Investigating Patterns in Music Preference and Mental Health

MGT6203: Data Analytics in Business
Spring 2023

<u>Team 4</u>
Paul Hee Jai Kim
Thuy Thu Nguyen
Cynthia Campbell
Charles Westberg
Shuqi Xiao

Introduction

Background

A study funded by the National Institute of Mental Health estimated that in the US, almost two hundred Billion dollars in lost earnings is caused by serious mental illness, and depression has been attributed to 400 million lost days of work annually [1]. On a global scale, depression and anxiety cause an estimated $1 trillion in lost productivity each year [2]. Music therapy is a lucrative niche market with a value of $2.4 billion in 2021 and is expected to grow to $4.42 billion by 2028 [3]. Based on research from Harvard Health, music intervention has a significant impact on mental health, and numerous studies have shown that some types of music have been used to treat mental problems such as anxiety, depression, stress [4].

Researchers have studied the significant impact of music intervention on mental health and smaller improvements in physical health [5]. However, while music has been found to have a positive impact on the mental health and quality of life for people, there is no best method of intervention or 'dose' of music that works best for all people. As music has been shown to have an uplifting effect on people and is a constant in many lives, it is worth exploring what kinds of patterns exist between peoples' choice in music and their self-reported mental health. Additionally, identifying how the profile of music varies between genres may reveal meaningful insights in regards to mental health. The findings may have profound impact for multiple business sectors, including streaming platforms, advertisers, and public health.

Consider the possibility of utilizing modeling techniques to look at trends across popular songs, combined with mental health survey questions to aid in helping individuals cope with potential mental health issues. This could be another tool in assisting those who suffer from anxiety or depression, but are uncomfortable speaking to a healthcare provider. Furthermore, music services could suggest curated song/playlists to 'balance' your mood. Balancing out the mental state of individuals would lead to an overall better workforce, which would increase productivity in all sectors. The goal here is to have a significant impact in reducing the aforementioned loss in productivity ($1 trillion/400 million lost work days).

Additionally, from another angle, advertisers for pharmaceutical companies may be interested in identifying key target demographics to reach. Working in tandem with music streaming platforms, there could be significant financial incentives to identify particular audiences who may benefit from certain medications.

Problem Statement & Hypothesis

The purpose of this investigation is to identify patterns between mental health scores and music preference, and to determine if certain popular music genres can be used to profile and aid as a prediction tool for potential mental health issues.

The initial hypotheses are as follows:

1. There **is** a relationship between listening habits and reported mental health status (e.g., those who listen to rock often may report lower mental health scores than ones who frequent jazz).
2. Music genres are identifiable by their aspects (e.g., country music has high acoustic scores; dance has high energy).
3. Due to recent pandemic-induced social isolation and economic challenges, most popular songs in the last few years are linked to lower mental health scores.

Data

Data Description

1. *Music and Mental Health Survey [Mental Health Survey Data] (Appendix A)*

This dataset contains background information and listening habits of 736 respondents, their listening frequency of 16 music genres (ranked Never, Rarely, Sometimes, and Very Frequently), and their experience] to 10 [experience regularly/to an extreme]).

2. *Top Hits 1980-2022 Compiled [Top 100 Hits of 1980 & Spotify Playlist Analyzer] (Appendix B)*

The second set of data was compiled manually. A playlist analyzer for Spotify was utilized, which analyzes songs in a Spotify playlist and identifies their musical attributes as well as genre. The top 100 songs of each year from 1980 to 2022 were analyzed this way, totaling 4,153 songs.

Data Cleaning

Data cleaning was facilitated through Excel and initial analysis of data quality was performed using pivot tables. New CSV files were generated with the selected columns that will be used in the analyses.

1. *Steps taken to clean up the Music and Mental Health Survey data*
    a. Removed columns: Timestamp, Instrumentalist, Composer, Fav genre, Exploratory, Foreign, Permissions
    b. Removed the listed columns based on song genre frequency in playlist, keeping the most frequent genres: Frequency [Classical], Frequency [Gospel], Frequency [Jazz], Frequency [K pop], Frequency [Lofi], Frequency [Rap], Frequency [Video game music]
    c. Transformed responses for Frequency questions using variables indicated in Appendix C. Created new columns for Country, EDM, Folk, Hip Hop, Latin, Metal, Pop, R&B, Rock
    d. Created a composite score column averaging the ratings for Anxiety, Depression, Insomnia & OCD, weighing each status equally (25% each)
    e. Cleaned dataset is displayed in Appendix D
2. *Top Hits 1980-2022 Compiled*
    a. Removed duplicate songs and the following columns: #, Popularity, Dup Check, Genres (kept Parent Genre), Album, Album Date, Time, Loud, Key, Time Signature, Added At, Album Label, Camelot, Spotify Track Img, Song Preview
    b. Duplicated song information & attributes for every listed parent genre for the song (i.e., if a song had multiple genres, row was duplicated for each listed genre). This generated a list of songs with only one listed genre
    c. Chose genres for the analyses based on the frequency of the genre in the Top Hits data (Appendix E)
    d. Removed songs with genre listed as “undefined” or “blank.
    e. Cleaned dataset is displayed in Appendix F

Key Variables

Independent: Music Genre, Beats per Minute (BPM), Danceability, Energy, Acousticness, Happiness, Speechiness, Liveliness, Listening Frequency.

Dependent: Mental Health Score (Anxiety, Depression, Insomnia, OCD) individually and as a composite score.

Exploratory Data Analysis

It was found that each attribute is statistically correlated with every other attribute (significance at 0.1% level) with correlation coefficients between 0.20 to 0.52 (Figure A). The strength of their relationships are not exceedingly large, so individual relationships between music and each of these conditions were evaluated.

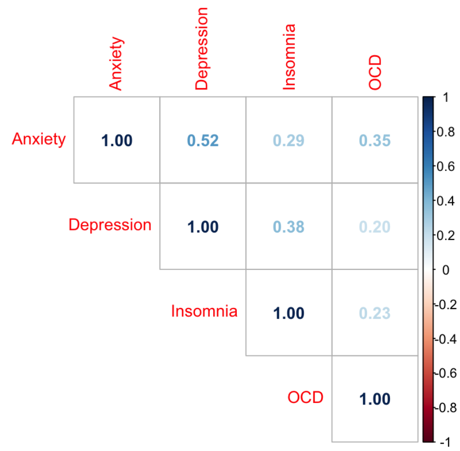


Figure A. Correlations between mental health attributes

With the Top Hits dataset, box plots (Figure B) for all musical attributes were generated to visualize whether the attributes are distinct. Unfortunately, the musical attributes are *not* especially distinct per genre, with the exception of Liveliness for the "New Age" genre (attributes broken down by genre are shown in Appendix G). The range for BPM is quite narrow, and apart from a select few genres, the distributions for most attributes are very similar.

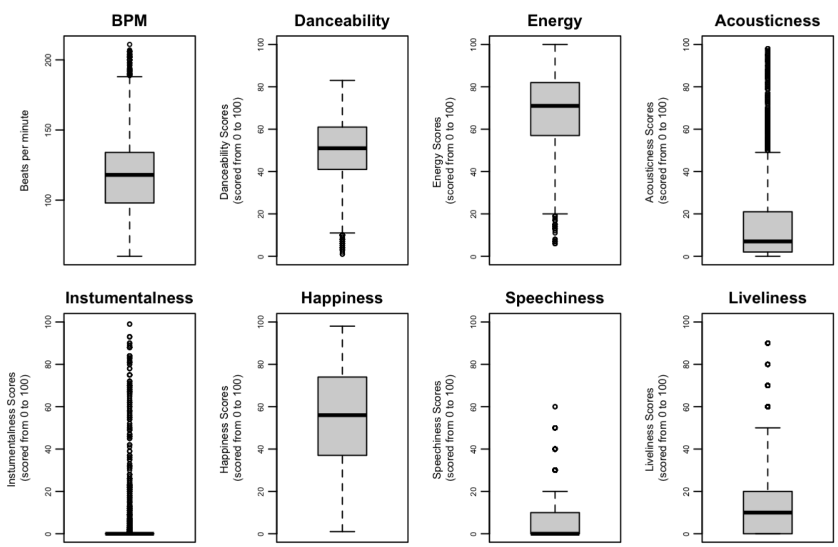


Figure B. Box plots of music attributes for **all songs** in Top Hits dataset

Continuing, principal component analysis (PCA) was performed on the Top Hits data, to see if it is at least possible to visualize distinct clusters for the songs in 2-dimensions. Unfortunately, as hinted at by the previous finding, it appears that the songs are all very closely grouped together, and it is difficult to identify any unique groupings (Appendix H). An additional principal component (PC) was added to visualize in 3D (Appendix I), and while it appears that there *may* be 2 new clusters that begin to separate out along the 3rd axis, those groups are small in size, and the majority of songs remain clumped together. This suggests that it will be rather difficult to distinguish and identify specific musical characteristics about each genre.

