Music Patterns as Mood Indicators

Music listening is a universal behavior with drastic emotional and cognitive effects. In this project, I'll investigate whether self-reporting listening patterns (genre preferences, diversity of musical exposure, daily listening time, and choice of streaming platform) convey predictive signals of mental-health risk across four specific domains: depression, anxiety, insomnia, and obsessive-compulsive disorder (OCD). By combining detailed survey data from 736 participants, my goal is to answer the following:

What listening patterns - in terms of genre, frequency, or streaming habits - signal elevated mental health risk?

Some key findings from our exploratory analysis include the following:

- **Genre Diversity**
  - Tends to increase with symptom severity (e.g. those reporting high insomnia explore around 0.6 more genres on average than low-insomnia respondents
- **Listening Duration:**
  - Shows a clear upward trend: higher symptom groups spend more hours per day with music (median hours climb from Low to Moderate to High in terms of severity of symptoms across all four conditions)
- **Genre-level associations:**
  - These are strongest for energetic styles such as Rock, Pop, Metal, and Rap. This is especially true in the depression and anxiety domains.
- **Streaming platform patterns:**
  - Spotify users are disproportionately represented among high-symptom groups, while Pandora and "no streaming" tend toward lower severity levels.

In the sections that follow, I will describe the actual dataset, cleaning the data, as well as the exploratory data analyses I conducted before turning to predictive modeling.

Data Description:

The survey dataset (mxmh_survey_results.csv) contains 736 complete responses and 31 post-cleaning variables. The data comes from the *2022 Music & Mental Health (MXMH)* survey.

Demographics and Usage:

- Age (years, float)
- Primary streaming service (Spotify, YouTube Music, Pandora, Apple Music, Other, None)

- Hours per day listening to music (float)

Musical Preferences & Behaviors:

- Favorite genre, Exploratory (yes/no), Foreign languages (yes/no)
- BPM (beats per minute; around 14% missing)
- Music effects (self-reported impact on mood; 8 missing)

Genre-Frequency Rating:

- Sixteen genres (e.g., Classical, Country, EDM, Folk, Gospel, Hip hop, Jazz, K-pop, Latin, Lofi, Metal, Pop, R&B, Rock, Video Game music) rated on a 4-point scale ("Never" to "Very frequently")

Mental-Health Scores

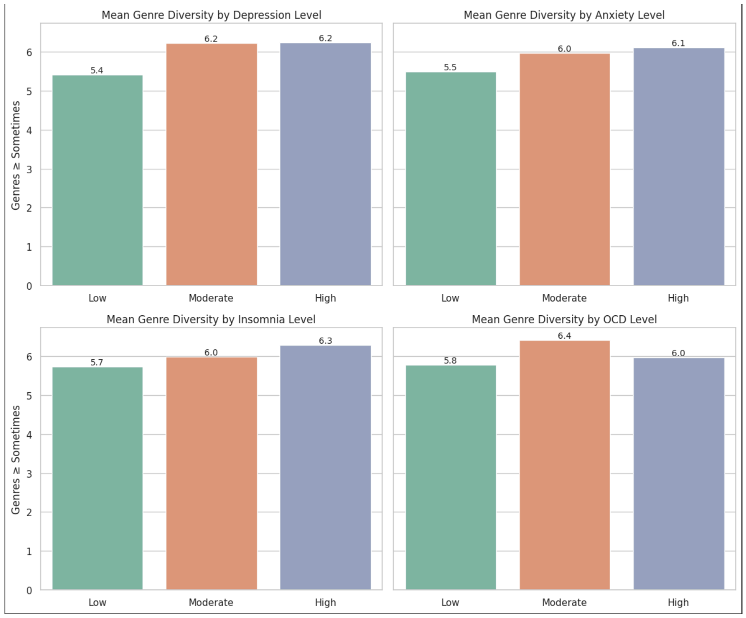
- Four continuous scales (0-10) self reporting Anxiety, Depression, Insomnia, and OCD severity.

After dropping the timestamp and permission fields, I addressed missingness, transformed genre frequencies into numeric values, and derived summary features to support further advanced modeling. I also engineered the following:

- 3-level severity labels (Low for scores less than or equal to 3, Moderate for 4-6, and High for score greater than or equal to 7) for each symptom
- Genre Diversity: Count of genres listened to "Sometimes" or more (0-16).
- Average Genre Frequency: Mean of the 16 mapped genre values (0-3).

