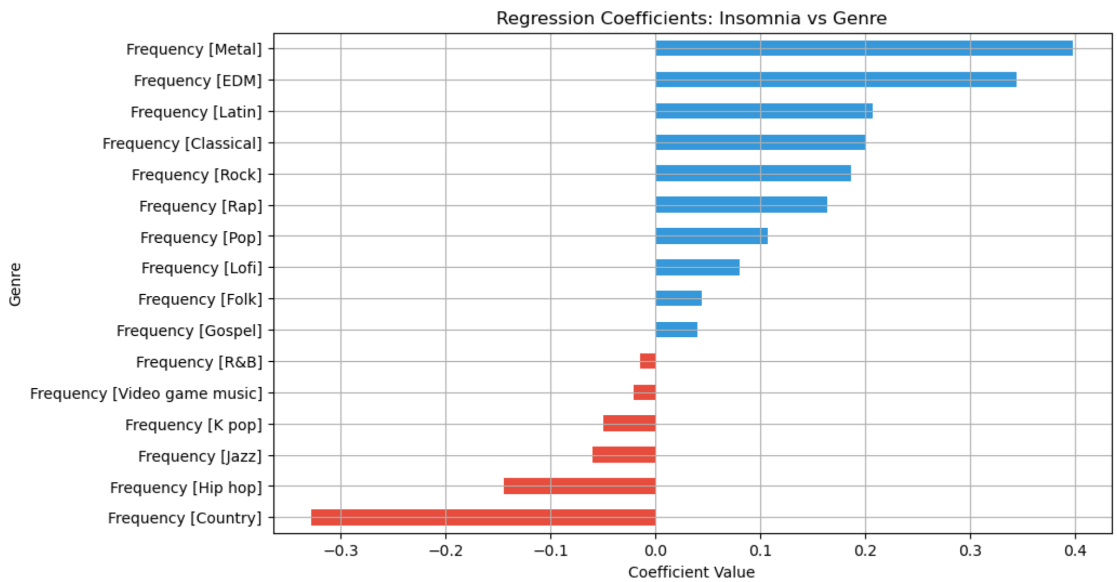
[33]:

```python
df_encoded = df[genre_columns + ["Anxiety"]].replace(freq_map).dropna()

# Regression
X = df_encoded[genre_columns]
y = df_encoded["Anxiety"]
model = LinearRegression()
model.fit(X, y)

# Plot coefficients
coefficients = pd.Series(model.coef_, index=X.columns).sort_values()
coefficients.plot(kind="barh", figsize=(10, 6), color=['#e74c3c' if c < 0 else␣
 ↪'#3498db' for c in coefficients])
plt.title("Regression Coefficients: Anxiety vs Genre")
plt.xlabel("Coefficient Value")
plt.ylabel("Genre")
plt.grid(True)
plt.show()
print(pd.Series(model.coef_, index=genre_frequency_cols))
```



```
Frequency [Classical]          -0.202433
Frequency [Country]            -0.156407
Frequency [EDM]                -0.001402
Frequency [Folk]                0.387302
Frequency [Gospel]             -0.206883
Frequency [Hip hop]            -0.104538
Frequency [Jazz]                0.111183
Frequency [K pop]               0.009280
Frequency [Latin]              -0.053198
Frequency [Lofi]               -0.102189
Frequency [Metal]               0.074817
Frequency [Pop]                 0.242546
Frequency [R&B]                 0.120146
Frequency [Rap]                -0.069791
Frequency [Rock]                0.075949
Frequency [Video game music]    0.055200
dtype: float64
```

[34]:
```python
df_encoded = df[genre_columns + ["Insomnia"]].replace(freq_map).dropna()

# Regression
X = df_encoded[genre_columns]
y = df_encoded["Insomnia"]
model = LinearRegression()
model.fit(X, y)

# Plot coefficients
coefficients = pd.Series(model.coef_, index=X.columns).sort_values()
coefficients.plot(kind="barh", figsize=(10, 6), color=['#e74c3c' if c < 0 else
 ↪'#3498db' for c in coefficients])
plt.title("Regression Coefficients: Insomnia vs Genre")
plt.xlabel("Coefficient Value")
plt.ylabel("Genre")
plt.grid(True)
plt.show()
print(pd.Series(model.coef_, index=genre_frequency_cols))
```