Methodology

Simple linear regression

Simple linear regression was run 55 times, each time with:

- Frequency of listening to a music genre or the respondents' age (Age) or the amount of time listening to music per day (Hours.per.day) as the predictor
- A mental health score (including the composite score) as the response

Out of 55 regressions, 25 show significance at the 95% confidence interval. The coefficients for these relationships range from -0.026 to 0.564. This shows that even when there is a correlation between the predictors and mental health states, the effects are quite marginal. The adjusted R^2 only goes up to 3.58%, further demonstrating that the relationship is weak (Appendix J).

Multivariate linear regression

Next, multivariate linear regression was used, accompanied by stepwise regression using AIC as the criteria for variable selection.

Table 1. Select multivariate linear regression results

Response	Predictors selected with best AIC	Adjusted R^2
Anxiety	Age + Folk + Pop + Rock	0.0467
Depression	Age + Hours.per.day + Country + Folk + Hip.Hop + Metal + Rock	0.0778
Insomnia	Hours.per.day + Country + Latin + Metal	0.0438
OCD	Age + Hours.per.day + Country + EDM	0.0351
Composite	Age + Hours.per.day + Country + EDM + Folk + Metal + Pop	0.0656

Interesting points to note are:

- The respondents' age (Age), amount of time listening to music per day (Hours.per.day), and frequency of listening to Country music (Country) are constantly shown to have an effect on the respondents' reported mental health status - each appearing 4 out of 5 times
- On the other hand, the respondents' frequencies of listening to Hip-hop, Latin, and R&B are selected only once or less.

The best adjusted R^2 only goes up to 7.8%, which is slightly better than simple linear regression but still quite low.

Logistics regression

The dependent variables are given as a range between 0-10, so conversion to binary form (0 and 1) is required before running logistic regression. The median self-reported score for each mental health issue was chosen as the threshold to ensure approximate class balance.

Logistic regression was run for each response, with predictors selected using AIC. The variables selected for each model and corresponding AUCs are given below.

Table 2. AUC for logistic regressions for each response variable evaluated

logistic_regression_selected	auc
Anxiety_binary ~ Age + Folk + Rock	0.5982
Depression_binary ~ Age + Hours.per.day + Country + EDM + Latin + Metal + Rock	0.6507
Insomnia_binary ~ Hours.per.day + Country + EDM + Latin + Metal	0.6135
OCD_binary ~ Age + Hours.per.day + Country + EDM	0.6201
Composite_binary ~ Age + Hours.per.day + Country + EDM + Folk + Metal + Rock	0.6395

AUCs range between 0.59 - 0.66, indicating weak predictive power. However, it's notable that Age, Hours.per.day, and Country are once again selected more often.

Decision Trees/Random Forest

Next, decision trees were explored as a way to categorize and predict mental health composite scores from the dataset. First, using stepwise regression to limit the number of factors going into predicting the responses, the best factors that were used to predict the composite scores were the age, frequency of listening to metal, frequency of listening to folk, frequency of listening to EDM, and hours of listening to music per day. For this decision tree (Figure C), folk and EDM were determined not to have a significant decision point to categorize the data. However, the calculated R^2 value for this model was only 11%. This means that even the best predictors were not great at explaining the mental health scores with a single decision tree.

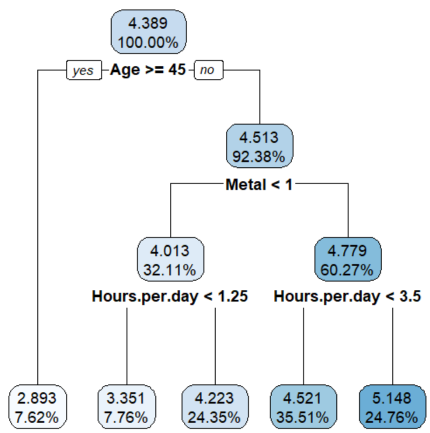


Figure C. Decision tree for Composite scores after stepwise variable selection.

Sampling multiple decision trees to generate a random forest model is the next logical step to look for a better predictor. However, the random forest model only yielded a R^2 value of 12% so was negligibly better than the singular decision tree (Figure D).

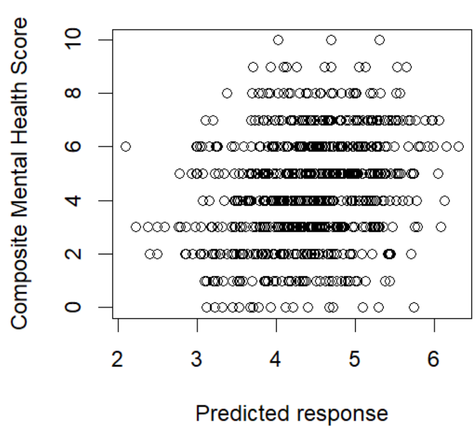


Figure D. Predicted composite response from random forest model vs the actual Composite response.

The composite score being an overall combination of mental health scores may not be the best way to find patterns within this data. Using the same process, decision trees for each type of mental health were also generated. Through stepwise variable selection, the predictors for each tree (see Appendix K) were used to create a decision tree for each mental health response (Appendix L). Overall, the decision tree R^2 values were poor (Anxiety 9.6%, Depression 12.8%, OCD 3.4%, and Insomnia 5.1%) . However, it is interesting to consistently see that age and hours listened per day had the most impact on many of these decision trees. This may suggest that the type of music matters less for mental health than an individual's age and their time spent listening to music.

To further explore the idea that younger and older people surveyed may have different decision trees, the data was split into two groups: one for people 25 and younger and one for people older than 25 (since 25 is the mean age of the respondents). These groups were then used to construct separate decision trees. The younger decision tree models did not produce good models and the older decision tree models yielded only slightly better R^2 values (Composite 20.7%, Anxiety 18.5%, Depression 14.4%, OCD 26.9%, and Insomnia 26%) (Figure E). Since these models are with a smaller sample size, these models are most likely overfit to the data. With a larger sample size of data, it would be interesting to see if older adults tend to have better defined trends than the younger adults as these models suggest.

To test the idea that respondents that listened to different amounts of music may have different decision trees, the data was split into two groups: one for people who listen to music for 3 hours and less and one for people who listen for more than 3 hours (since the median hours per day of respondents was 3). The respondents who listened to 3 hours or less of music per day yielded poor models however those that listened to more than 3 hours a day had better models (Composite 27.6%, Anxiety 16.1%, Depression 14.4%, OCD 12.4%, and Insomnia 20.2%) (Figure F). These models also are likely to be overfit as these models are from a smaller subset of data. However, the difference in these model R^2 values suggests that better trends between mental health and music can be seen in those that listen to music more throughout their day.

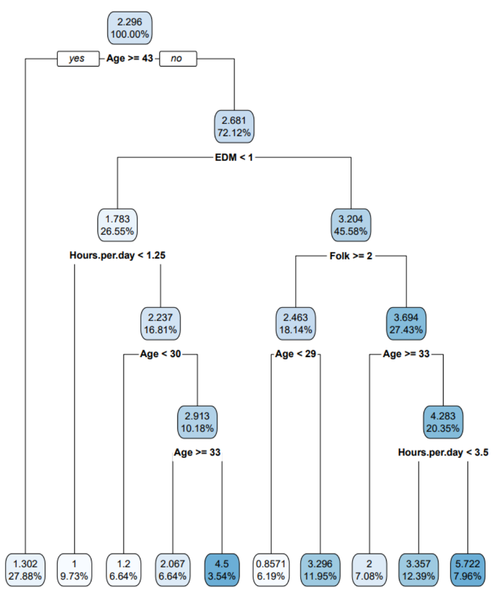


Figure E. Best model examples were OCD (left), Insomnia (center), and Composite (right) decision trees for those older than 25.