Models & Methods: Exploratory Data Analysis

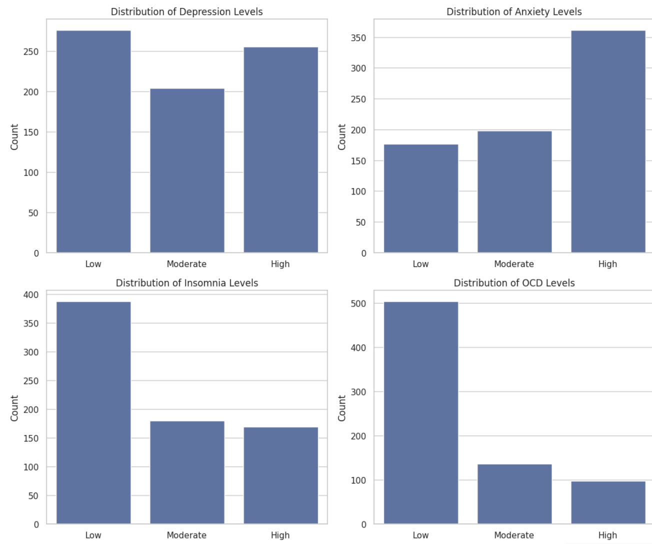
Mean Genre Diversity by Severity



1) **Depression:** Those with Moderate or High depression listen to about 6.2 genres "Sometimes"+, versus 5.4 for Low. This is nearly a full genre jump.
2) **Anxiety:** A clear upward trend: High-anxiety listeners hit around 6.1 genres vs 5.5 genres for Low
3) **Insomnia:** There is a steadier climb here. The High-insomnia group hits around 6.3 dangerous vs 5.7 genres for Low, showcasing broader exploration as sleep worsens.
4) **OCD:** This symptom peaks at around 6.4 genres, but High-OCD dips to around 6, possibly conveying that extreme OCD coincides with narrower tastes.

The takeaway here is that individuals with higher depressive, anxious, or insomnia symptoms explore a broader set of genres, sampling roughly one or more genre "Sometimes"+, while those with extreme OCD show a narrowing in tastes.

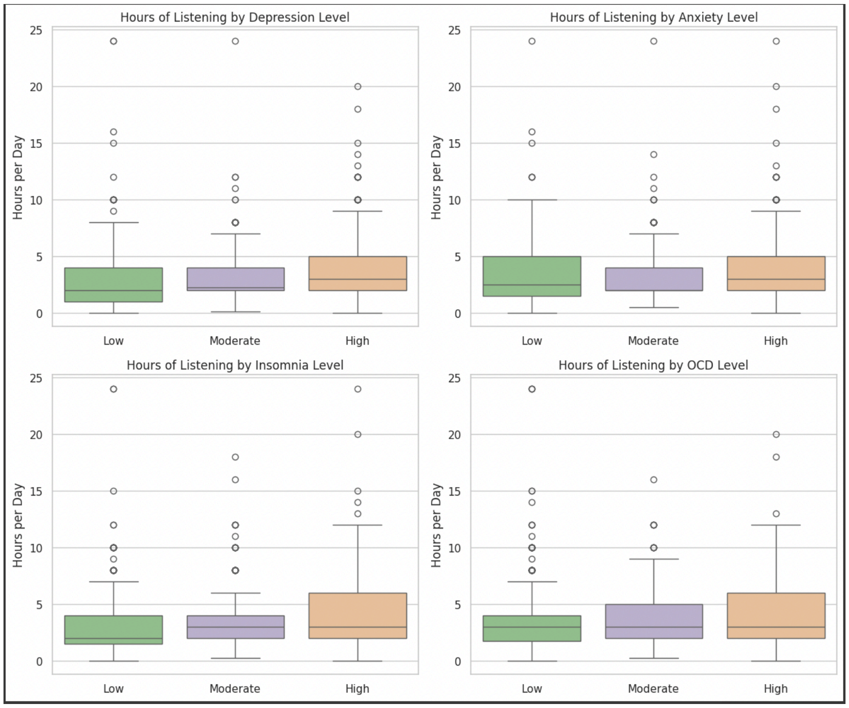
Distribution of Mental-Health Severity Levels



1) **Depression:** Roughly equal Low (275) and High (255), with a smaller Moderate (205) segment.
2) **Anxiety:** Strong skew toward High (360), Moderate (200), and few Low (175). High anxiety is most prevalent.
3) **Insomnia:** Majority report Low (385), fewer Moderate (180) and High (170). Sleep issues are less common overall.
4) **OCD:** Over two-thirds Low (505), 135 Moderate, 100 High. OCD symptoms are relatively rare here.

