Investigating Mental Health Through Music Listening Habits and Characteristics

Data Science Major Capstone | Ayesha Tariq, Class of 2024

WELLESLEY W

Background

It is crucial to understand factors and habits that may correlate with mental health, as mental health conditions are massively prevalent in society. Given the great historical and current presence of music throughout society, it is worthwhile to understand its connections with mental health outcomes.

This Capstone project seeks to answer the following questions:
- **How do music listening habits and characteristics of music relate to mental health outcomes?**
- **What variables correlate most greatly with anxiety outcomes?**

This study examines musical characteristics (such as energy, danceability level, valence) and listening habits (such as number of hours one listens to music) in relation to mental health outcomes, with a focus on anxiety, the most common mental health disorder.[1]

Data

Data was taken from an open-sourced dataset on Kaggle titled "Music and Mental Health Survey Results" as well as from Spotify using the spotifyr package.
- The survey data included:
  - Self-reported mental health scores on scales from 1 to 10
  - Frequency levels of listening to various genres
  - Age, Number of hours one listens to music, Favorite genres
  - Whether someone plays an instrument ("Instrumentalist")
- The Spotify data contained (for 32,000 songs):
  - Genre, Energy level, Danceability level, Valence (measure of musical positivity), Key information
- The two datasets were merged in the following manner:
  - The Spotify data was collapsed into rows separated by genre, with average characteristic values for the songs in the genre.
  - Weighted averages of genre listening frequency were calculated, multiplying the listening level of each genre by the average of the spotify music characteristic to get a characteristic score for each individual per music characteristic.
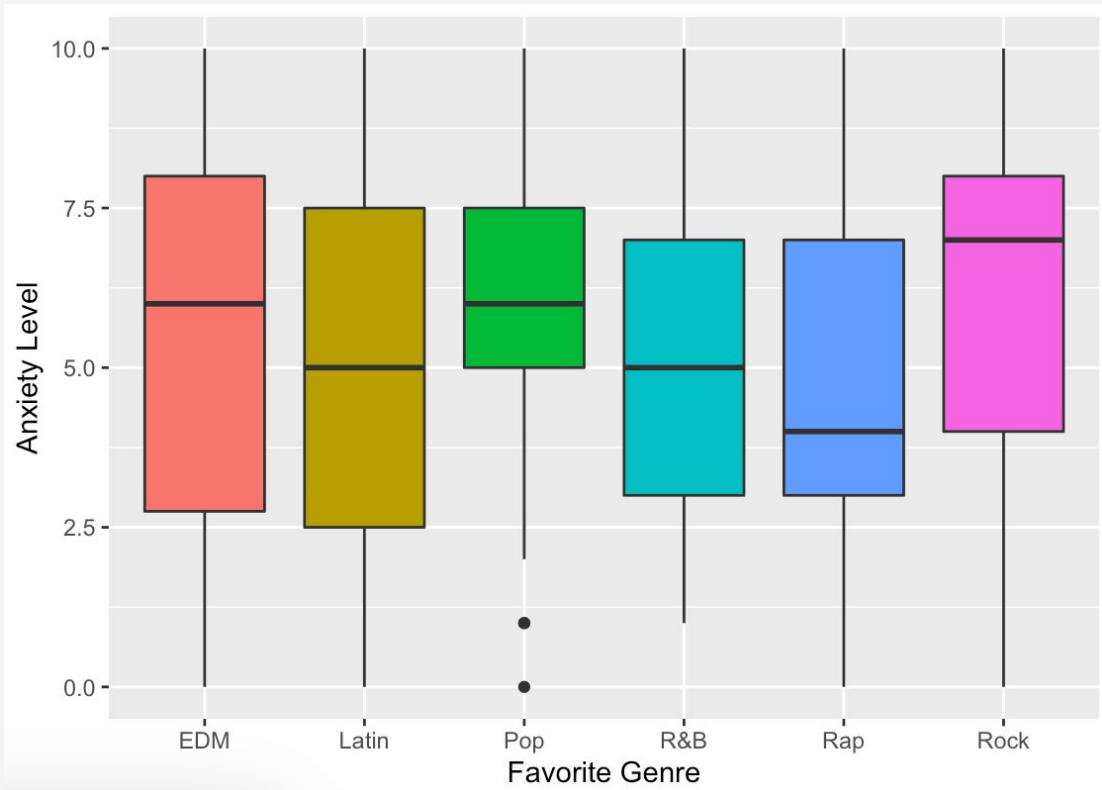