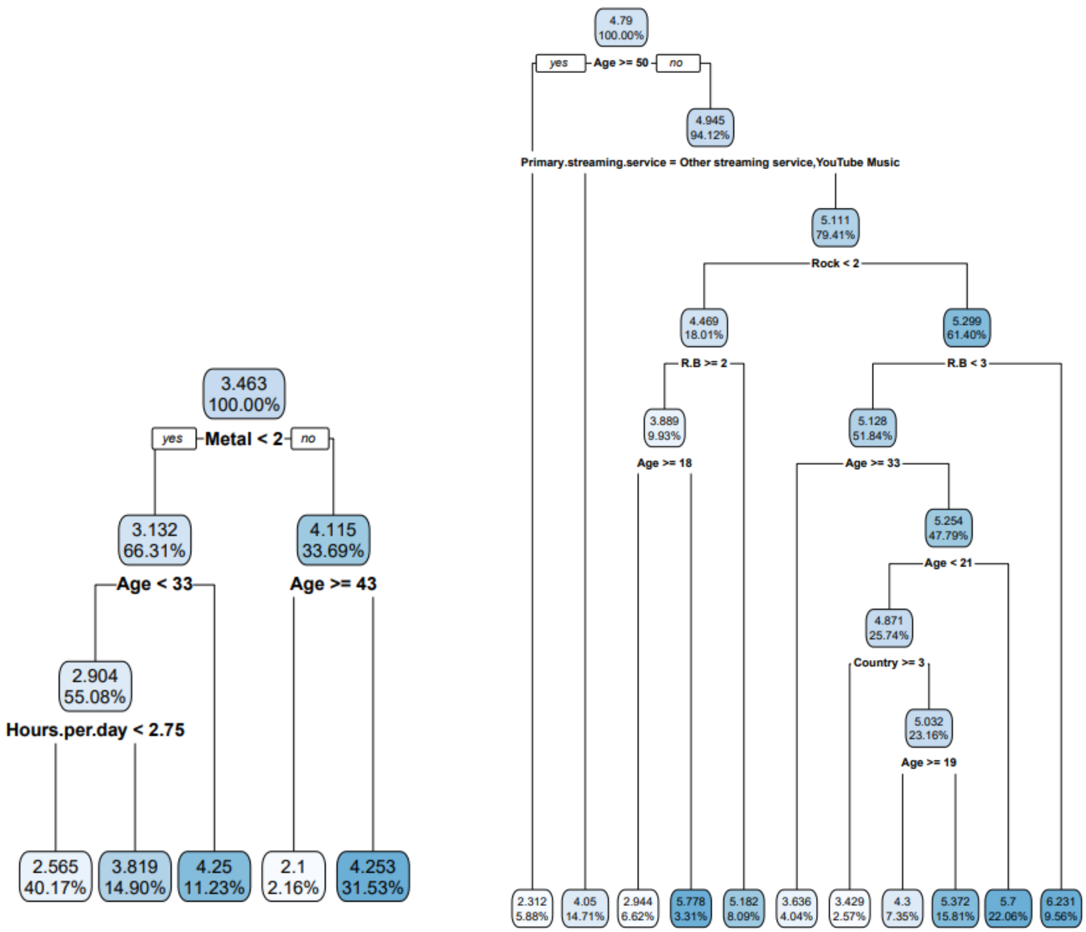


Figure F. Best model examples were Insomnia (left) and Composite (right) decision trees for those that listen to more than 3 hours of music.

Clustering/Classification

K-means clustering algorithm was implemented to identify how many clusters should be created for the Top Hits music data. An elbow plot (Appendix M) was generated, which indicated that between 4-6 clusters should be used. However, upon clustering the dataset into 6 groups and plotting them in a 2-dimensional plot (Appendix N), it was evident that the clusters did not separate out the songs in any meaningful way, at least as visible in 2 dimensions. This is likely because the songs are not distinct enough in their characteristics.

Next, two different classification models (KNN and SVM) were tuned to determine how well a model can be built to predict the parent genre of a song, given the music attributes (BPM, Danceability, Energy, etc) as predictors. The accuracy plots for both KNN and SVM classification models show the optimal hyperparameters for each respective model (Figure G).

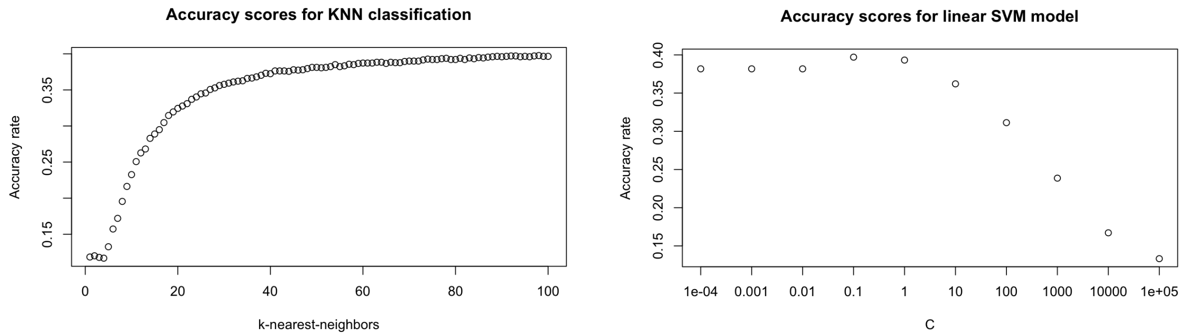


Figure G. Tuning for optimal hyperparameters for k-Nearest Neighbor (KNN) model

The selection of the hyperparameters significantly impacted the models' qualities, but even at their optima (k>60 or C=0.1), neither model was able to achieve better than 40% accuracy. While this is not outstanding performance, it is much better than initially expected based on the difficulty in clustering the dataset. With 15 original "Parent.Genres," a *random guess* would expect ~6.67% accuracy, so it could be argued that the models do a reasonable job at identifying the correct genre of the songs. Nonetheless, if it is not possible to link specific genres of music to certain mental health outcomes, these classification models become less meaningful.

Further, to expand on the analysis of the Top Hits dataset, it was sought to determine whether there have been any trends in the music attributes over the years (from 1989 to 2022). The mean values for each feature was plotted for each year to visually gauge whether there have been any significant fluctuations (Figure H).

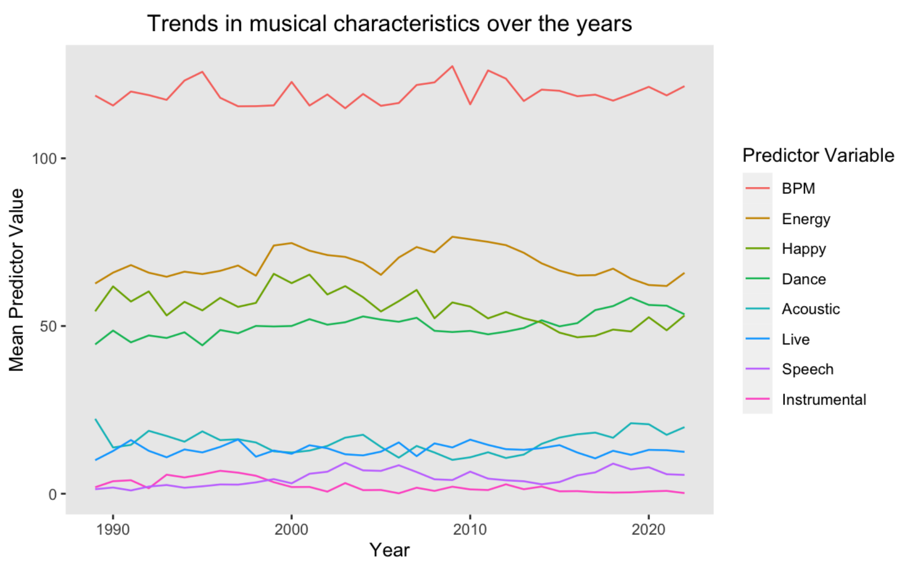


Figure H. Changes in average musical attribute scores from 1989 to 2022

It was observed that, while there are not *obvious* changes/trends in these features, some of them may exhibit some subtle shifts over time. To confirm whether there are any statistically significant variations over time, regressions were generated for each attribute against the years. Overall, it was found that 4 attributes did possess some significant change over the years (Table 2).

Table 3. Attributes with significant change over the years

term	estimate	std.error	statistic	p.value
Dance	0.2492166	0.01925293	12.944347	7.894741e-38
Instrumental	-0.1376856	0.01450301	-9.493583	3.121959e-21
Happy	-0.3271647	0.03012439	-10.860460	3.144505e-27
Speech	0.1452135	0.01178741	12.319369	1.817838e-34

Most notably, the “Happiness” attribute had the largest decrease at 0.33 units/year. This suggests that, over the 33 years included in the data, the average Happiness score for the songs that became Top Hits fell by roughly 11 units. Moreover, the remaining findings appear to confirm anecdotal observation that the most popular songs in recent years have become more geared toward Dancing (think EDM, hip-hop, etc), with less Instrumental characteristics. This is a rather interesting find that may enable future exploration. Since the most popular hits over the years have trended in this manner, future research that can yield better insights into whether “Happy” songs are actually correlated with improved mental health outcomes, for instance, may help define future health trajectories based on these markers.