Anxiety is overwhelmingly skewed toward High, depression splits roughly evenly between Low and High, insomnia concentrates in Low, and OCD is predominantly Low, highlighting varying base‑rates of each symptom in our sample.

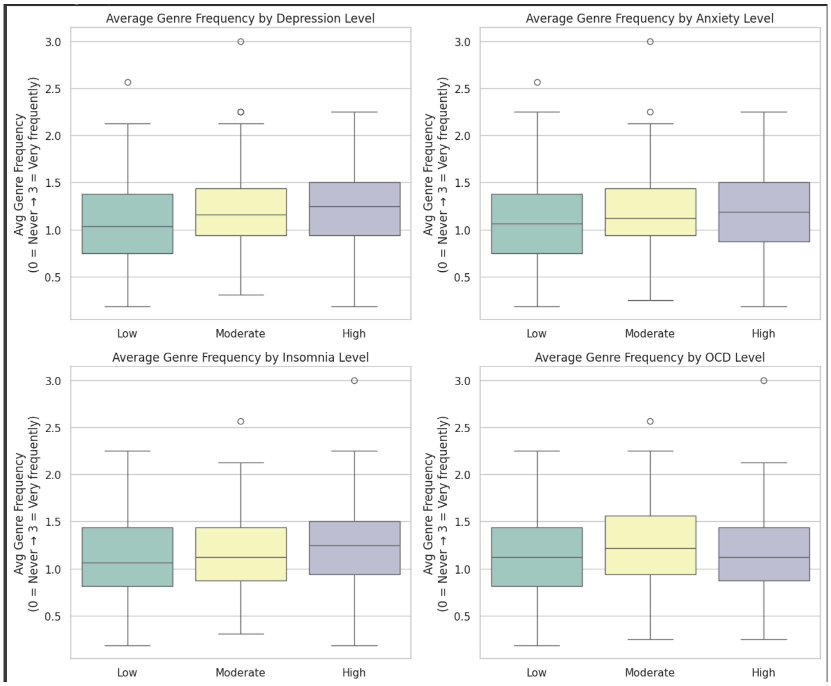
Hours of Listening by Severity



- In General**,** median daily listening steadily rises from Low→Moderate→High severity across all four conditions.
- **Depression & Anxiety:** High-severity groups show expanded IQR and extreme outliers (up to 24 hrs/day), indicating some use music heavily.
- **Insomnia:** A longer upper whisker for High suggests music as a coping or sleep-aid tool.
- **OCD:** High-OCD outliers reach 20 hrs/day, revealing a small group relying intensely on music.

High-severity Depression and Anxiety groups showing heavy outliers (up to 24 hrs/day) and insomnia sufferers extending their upper whiskers of their boxplots. This showcases music as a coping mechanism.