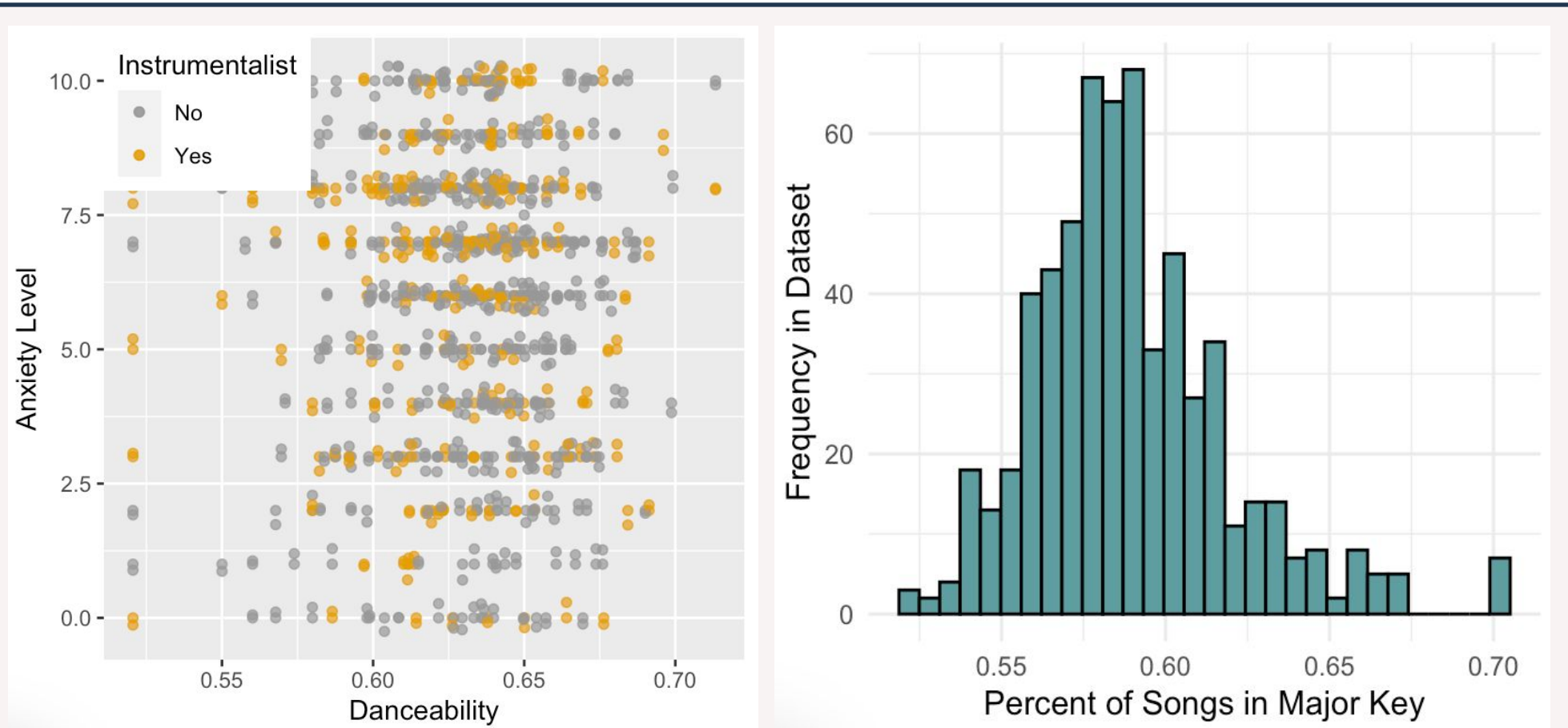Figure 1: Dataset Visualization - Favorite Genre and Anxiety Level

Figure 2: Anxiety level by Danceability level, categorizing by Instrumentalist variable

Figure 3: Distribution of weighted averages of songs in major key (based on genre)

Methods

Multiple first-order linear regression models were created to understand the relationship between the predictor variables and Anxiety outcome levels. Several musical characteristic variables were examined, such as energy, valence, and danceability, and survey variables such as age and hours of listening per day were also examined. Indirectly, genre listening frequency was somewhat accounted for due to the weighted averages. Models were then compared using ANOVA testing.

	Predictors	R^2	Adjusted R^2	BIC
Model 1	Age, Hours Per Day, Valence	0.04327	0.03852	2961.36
Model 2	Depression, Insomnia, OCD	0.3582	0.355	2718.222
Model 3	Age, Hours Per Day, Depression, Insomnia, OCD	0.3748	0.3696	2715.1
Model 4	Age, Hours Per Day, Valence, Depression	0.3091	0.3045	2769.566
Model 5	Age, Hours Per Day, Valence, Depression, Insomnia, OCD	0.3749	0.3686	2721.42