Discussion

While performing exploratory data analysis upon the data set it was discovered that some initial assumptions of the data were not present. In the Top Hits data, the genres did not show the expected unique separation. Across all genres, the factors used to score the genres were not very useful for performing clustering analysis as the survey results did not show enough of a distribution across multiple factors. Moving forward in future research it could prove fruitful to combine the survey results with a true ‘happiness score’ that was formulated based off of a coded hedonometer [6], which has created happiness evaluations on over 10,000 words. This would provide a more substantial scoring metric than the arbitrary scoring system from the survey. Diving deeper into the song components could also provide more valuable insight and tie a listener's mental health to songs, as research has suggested that major and minor chords and their prevalence is tied with happiness and sorrow in songs [7]. The most critical component of a new survey would be its ability to separate out genre characteristics at a level that is understandable to participants, yet able to provide quality data.

The factor song characteristics could also have been collected using a ‘baseline’ question or two to help organize the participants. Listing out several songs and having respondents circle what emotions those songs convey could help in setting a starting point, or provide some extra critical clustering potential. What kind of emotions a participant feels when songs are mentioned would be interesting to compare to their favorite genre. An example could be asking if someone felt amused by the song ‘Yakkity Yak’. A study from the University of California at Berkeley identified thirteen unique emotions provoked by music [8]. Attempting to correlate this ‘baseline’ with their mental health level and music choices could provide interesting feedback and aid in building a stronger predictive/prescriptive model.

The mental health survey also shows a similar weakness, which is a known disadvantage of surveys. With most of the responses of this survey being very median-centric, having more quantifiable questions would be more beneficial here as well. Rather than the frequency factors of ‘Never,’ ‘Rarely,’ ‘Sometimes,’ and ‘Very Frequently,’ answers based on how many songs/times a week one listens to certain genres could be more impactful. Adding numerical values for frequency could provide more insight and also show true separation between the answers. Further, the composite scores of mental health did not represent a very useful distribution, with a vast majority of the scores all falling close to the mean. Thus, a more robust survey is likely needed to develop meaningful models for business use.

The standardized 0-10 scoring system used to gather information on the anxiety, depression, OCD, and insomnia of the respondents could also benefit from an adjustment. Given a scale of 0-10 for anxiety, many people may put themselves around the middle. Does a respondent who likes a tidy house report a high level of OCD? Should that be graded the same as someone undergoing treatment for OCD or psychosis? Having a smaller answer set with better descriptions of what each answer represents could be beneficial. Better yet, a more controlled experiment/study can be conducted by collecting information from individuals who have been clinically diagnosed with a certain condition and having that be in the 8-10 range as a point of reference for participants.

This report serves as a pilot for future research and provides insight into developing a more powerful research project to gather more intelligence regarding the effects of music on the mental health of the population. A survey that wishes to focus on correlation between music and its effects on mental health would have to build questions that are understandable with fewer open-ended questions. Given the need for gathering results as honestly as possible, it may behoove future researchers to utilize interviewers, or survey aids to help walk respondents through questions. This option is more expensive, but could clear some of the hurdles discussed above and elicit more honest responses and potentially move away from everyone scoring in the middle. Reducing the known weaknesses of surveys, but still trying to utilize their strengths will be the challenge. Balancing questions to minimize dishonesty, confusion, or responses that paint one in a seemingly unfavorable light will be the key to gathering truly meaningful data.

The financial incentive for companies to get into the music therapy market is still substantial and should not be overlooked. Reducing the one trillion in lost revenue due to mental health illness in the workplace is incentive enough. The loss in productivity due to mental stressors is well worth the investment in the workforce by company leadership, and having a drug-free means of providing aid for sufferers of mental illnesses would help the overall population live happier lives.

References

[1] Kessler, RC, Heeringa S, Lakoma MD, Petukhova M, Rupp AE, Schoenbaum M, Wang PS, Zaslavsky AM. “The individual-level and societal-level effects of mental disorders on earnings in the United States: Results from the National Comorbidity Survey Replication.” *American Journal of Psychiatry*, 2008 May.

[2] Sime, C., “The Cost of Ignoring Mental Health in the Workplace.” Published in Forbes.com. 17 April 2019. Accessed at: https://www.forbes.com/carleysime/the-cost-of-ignoring-mental-health.

[3] SkyQuest Technology Consulting Pvt. Ltd., “Music Therapy Market to Generate Revenue of $4.42 Billion by 2028 | Over 50% Business in Music Therapy are Less than 5 Years Old.” Published in GlobeNewswire. 21 September 2022. Accessed at: https://www.globenewswire.com/Music-Therapy-Market-to-Generate-Revenue-of-4-42-Billion-by-2028

[4] Kubicek, L., “Can music improve our health and quality of life?” Published in Harvard Health Publishing, Harvard Medical School. 25 July 2022. Accessed at: https://www.health.harvard.edu/can-music-improve-our-health-and-quality-of-life.

[5] Frost, A., “How Music Can Improve Your Mental Health.” Published by The Jed Foundation. Accessed at: https://jedfoundation.org/resource/how-music-can-improve-your-mental-health.

[6] Dodds, P, et.al., “Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter.” Published by Plos One. 7 December 2011. Accessed at: https://journals.plos.org/plosone/article

[7] Wagner, Andrew.,”What makes a happy song? Chances are it has more seventh chords.” Published in Science.org. 14 November 2017. Accessed at: https://www.science.org/what-makes-happy-song

[8] Cowen, A.,”What music makes us feel: at least 13 dimensions organize subjective experiences associated with music across different cultures. Published in Proceedings for National Academy for Sciences. 25, June 2019. Accessed at: https://www.pnas.org/doi/10.1073/pnas.1910704117

Appendices

Appendix A: Music and Mental Health Survey

Frequency	Frequency	Frequency	Frequency	Frequency	Frequency	Frequency	Frequency	Frequency	Frequency	Anxiety	Depressio	Insomnia	OCD	Music effe	Permissions
Never	Very frequ	Very frequ	Rarely	Never	Very frequ	Sometime	Very frequ	Never	Sometime	3	0	1	0		I understand.
Very frequ	Rarely	Sometime	Rarely	Never	Sometime	Sometime	Rarely	Very frequ	Rarely	7	2	2	1		I understand.
Rarely	Very frequ	Never	Sometime	Sometime	Rarely	Never	Rarely	Rarely	Very frequ	7	7	10	2	No effect	I understand.
Very frequ	Sometime	Very frequ	Sometime	Never	Sometime	Sometime	Never	Never	Never	9	7	3	3	Improve	I understand.
Never	Very frequ	Sometime	Sometime	Never	Sometime	Very frequ	Very frequ	Never	Rarely	7	2	5	9	Improve	I understand.
Very frequ	Very frequ	Rarely	Very frequ	Rarely	Very frequ	Very frequ	Very frequ	Very frequ	Never	8	8	7	7	Improve	I understand.
Sometime	Never	Rarely	Rarely	Rarely	Rarely	Rarely	Never	Never	Sometime	4	8	6	0	Improve	I understand.
Rarely	Very frequ	Never	Sometime	Never	Sometime	Sometime	Rarely	Never	Rarely	5	3	5	3	Improve	I understand.
Never	Never	Never	Never	Very frequ	Never	Never	Never	Very frequ	Never	2	0	0	0	Improve	I understand.

Appendix B: Top Hits 1980-2022 Compiled

1	Year	#	Song	Artist	Popularity	BPM	Genres
2	2015	16	Don't	Bryson Tiller	80	97	kentucky hip hop, pop, r&b, rap, trap
3	1999	92	Don't Call Me Baby	Madison Avenue	54	125	disco house, vocal house
4	1995	12	Don't Look Back In Anger - Remastered	Oasis	69	163	beatlesque, britpop, madchester, permanent wave, pop rock, rock
5	2019	38	Don't Start Now	Dua Lipa	76	124	dance pop, pop, uk pop
6	2009	14	Down	Jay Sean,Lil Wayne	75	132	dance pop, pop rap, post-teen pop, urban contemporary, hip hop, new orleans
7	1994	59	Dreamer - Janice Robinson Vocal	Livin' Joy	56	127	disco house, diva house, hip house, vocal house
8	2004	14	Drop It Like It's Hot	Snoop Dogg,Pharrell Williams	74	92	g funk, gangster rap, hip hop, pop rap, rap, west coast rap, pop
9	2020	8	Dynamite	BTS	0	114	k-pop, k-pop boy group
10	2021	12	Easy On Me	Adele	81	142	british soul, pop, pop soul, uk pop
11	1998	45	Every Morning	Sugar Ray	64	110	alternative metal, alternative rock, funk metal, neo mellow, pop rock, rock
12	1984	29	Everything She Wants	Wham!	66	115	europop, new romantic, new wave pop
13	2015	100	Fast Car	Jonas Blue,Dakota	64	114	dance pop, edm, pop, pop dance, tropical house, uk dance,
14	2000	60	Fill Me In	Craig David	59	132	british soul, dance pop, pop rap, r&b, urban contemporary
15	1998	17	Fly Away	Lenny Kravitz	71	160	permanent wave, pop rock, rock
16	1992	35	Found Out About You	Gin Blossoms	58	135	alternative rock, neo mellow, permanent wave, pop rock, post-grunge, rock, te
17	1999	93	Get It On Tonite	Montell Jordan	54	99	contemporary r&b, dance pop, hip hop, hip pop, new jack swing, pop rap, r&b,
18	1982	29	Give It Up	KC & The Sunshine Band	66	125	disco, funk, soft rock
19	2021	5	good 4 u	Olivia Rodrigo	23	169	pop
20	2007	38	Hate That I Love You	Rihanna,Ne-Yo	70	94	barbadian pop, pop, urban contemporary, dance pop
21	2020	65	Head & Heart (feat. MNEK)	Joel Corry,MNEK	76	123	dance pop, edm, pop, pop dance, tropical house, uk dance, uk contemporary r8
22	2021	22	Heat Waves	Glass Animals	77	81	gauze pop, indietronica, shiver pop
23	2015	65	Hello	Adele	71	79	british soul, pop, pop soul, uk pop
24	2015	86	Here	Alessia Cara	67	120	canadian contemporary r&b, canadian pop, dance pop, electropop, pop, post-t
25	1995	61	Hey Lover	LL COOL J,Boyz II Men	57	88	east coast hip hop, golden age hip hop, hardcore hip hop, hip hop, hip pop, old