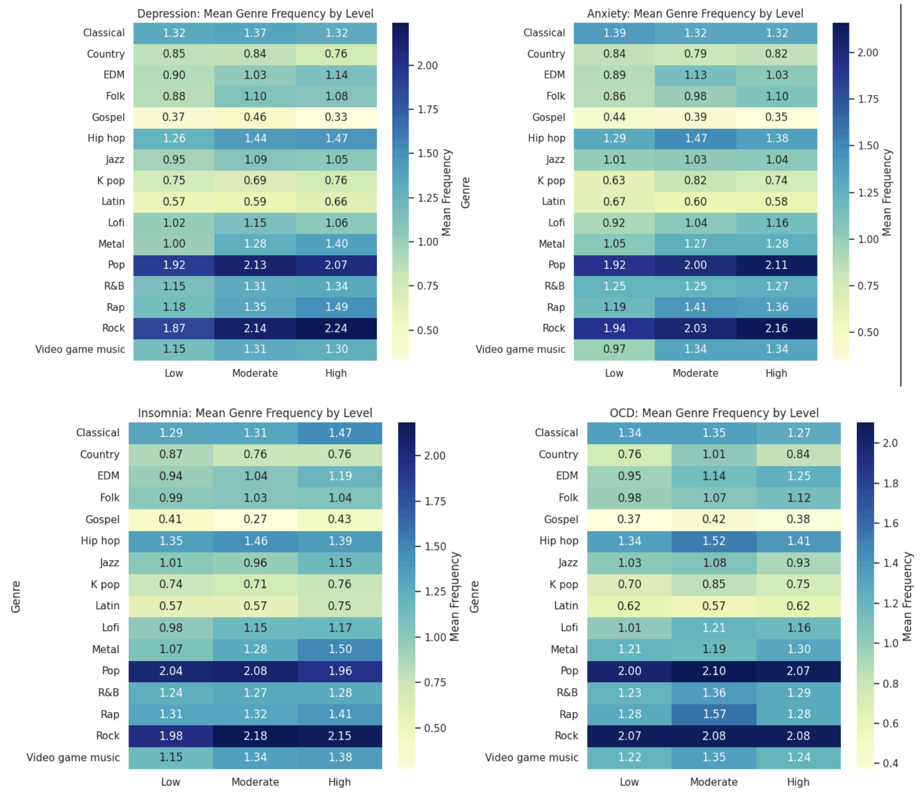
Average Genre Frequency by Severity



- **Depression & Anxiety:** High-severity IQRs shift upward. There are outliers at the 3.0 max, conveying deep engagement across all genres.
- **Insomnia:** Moderate uptick in median, but there's less spread than mood symptoms. Genre breadth grows at a steady rate.
- **OCD:** There are nearly overlapping boxes. Genre frequency remains stable regardless of OCD severity.

Those with severe depression or anxiety not only sample more genres but listen to them more frequently (IQR shifts upward, outliers at the 3.0 max), whereas insomnia shows a steady rise and OCD remains flat.

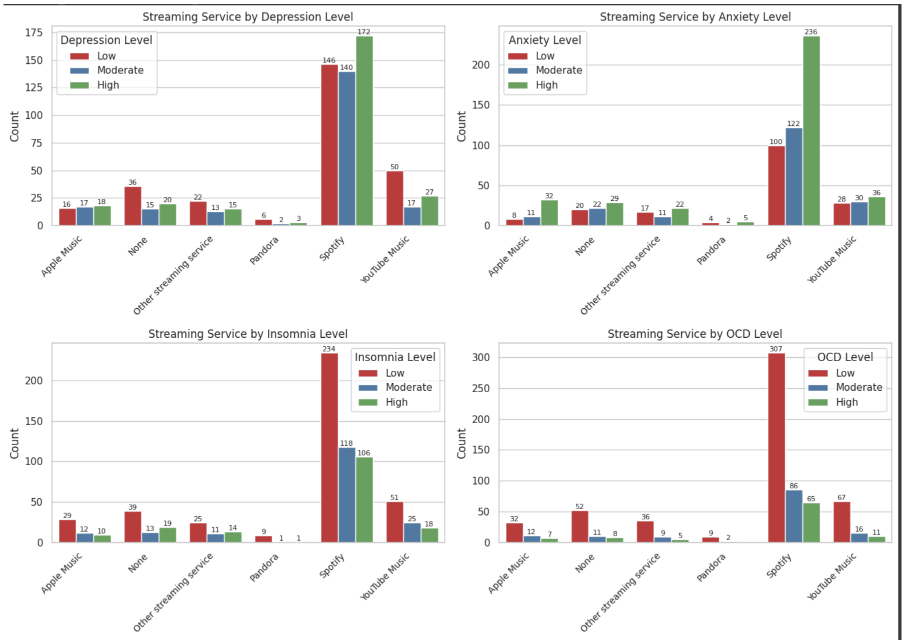
Mean Genre Frequency Heatmaps



- **Depression:** Biggest jumps are in Rock, Metal, and Pop. Softer genres such as Gospel dip.
- **Anxiety:** Just like depression, Rock and Pop rise most. EDM and Lofi also climb. Therefore, both high-energy and chill-electronic styles correlate here.
- **Insomnia:** Metal and Rock increases point to intense music for wakefulness. Classical/Lofi climb upward for winding down as well.
- **OCD:** There are little patterns here. Only small Rap/Pop increases, indicating little genre-level signal.