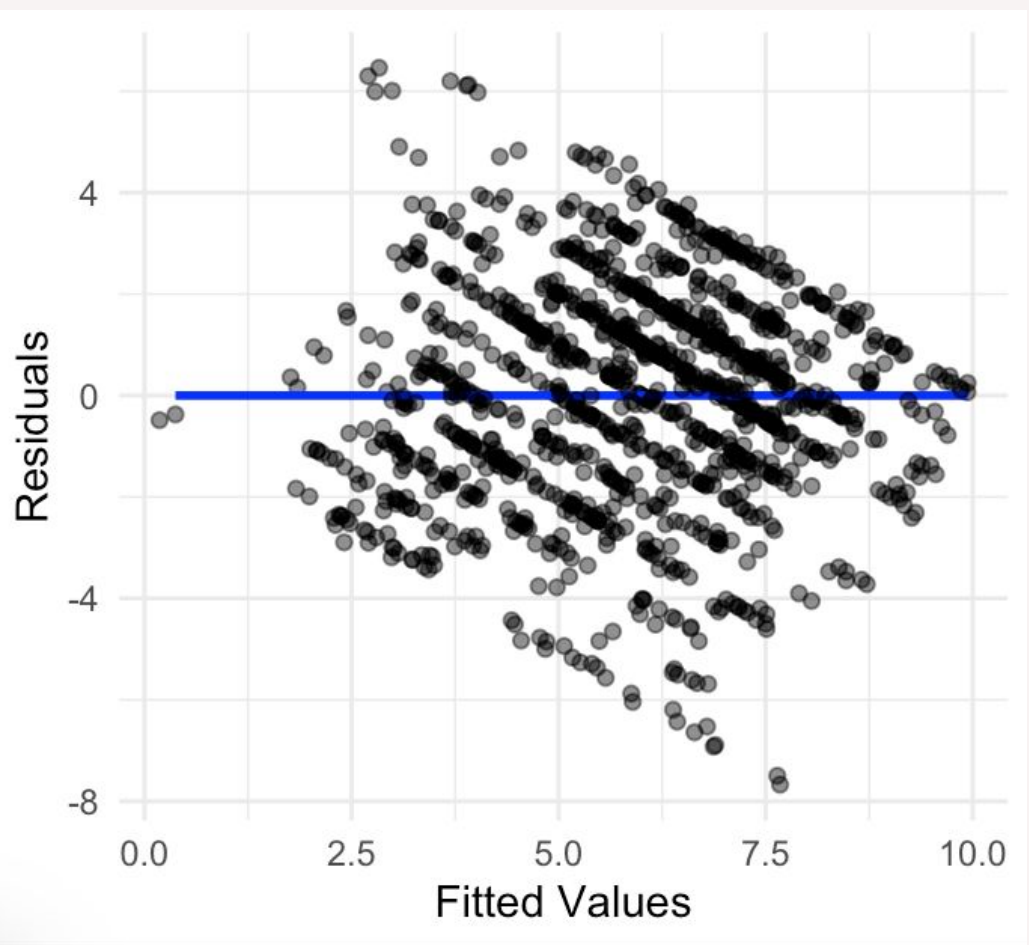Figure 4: R squared, Adjusted R squared, and BIC values for fitted models

Figure 5: Residuals vs Fitted Values of Model 3

Results

- The model with the lowest BIC and highest Adjusted R^2 was Model 3. The key variables to creating the best model were Age, Depression, and OCD, all for which ($p<0.001$). All other predictors had $p>0.05$.
- The musical characteristic variable of variance did not improve Model 5's fit as compared to Model 3 ($p=0.5336$).
- The listening habit variable of Hours Per Day did improve the fit of Model 3 from Model 2.

	Anxiety		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	3.81	(3.22 , 4.39)	**<0.001**
Age	-0.03	(-0.05 , -0.01)	**<0.001**
Hours per day	-0.03	(-0.09 , 0.03)	0.280
Insomnia	0.06	(-0.01 , 0.12)	0.076
Depression	0.43	(0.36 , 0.49)	**<0.001**
OCD	0.24	(0.17 , 0.30)	**<0.001**
Observations	609		
R^2 / R^2 adjusted	0.375 / 0.370		

Figure 6: Coefficients, Confidence Intervals, and P Values of Model 3

Discussion

Ultimately, certain predictors were most greatly correlated with Anxiety outcomes: age, Depression levels, and OCD levels. These findings correspond with literature examining the correlation of mental health conditions, which has revealed mental health comorbidity to be common.[2] Future work addressing the following limitations may better understand the musical characteristic variables.

Limitations/Considerations
- The mental health scores were self-reported, raising concern about data validity
- Weighted averages did not account for music listening times (though the Hours Per Day variable was included in modeling)
- Spotify dataset only contained 6 genres, survey dataset included 16

Acknowledgements

Thank you to Dr. Cassandra Pattanayak, my Capstone advisor, for your help and guidance.

References

[1] "Anxiety Disorders." World Health Organization, 27 Sept. 2023, https://www.who.int/news-room/fact-sheets/detail/anxiety-disorders
[2] "Comorbidity within mental disorders: a comprehensive analysis based on 145 990 survey respondents from 27 countries." National Library of Medicine, 12 Aug. 2020, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7443806/
[3] "Music & Mental Health Survey Results." Kaggle, https://www.kaggle.com/datasets/catherinerasgaitis/mxmh-survey-results.