1	Year	#	Song	Parent Genres	Album	Album Date
2	2015	16	Don't	Hip Hop, Pop, R&B	TRAPSOUL	10/2/2015
3	1999	92	Don't Call Me Baby	Dance/Electronic	Don't Call Me Baby	1/1/1999
4	1995	12	Don't Look Back In Anger - Remastered	Folk/Acoustic, Rock, Pop	(What's The Story) Morning Glory? (Deluxe Remastered Edition)	1995
5	2019	38	Don't Start Now	Pop	Don't Start Now	10/31/2019
6	2009	14	Down	Pop, Hip Hop	All Or Nothing	1/1/2009
7	1994	59	Dreamer - Janice Robinson Vocal	Dance/Electronic, Pop	Donâ€™t Stop Movinâ€™	1/1/1997
8	2004	14	Drop It Like It's Hot	Hip Hop, Pop	R&G (Rhythm & Gangsta): The Masterpiece	1/1/2004
9	2020	8	Dynamite	Pop	Dynamite	8/21/2020
10	2021	12	Easy On Me	R&B, Pop	Easy On Me	10/14/2021
11	1998	45	Every Morning	Metal, Rock, Folk/Acoustic, Pop	14:59	12/23/1998
12	1984	29	Everything She Wants	undefined	Make It Big	10/23/1984
13	2015	100	Fast Car	Pop, Dance/Electronic	Fast Car	12/4/2015
14	2000	60	Fill Me In	R&B, Pop, Hip Hop	Born to Do It	8/14/2000
15	1998	17	Fly Away	Rock, Pop	5	5/12/1998
16	1992	35	Found Out About You	Rock, Folk/Acoustic, Pop, Metal	New Miserable Experience	1/1/1992
17	1999	93	Get It On Tonite	Pop, Hip Hop, R&B	Best Of Montell Jordan	9/25/2015
18	1982	29	Give It Up	undefined	All In a Night's Work (Expanded Version)	3/11/2016
19	2021	5	good 4 u	Pop	good 4 u	5/14/2021
20	2007	38	Hate That I Love You	Pop	Good Girl Gone Bad: Reloaded	6/2/2008
21	2020	65	Head & Heart (feat. MNEK)	Pop, Dance/Electronic, R&B	Head & Heart (feat. MNEK)	7/3/2020
22	2021	22	Heat Waves	Pop, Rock	Dreamland	8/7/2020
23	2015	65	Hello	R&B, Pop	25	11/20/2015
24	2015	86	Here	R&B, Pop	Know-It-All	11/13/2015
25	1995	61	Hey Lover	Pop, Hip Hop, R&B	Mr. Smith (Deluxe Edition)	11/20/1995

Appendix B: Top Hits 1980-2022 Compiled (continued)

1	Year	#	Song	Time	Dance	Energy	Acoustic	Instrumental	Happy	Speech	Live	Loud	Key	Time S
2	2015	16	Don't	3:18	77	36	22	0	19	20	10	-6	B Minor	4
3	1999	92	Don't Call Me Baby	3:48	81	98	6	1	96	0	30	-7	D#/Eâ™- Minor	4
4	1995	12	Don't Look Back In Anger - Remastered	4:49	33	94	7	0	31	0	10	-3	C Major	4
5	2019	38	Don't Start Now	3:03	79	79	1	0	68	0	10	-5	B Minor	4
6	2009	14	Down	3:32	73	68	1	0	73	0	0	-4	D Major	4
7	1994	59	Dreamer - Janice Robinson Vocal	3:43	74	76	1	0	83	0	20	-7	F Minor	4
8	2004	14	Drop It Like It's Hot	4:26	89	63	19	0	66	20	10	-4	Câ™¯/Dâ™- Major	4
9	2020	8	Dynamite	3:19	75	77	1	0	74	10	0	-4	F#/Gâ™- Minor	4
10	2021	12	Easy On Me	3:44	60	37	58	0	13	0	10	-8	F Major	4
11	1998	45	Every Morning	3:39	83	68	8	0	98	0	0	-4	G#/Aâ™- Major	4
12	1984	29	Everything She Wants	5:02	90	46	32	0	96	0	10	-16	F#/Gâ™- Minor	4
13	2015	100	Fast Car	3:32	64	57	48	0	53	0	30	-7	A Major	4
14	2000	60	Fill Me In	4:17	68	74	38	1	83	0	0	-7	G#/Aâ™- Major	4
15	1998	17	Fly Away	3:41	59	87	2	0	74	0	60	-5	G Major	4
16	1992	35	Found Out About You	3:53	54	83	0	5	67	0	0	-8	G Major	4
17	1999	93	Get It On Tonite	4:37	81	50	26	0	86	0	0	-10	A#/Bâ™- Minor	4
18	1982	29	Give It Up	4:14	86	62	2	0	88	0	0	-12	D#/Eâ™- Major	4
19	2021	5	good 4 u	2:58	56	66	30	0	67	20	10	-5	F#/Gâ™- Minor	4
20	2007	38	Hate That I Love You	3:38	64	73	32	0	73	0	10	-5	F Minor	4
21	2020	65	Head & Heart (feat. MNEK)	2:46	73	87	17	0	91	0	0	-3	G#/Aâ™- Major	4
22	2021	22	Heat Waves	3:58	76	53	44	0	53	0	0	-7	B Major	4
23	2015	65	Hello	4:55	58	43	33	0	29	0	0	-6	F Minor	4
24	2015	86	Here	3:19	38	82	8	0	33	10	0	-4	C Major	4
25	1995	61	Hey Lover	4:44	71	43	34	0	54	0	10	-12	F Minor	4