High-depression and High-anxiety respondents exhibit increases in energetic genres (Rock, Metal, Pop, EDM, Lofi), insomnia aligns with both intense (Metal, Rock) and mellow (Classical, Lofi) styles, and OCD shows minimal shifts across genres.

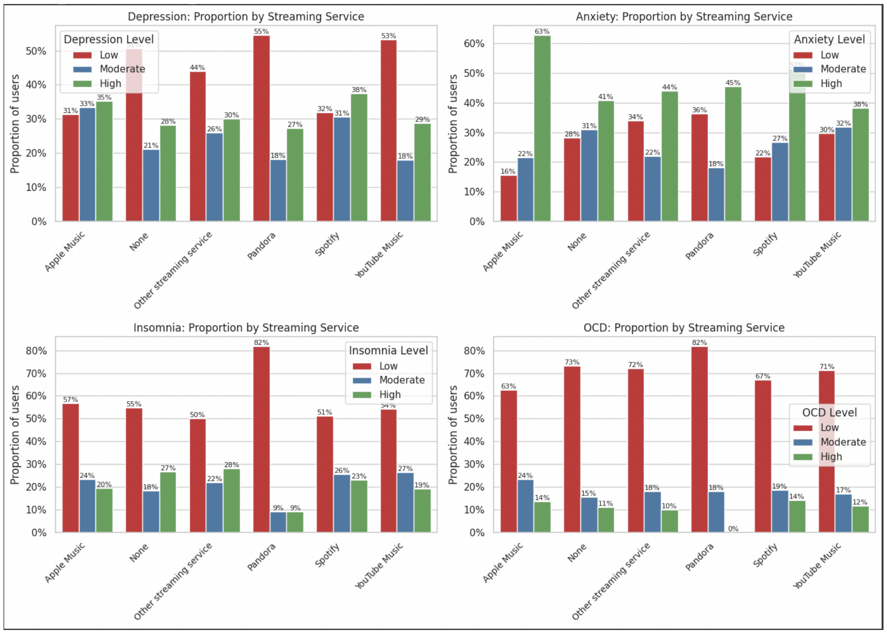
Streaming Service Usage (Counts)



- Spotify dominates across every group, with counts rising from Low→High for Depression & Anxiety.
- YouTube Music shows a smaller but steady increase with severity.
- Pandora and Other services have very low High-severity counts. They typically skew toward lower symptom groups.

Funny enough, Spotify's user base grows with symptom severity, especially for Depression and Anxiety. YouTube Music rises slowly, and Pandora/Other services cluster among lower-symptom individuals. Therefore, (based on the sample alone) you are more likely to use Spotify if you have severe Depression and Anxiety.

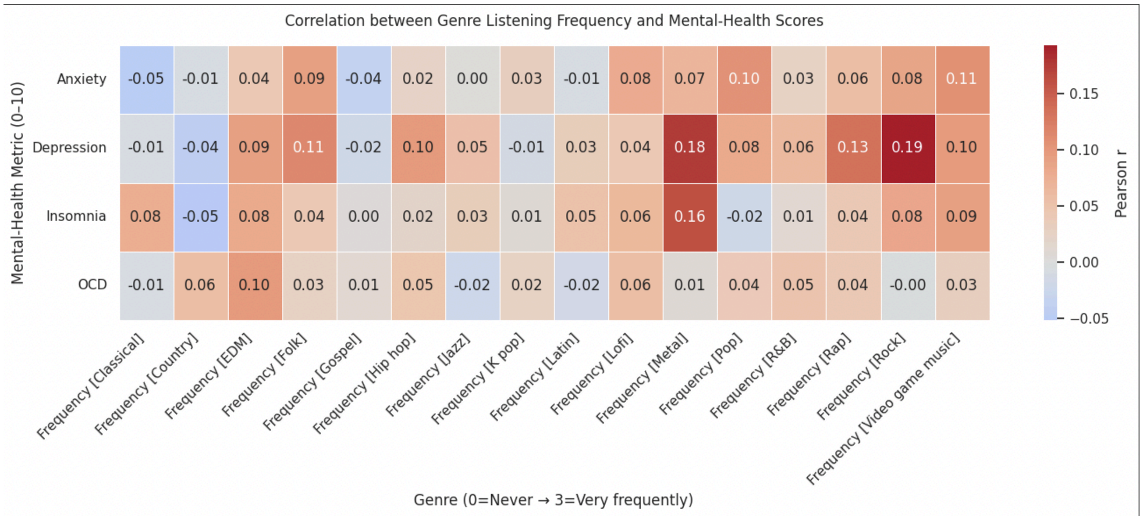
Streaming Service Usage (Proportions)



- **Spotify:** There's a higher share of High-severity users (up to 53–63%), especially in Anxiety and Depression. This makes sense when looking at the preceding Counts plot.
- **Pandora & Other:** Concentrated among Low-severity (70–82%). There's almost none in High for OCD.
- **Apple Music & YouTube:** These are more balanced but YouTube edges up in the High groups for the most part.