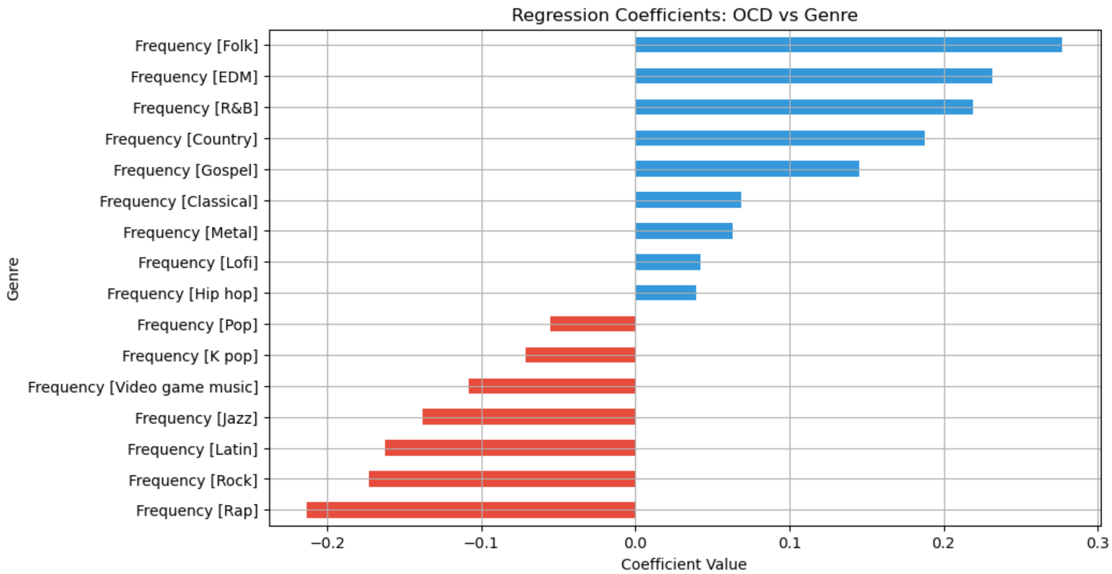



```
Frequency [Classical]           0.200083
Frequency [Country]            -0.328618
Frequency [EDM]                 0.344938
Frequency [Folk]                0.044008
Frequency [Gospel]              0.040448
Frequency [Hip hop]            -0.144732
Frequency [Jazz]               -0.059605
Frequency [K pop]              -0.049997
Frequency [Latin]               0.207056
Frequency [Lofi]                0.080030
Frequency [Metal]               0.398557
Frequency [Pop]                 0.107264
Frequency [R&B]                -0.014846
Frequency [Rap]                 0.164238
Frequency [Rock]                0.187048
Frequency [Video game music]   -0.020722
dtype: float64
```

[35]:

```
df_encoded = df[genre_columns + ["OCD"]].replace(freq_map).dropna()

# Regression
X = df_encoded[genre_columns]
y = df_encoded["OCD"]
model = LinearRegression()
model.fit(X, y)

# Plot coefficients
coefficients = pd.Series(model.coef_, index=X.columns).sort_values()
```

```
coefficients.plot(kind="barh", figsize=(10, 6), color=['#e74c3c' if c < 0 else
  '#3498db' for c in coefficients])
plt.title("Regression Coefficients: OCD vs Genre")
plt.xlabel("Coefficient Value")
plt.ylabel("Genre")
plt.grid(True)
plt.show()
print(pd.Series(model.coef_, index=genre_frequency_cols))
```




```
Frequency [Classical]           0.068546
Frequency [Country]             0.187615
Frequency [EDM]                 0.231243
Frequency [Folk]                0.277035
Frequency [Gospel]              0.145571
Frequency [Hip hop]             0.039797
Frequency [Jazz]               -0.137712
Frequency [K pop]              -0.071170
Frequency [Latin]              -0.162756
Frequency [Lofi]                0.041938
Frequency [Metal]               0.063273
Frequency [Pop]                -0.055169
Frequency [R&B]                 0.218837
Frequency [Rap]                -0.213405
Frequency [Rock]               -0.173100
Frequency [Video game music]   -0.108343
dtype: float64
```

7 Reflections/Conclusions/Next Step

What did we find? - After analyzing our data, we found that we need to carefully interpret the data and be cautious when drawing conclusions because there are a lot of limitations to the dataset that we are working with. - We also found that individuals who reported a high score for a certain mental health condition have also often reported a similar score for other mental health conditions, suggesting a possible correlation among those conditions. - We also found that people tend to listen to rock, metal, or pop rather than listen to other music genres, even though 74% of respondents claimed they explore other music genres/artists. - Finally, from our Linear Regression Models, the results show that certain genres — such as Pop, Folk, EDM, and Rock — tend to have higher positive associations with mental health conditions like depression and OCD. On the other hand, genres such as Country, Gospel, and Classical often show negative or milder relationships with mental distress. Albeit, we concluded that although there might be some association between music genre preference and mental health, it is not a strong standalone predictor. There could be other factors, such as genetics, environment, stress, etc., that could also contribute to mental health, and these differences don't imply causation.

Limitations/Challenges - We realized that there were a lot of limitations with the data, including how the data was self-reported as well as how small the entire dataset was to begin with

Next Steps? - We would need a larger dataset if we were to want a more accurate result. We could also include more factors that correlate to mental health.

8 Contributions

Each person in this group contributed equally to the assignment. The work was distributed equally.

9 Work Cited

- https://www.apa.org/monitor/2019/01/gen-z
- https://www.apa.org/news/press/releases/stress/2023/collective-trauma-recovery
- https://www.statista.com/topics/13068/gen-z-audio-consumption-in-the-us/#topicOverview -https://www.tmh.org/healthy-living/blogs/healthy-living/how-music-affects-your-mind-mood-and-body -https://www.kaggle.com/datasets/catherinerasgaitis/mxmh-survey-results?resource=download