1	Year	#	Song	Added At	Spotify Track Id	Album Label	Camelot	Spotify Track Img	Song Preview
2	2015	16	Don't	6/20/2020	3pXF1nA74528Edde4of9CC	TrapSoul/RCA Records	10A	https://p.scdn.co/mp3-pr	https://i.scdn.co/image/ab67616d00004851d5f3cea8affdca01a0dc754f
3	1999	92	Don't Call Me Baby	4/22/2021	4fazGt0v2v2zkzddQOcDZD	Vicious	2A	https://p.scdn.co/mp3-pr	https://i.scdn.co/image/ab67616d00004851b8808f12af8bac493f2933bb
4	1995	12	Don't Look Back In Anger - Remastered	1/19/2021	12dU3vAh6AFoJkisorfoUl	Big Brother Recordings Ltd	8B	https://p.scdn.co/mp3-pr	https://i.scdn.co/image/ab67616d000048517a4c8c59851c88f6794c3cbf
5	2019	38	Don't Start Now	6/22/2020	6WrI0LAC5M1Rw2MnX2ZvEg	Warner Records	10A	https://p.scdn.co/mp3-pr	https://i.scdn.co/image/ab67616d000048518583df1a14bea9175f9ac520
6	2009	14	Down	8/7/2020	6cmm1LMvZdB5zsCwX5BjqE	Cash Money	10B		https://i.scdn.co/image/ab67616d00004851e207a14471e5356294146e9d
7	1994	59	Dreamer - Janice Robinson Vocal	1/19/2021	6wetvpPWooBdmAEOKnDhpo	UMC (Universal Music Catalogue)	4A		https://i.scdn.co/image/ab67616d000048517d1b4387ef61788f72715186
8	2004	14	Drop It Like It's Hot	8/7/2020	2NBQmPrOEEjA8VbeWOQGxO	Geffen	3B		https://i.scdn.co/image/ab67616d00004851e803716268c173c3f9a0c057
9	2020	8	Dynamite	8/21/2020	0v1x6rN6JHRapa03JElljE	2020 BigHit Entertainment	11A		https://i.scdn.co/image/ab67616d000048512f86d9710377e63bfbc82ba8
10	2021	12	Easy On Me	10/14/2021	0gplL1WMoJ6iYaPgMCL0gX	Columbia	7B	https://p.scdn.co/mp3-pr	https://i.scdn.co/image/ab67616d0000485150dba34377a595e35f81b0e4
11	1998	45	Every Morning	1/19/2021	2ouURa1AIXp3AvkS52Jry5	RT Industries	4B	https://p.scdn.co/mp3-pr	https://i.scdn.co/image/ab67616d00004851d4d4c752eccb62aa42f6487d
12	1984	29	Everything She Wants	1/25/2021	5hXEcqQhEjfZdbIZLO8mf2	Columbia	11A	https://p.scdn.co/mp3-pr	https://i.scdn.co/image/ab67616d00004851a2fc41b0dd6ce4f0d16a4c46
13	2015	100	Fast Car	6/20/2020	2mCF8L0brIs88eH6Kf2h9p	Positiva	11B		https://i.scdn.co/image/ab67616d00004851c1faca501ed4c2752524da96
14	2000	60	Fill Me In	8/7/2020	0UzsDmdpw0Q14KU4hieQss	Sony Music UK	4B	https://p.scdn.co/mp3-pr	https://i.scdn.co/image/ab67616d000048517c2e92fb2302f8e8fcd9b389
15	1998	17	Fly Away	1/19/2021	1OxclUqVmVYxT6427tbhDW	Virgin Records	9B		https://i.scdn.co/image/ab67616d00004851f1157b7dcd21bae0c2c75d89
16	1992	35	Found Out About You	2/1/2023	5WmDRnuGYo31xrWNDcYaps	A&M	9B		https://i.scdn.co/image/ab67616d00004851e3100bdcdc758b5fab7e4894
17	1999	93	Get It On Tonite	4/22/2021	0AcLrSfAEBQcUnHOTm5pXg	Def Soul	3A		https://i.scdn.co/image/ab67616d00004851430debb283066ac6725483cf
18	1982	29	Give It Up	1/25/2021	3yDhZq8f17SmumVmEyCaRN	Epic/Legacy	5B	https://p.scdn.co/mp3-pr	https://i.scdn.co/image/ab67616d00004851070bc5edcb93d3f394b9445d
19	2021	5	good 4 u	5/18/2021	6PERP62TejQjgHu81OHxgM	Olivia Rodrigo PS	11A		https://i.scdn.co/image/ab67616d00004851670ec029374e082f921f9f74
20	2007	38	Hate That I Love You	8/7/2020	7iu0WYLdo4yksKf3seaxzI	Def Jam Recordings	4A		https://i.scdn.co/image/ab67616d00004851f9f27162ab1ed45b8d7a7e98
21	2020	65	Head & Heart (feat. MNEK)	8/20/2020	6cx06DFPPHchuUAcTxznu9	Atlantic Records UK	4B	https://p.scdn.co/mp3-pr	https://i.scdn.co/image/ab67616d0000485191e93c59bacfe819db9601eb
22	2021	22	Heat Waves	9/20/2021	3USxtqRwSYz57Ewm6wWRMp	Polydor Records	1B		https://i.scdn.co/image/ab67616d00004851712701c5e263efc8726b1464
23	2015	65	Hello	6/20/2020	62PaSfnXSMyLshYJrlTuL3	XL Recordings	4A	https://p.scdn.co/mp3-pr	https://i.scdn.co/image/ab67616d000048512d51dcc321ff0d476d3ec7e6
24	2015	86	Here	6/22/2020	5zUQZjVB6bfewBXWqsP9PY	EP Entertainment, LLC / Def Jam	8B		https://i.scdn.co/image/ab67616d00004851d7ef1ffbd2a5821152f35ce1
25	1995	61	Hey Lover	1/19/2021	5wG7d4cNogw0ETKaICPEYA	LL Cool J.	4A		https://i.scdn.co/image/ab67616d0000485118a12da5d84a8f12312698ab

Appendix C: Mental Health Data

Frequency Qualitative Answer:	Translated Number:
Never	0
Rarely	1
Sometimes	2
Very frequently	3

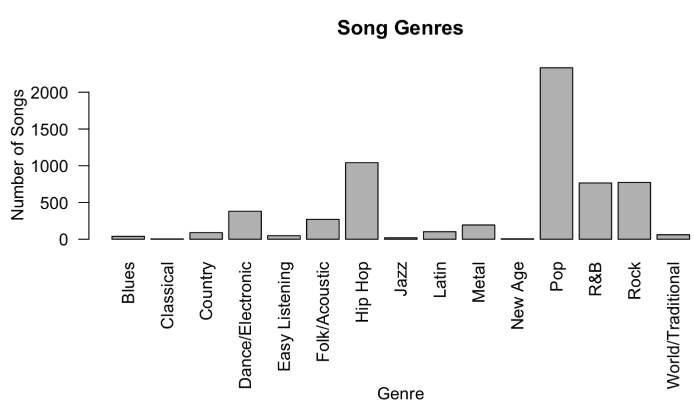
Appendix D: Cleaned Mental Health Survey Dataset

Age	Primary streaming service	Hours per day	While working	BPM	Frequency [Country]	Country	Frequency [EDM]	EDM	Frequency [Folk]	Folk	Frequency [Hip hop]	Hip Hop	Frequency [Latin]	Latin
18	Spotify	3	Yes	156	Never	0	Rarely	1	Never	0	Sometimes	2	Very frequently	3
63	Pandora	1.5	Yes	119	Never	0	Never	0	Rarely	1	Rarely	1	Sometimes	2
18	Spotify	4	No	132	Never	0	Very frequently	3	Never	0	Rarely	1	Never	0
61	YouTube Music	2.5	Yes	84	Never	0	Never	0	Rarely	1	Never	0	Very frequently	3
18	Spotify	4	Yes	107	Never	0	Rarely	1	Never	0	Very frequently	3	Sometimes	2
18	Spotify	5	Yes	86	Sometimes	2	Never	0	Never	0	Sometimes	2	Rarely	1
18	YouTube Music	3	Yes	66	Never	0	Rarely	1	Sometimes	2	Rarely	1	Rarely	1
21	Spotify	1	Yes	95	Never	0	Rarely	1	Never	0	Very frequently	3	Never	0
19	Spotify	6	Yes	94	Very frequently	3	Never	0	Sometimes	2	Never	0	Never	0
18	I do not use a stream	1	Yes	155	Rarely	1	Rarely	1	Rarely	1	Rarely	1	Rarely	1
18	Spotify	3	Yes		Very frequently	3	Never	0	Never	0	Never	0	Never	0

Age	Primary streaming service	Frequency [Metal]	Metal	Frequency [Pop]	Pop	Frequency [R&B]	R&B	Frequency [Rock]	Rock	Anxiety	Depression	Insomnia	OCD	Composite	Music effects
18	Spotify	Never	0	Very frequently	3	Sometimes	2	Never	0	3	0	1	0	4	
63	Pandora	Never	0	Sometimes	2	Sometimes	2	Very frequently	3	7	2	2	1	12	
18	Spotify	Sometimes	2	Rarely	1	Never	0	Rarely	1	7	7	10	2	26	No effect
61	YouTube Music	Never	0	Sometimes	2	Sometimes	2	Never	0	9	7	3	3	22	Improve
18	Spotify	Never	0	Sometimes	2	Very frequently	3	Never	0	7	2	5	9	23	Improve
18	Spotify	Rarely	1	Very frequently	3	Very frequently	3	Very frequently	3	8	8	7	7	30	Improve
18	YouTube Music	Rarely	1	Rarely	1	Rarely	1	Never	0	4	8	6	0	18	Improve
21	Spotify	Never	0	Sometimes	2	Sometimes	2	Never	0	5	3	5	3	16	Improve
19	Spotify	Very frequently	3	Never	0	Never	0	Very frequently	3	2	0	0	0	2	Improve
18	I do not use a stream	Never	0	Sometimes	2	Sometimes	2	Sometimes	2	2	2	5	1	10	Improve
18	Spotify	Never	0	Rarely	1	Rarely	1	Rarely	1	7	7	4	7	25	No effect