High-severity users disproportionately favor Spotify (over 50% of its audience), whereas Pandora and niche services skew heavily toward Low-severity groups. YouTube shows a slight High-severity uptick.

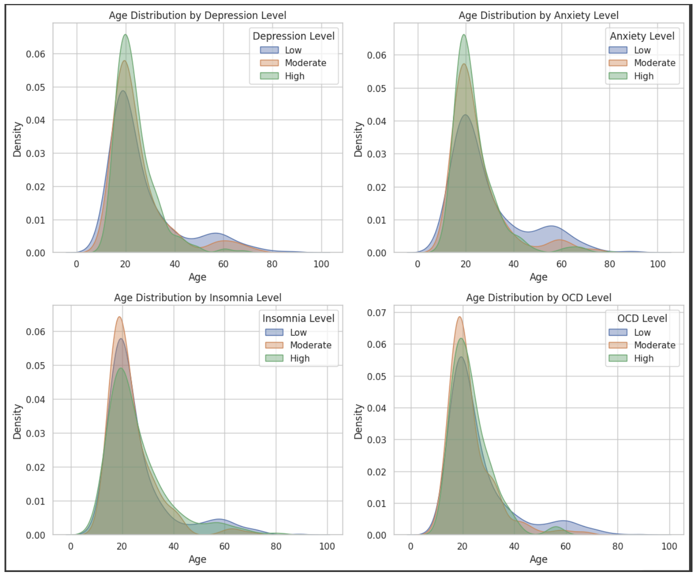
Genre-Mental-Health Correlation



- **Depression:** Strongest positive correlations with Rock (r = 0.19) and Metal (r = 0.18). The correlations are moderate with Rap, EDM.
- **Anxiety:** Highest with Rock (r = 0.13), Pop (r = 0.10), EDM (r = 0.09), and Video-game music (r = 0.11).
- **Insomnia:** Peaks at Metal (r = 0.16) and Rock (r = 0.08), plus mild Lofi (r = 0.06).
- **OCD:** The correlations are all near zero. There is no strong genre correlation here.

Rock and Metal have the strongest positive correlations with Depression (r = 0.19, r = 0.18) and Anxiety (r = 0.13, r = 0.10). EDM and Rap show weak but possible links. Insomnia aligns most with Metal (r = 0.16), while OCD correlations remain near zero.

Age Distribution by Severity



- **Depression & Anxiety:** High-severity curves are narrowly peaked in the early 20s, while Low-severity tails extend into older ages (40s–60s).
- **Insomnia & OCD:** All three curves largely overlap. There are slight bulges in younger ages for High severity, but age isn't a strong differentiator.

High-severity Depression and Anxiety are tightly centered in the early 20s, while Low-severity curves extend into older ages; insomnia and OCD age distributions overlap broadly, indicating age is less predictive for those symptoms.

EDA Summary

Our exploration reveals clear listening‑behavior signatures of mental‑health risk. Heavier and longer engagement with music correlates with higher symptom severity while energetic genres (Rock, Metal, EDM) and broad genre diversity mark depression and anxiety, with insomnia showing a hybrid pattern of intense and soothing styles. The streaming choice, especially Spotify usage, also rises with risk. Age and platform effects offer additional, though subtler, signals.

Models & Methods: Predictive Modeling

I trained and compared five classifiers (Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors, and a Neural Net) across four binary targets (High vs Not-High for Depression, Anxiety, Insomnia, and OCD). I used ROC AUC as the primary tuning metric since it evaluates ranking ability across all thresholds and is robust to class imbalance. I also report accuracy, precision, and recall at the default threshold of 0.5. Below are each model's setup, visuals (if available), and takeaways which tell the story of how, especially how well, listening behaviors map onto mental-health risk.