Appendix E: Song List Compiled

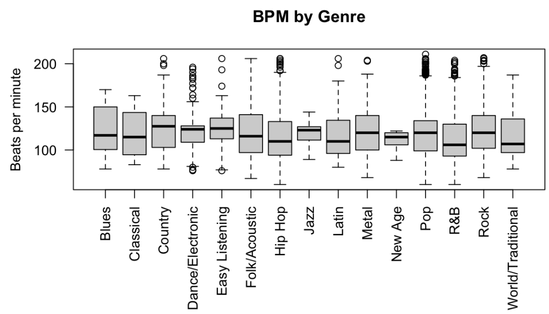
Row Labels	Count of Song	Data Clean-up Reasoning	Cleaned Genre:
Blues	39	Not in Mental Health Survey	n/a
Classical	4	Not in Mental Health Survey	n/a
Country	90	9th highest song count	Country
Dance/Electronic	382	5th highest song count	EDM
Easy Listening	49	Not in Mental Health Survey	n/a
Folk/Acoustic	270	6th highest song count	Folk
Hip Hop	1041	2nd highest song count	Hip Hop
Jazz	19	low count	n/a
Latin	103	8th highest song count	Latin
Metal	194	7th highest song count	Metal
New Age	6	Not in Mental Health Survey	n/a
Pop	2333	1st highest song count	Pop
R&B	766	4th highest song count	R&B
Rock	773	3rd highest song count	Rock
undefined	896	remove	n/a
World/Traditional	61	Not in Mental Health Survey	n/a
(blank)	28	remove	n/a
Grand Total	**7054**		

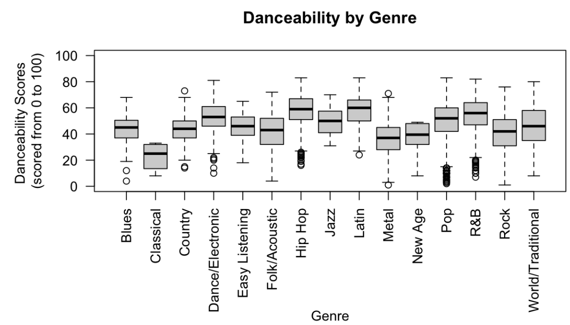


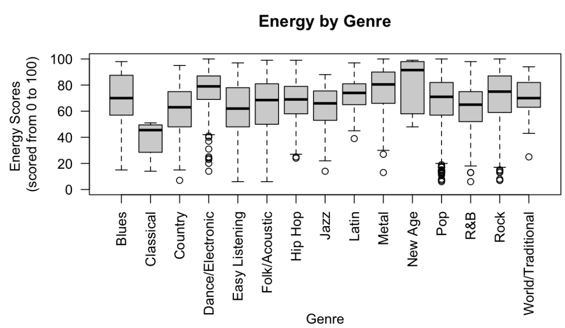
Appendix F: Cleaned Top Hits Data

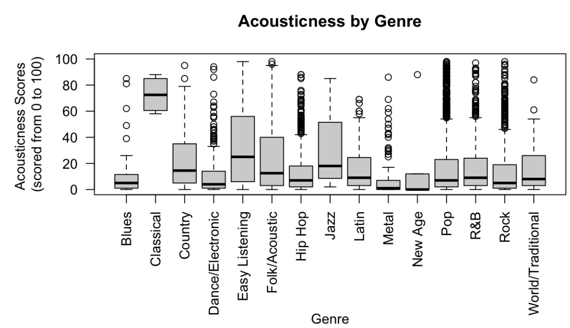
Year	Song	Artist	BPM	Parent Genres	Dance	Energy	Acoustic	Instrumental	Happy	Speech	Live	Spotify Track Id
2003	The Game of Love (feat. Michelle Branch) - Main / Radio Mix	Santana,Michelle Branch	120	Blues	59	93	7	0	67	0	0	41IShViF79gmulEkIHAjcb
2007	Brianstorm	Arctic Monkeys	165	Blues	42	98	0	0	44	20	20	5rTIpPWeO0IL4HWlGWrz5G
1996	Change the World	Eric Clapton	97	Blues	72	56	49	1	54	0	10	5Ds0VGkTSQ1jf4KzLUpZPb
2013	Do I Wanna Know?	Arctic Monkeys	85	Blues	55	53	19	0	41	0	20	5FVd6KXrgO9B3JPmC8OPst
2007	Fluorescent Adolescent	Arctic Monkeys	112	Blues	65	81	0	0	82	0	10	7e8utCy2JlSB8dRHKi49xM
2005	I Bet You Look Good On The Dancefloor	Arctic Monkeys	103	Blues	54	95	0	0	78	0	30	29EkMZmUNz1WsuzaMtVo1i
1998	My Father's Eyes	Eric Clapton	93	Blues	81	55	1	0	78	0	20	2GGskYwS4j8LDMSDUJ8vrl
1992	Tears in Heaven - Acoustic Live	Eric Clapton	79	Blues	69	33	81	0	46	0	70	3UqHlIBI771FNCiLY5MKrp
2006	When The Sun Goes Down	Arctic Monkeys	169	Blues	35	88	3	0	41	20	10	0ZRrJTPXDToRJ2iLo9oLrW
2013	Why'd You Only Call Me When You're High?	Arctic Monkeys	92	Blues	69	63	5	0	80	0	10	086myS9r57YsLbJpU0TgK9
1998	He Got Game	Public Enemy,Stephen Stills	97	Blues	83	64	7	2	87	10	0	48zajk3g8P3BMT7F9BbDoD
2000	Maria Maria (feat. The Product G&B)	Santana,The Product G&B	98	Blues	78	60	4	0	68	10	0	3XKIUb7HzIF1Vu9usunMzc
1999	Smooth (feat. Rob Thomas)	Santana,Rob Thomas	116	Blues	61	92	16	0	96	0	30	0n2SEXB2qoRQg171q7XqeW
1992	It's Probably Me	Sting,Eric Clapton	91	Blues	79	41	17	0	51	0	0	1X3Vb1oIAW6Ee22JZAEi59

Appendix G: Distribution of music attributes by genre

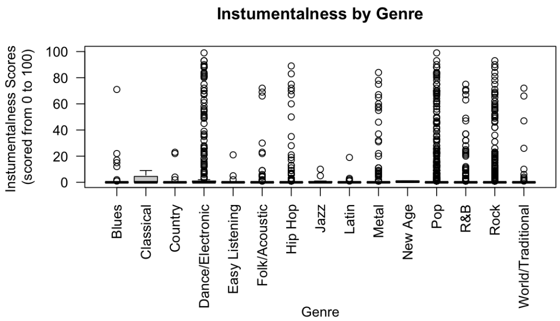


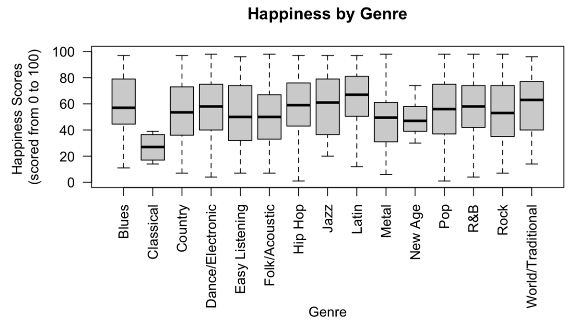


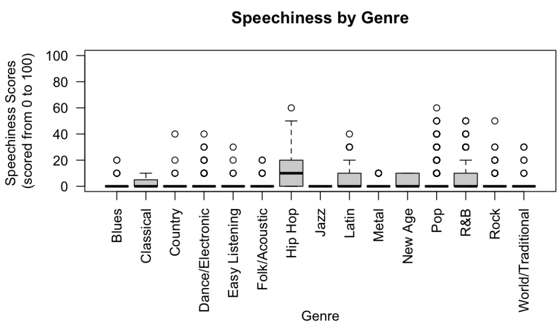


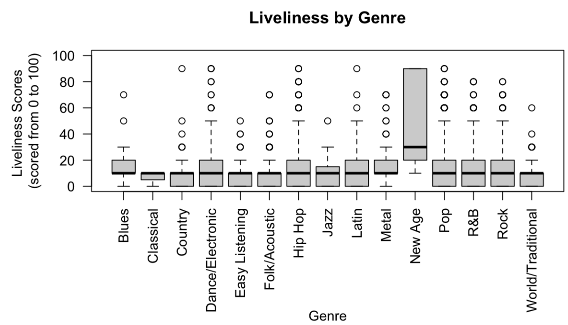


Appendix G: Distribution of music attributes by genre (continued)

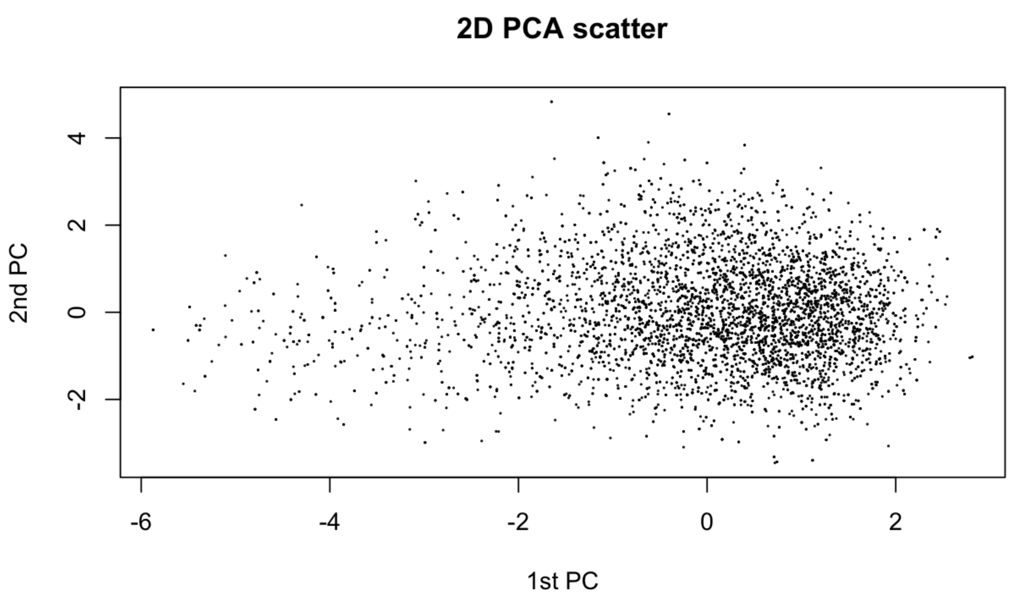




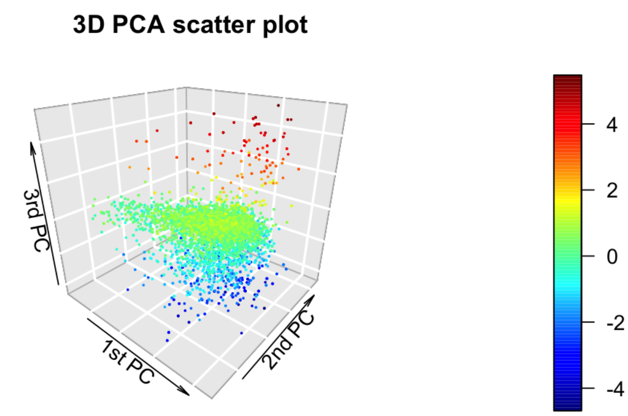




Appendix H: 2-dimensional PCA scatterplot



Appendix I: 3-dimensional PCA scatterplot



Appendix J: Results of simple linear regression with music genre, age or hours.per.day as predictor and reported mental health as response, showing only those with p-value <5%

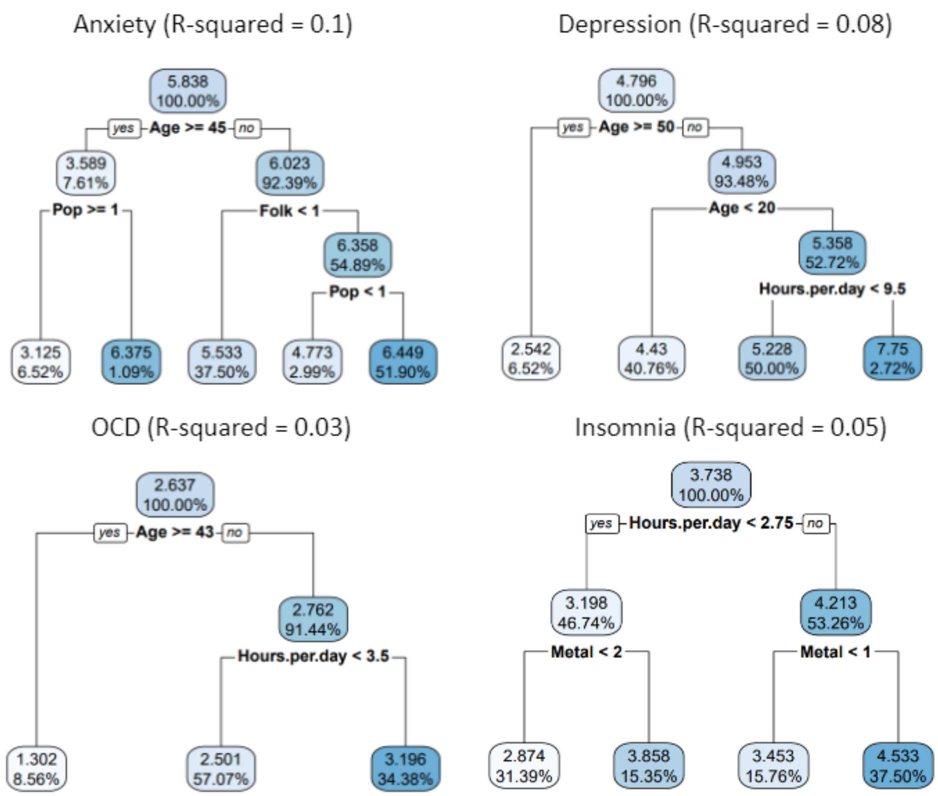
A data.frame: 25 × 5

predictor	response	coefs	p_values	adjusted_R_squared_percent
<chr>	<chr>	<chr>	<chr>	<chr>
Age	Composite	-0.026	0	2.08
Age	Depression	-0.031	0.001	1.343
Age	OCD	-0.031	0	1.559
Age	Anxiety	-0.041	0	2.989
Hours.per.day	Composite	0.096	0	1.839
Hours.per.day	Depression	0.111	0.003	1.087
Hours.per.day	OCD	0.111	0.001	1.275
Hours.per.day	Insomnia	0.145	0	1.878
Folk	Composite	0.196	0.01	0.769
EDM	Composite	0.221	0.002	1.106
Rock	Anxiety	0.228	0.022	0.575
EDM	Insomnia	0.241	0.026	0.535
Rock	Insomnia	0.251	0.023	0.569
Folk	Anxiety	0.255	0.012	0.714
Rock	Composite	0.26	0	1.547
EDM	Depression	0.262	0.014	0.689
Pop	Depression	0.267	0.026	0.541
EDM	OCD	0.269	0.007	0.848
Metal	Composite	0.27	0	2.048
Hip.Hop	Depression	0.29	0.007	0.844
Pop	Anxiety	0.312	0.005	0.954
Folk	Depression	0.338	0.002	1.136
Metal	Insomnia	0.437	0	2.449
Metal	Depression	0.471	0	2.985
Rock	Depression	0.564	0	3.58

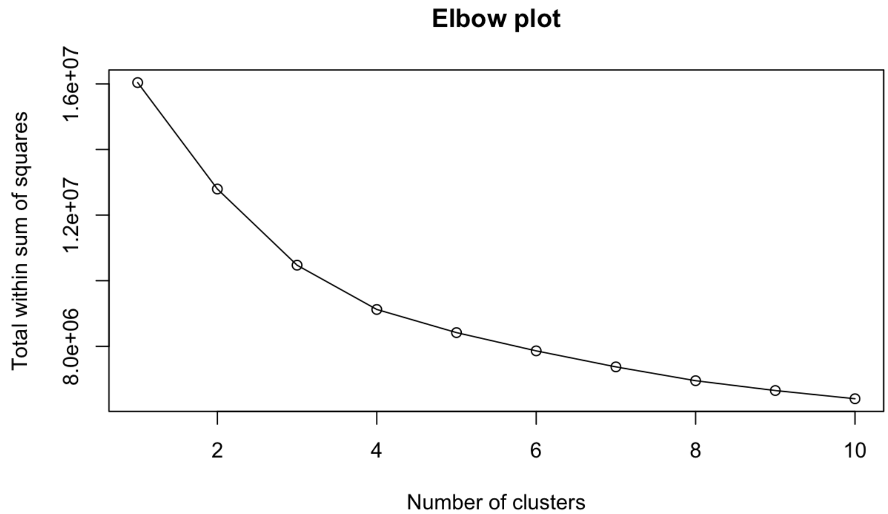
Appendix K: Decision tree variables table

Dependent variable	Independent variables selected
Depression	Age + Hours per day + Country + Folk
Anxiety	Age + Folk + Pop + Rock
OCD	Age + Hours per day + Country + EDM
Insomnia	Hours per day + Country + Latin + Metal

Appendix L: Decision trees for Anxiety, Depression, OCD, and Insomnia mental health scores after stepwise variable selection.



Appendix M: Kmeans elbow plot



Appendix N: Kmeans cluster plot of 6 clusters