Logistic Regression

Why?: The coefficients directly show each listening feature's direction and magnitude of risk. Additionally, this model sets a baseline against which to compare more complex models.

For the Pipeline:

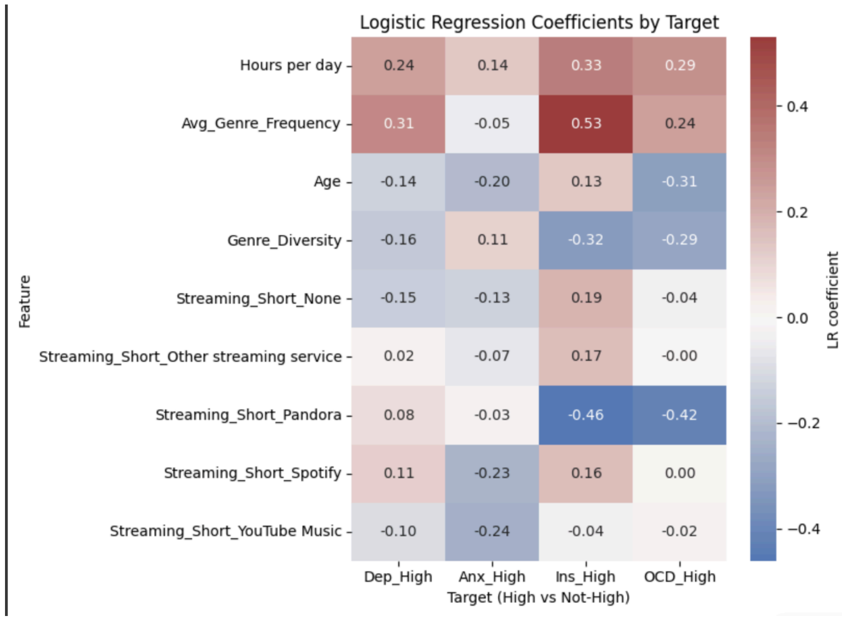
- I used a StandardScaler on numeric features (Hours/day, Avg_Genre-Frequency, Genre_Diversity, Age)
- LogisticRegression (class_weight = 'balanced', max_iter = 100)
- One-hot encoded the streaming platforms

The Performance was the following:

- Accuracy: 0.52
- Precision: 0.36
- Recall: 0.47
- ROC AUC: 0.51

This model defaults to "not high-depression", catching fewer than half of true positives and barely beating random chance with an AUC of around 0.50.

Before moving on, I wanted to know what moves the coefficients so I created the following visual:



- **Listening Volume:** Hours per day as well as Avg_Genre_Frequency both increase odds of "High" across conditions. People who listen more, and more often across genres, tend to report more severe symptoms
- **Genre Breadth:** Genre Diversity is positive for Depression but negative for OCD which means that depressed listeners sample more genres, where as those with OCD may have a narrower taste.
- **Age:** This metric is negatively associated with Depression/Anxiety in this sample but this trend flips for Insomnia.
- **Streaming Platforms:** The platform's usage increases Depression/Insomnia odds, while Pandora is protective for Insomnia/OCD

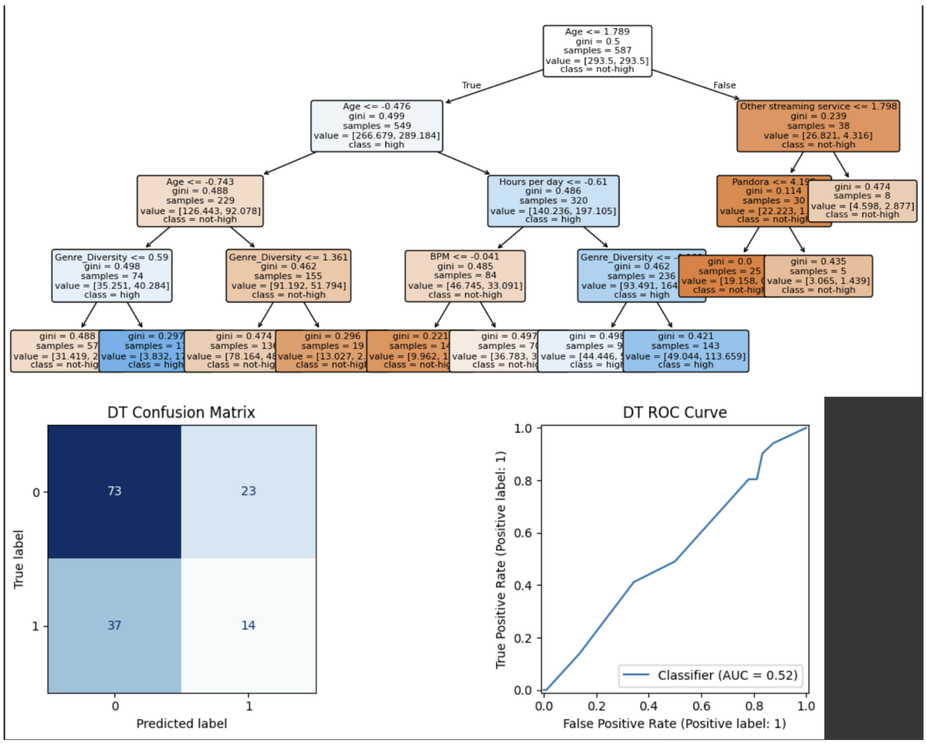
Therefore, a simple linear boundary such as Logistic Regression can't reliably screen "High" vs "Not-High" for most conditions. It can't screen here because risk does not grow proportionally in feature space. Listening behavior correlates weakly with symptoms and interact non-linearly, so this model struggles to distinguish the classes.

Decision Tree

Why?: I chose DT because this is rule-based, meaning it's intuitive in the sense it follows "if" → "then" logic. Additionally, this model is nonlinear so it may better capture simple interactions between listening metrics

For the Pipeline:

- I used a StandardScaler
- DecisionTreeClassifier (class_weight = 'balanced')
- Tune max_depth via a 5-fold CV on ROC AUC, then grid-search min_samples_leaf



- The tree captures a few splits (such as Age, Genre_Diversity), but generalizes no better than chance. Many of the rules overfit CV folds and vanish on new data.

With a decision tree, there are no simple "if-then" listening rules that are robust enough to detect high-severity across these features. The model overfits those few rules in CV but then generalizes no better than chance on the data. There simply isn't a small set of hard thresholds on listening features that can reliably flag high-risk individuals.

Random Forest

Why?: This model permits me to reduce overfitting (a problem from the Decision Tree model) and boost weaker signals. In other words, it can capture weak, distributed signals across many features.

For the Pipeline:

- I used 100 trees, class_weight = ‘balanced’
- Baseline vs Tuned (grid_search n_estimators, max_depth, min_samples_leaf by ROC AUC)

```
Baseline Random Forest:
  Accuracy : 0.59
  Precision: 0.32
  Recall   : 0.16
  ROC AUC  : 0.50

Top features:
 BPM                         0.267657
Age                         0.259579
Genre_Diversity             0.174603
Hours per day               0.174552
MusicEffects_Code           0.050742
Spotify                     0.025315
None                        0.015761
YouTube Music               0.015045
Other streaming service     0.011918
Pandora                     0.004827
dtype: float64
```

```
Tuned Random Forest — Test Metrics
  Accuracy : 0.61
  Precision: 0.43
  Recall   : 0.45
  ROC AUC  : 0.55

Top features (tuned RF):
 Age                         0.297056
Hours per day               0.188618
Genre_Diversity             0.158164
BPM                         0.140767
MusicEffects_Code           0.084533
Spotify                     0.049461
None                        0.039145
YouTube Music               0.027364
Other streaming service     0.011513
Pandora                     0.003380
dtype: float64
```

- I went from a Baseline AUC of 0.50 to a Tuned AUC of 0.55
- The top predictors were Age, Hours/day, Genre_Diversity, Avg_Genre_Frequency, BPM
- The best test metrics yielded the following results:
    - Accuracy: 0.61
    - Precision: 0.43
    - Recall: 0.45
    - ROC AUC: 0.55

Random Forest uncovers weak but consistent predictors: younger, heavier listeners who sample many genres, especially at higher tempos, are more likely to report high symptoms. This is the most successful model for Depression, lifting screening ability above random. This works the best for Depression because the higher complexity of a random forest lets it leverage subtle correlations that Logistical Regression and Decision Trees Miss. Nevertheless, even Random Forest can only edge above random here. This indicates that listening features along provide limited screening power without deeper contextual data.

K-Nearest Neighbor

Why?: This model predicts based on “neighboring” listeners in feature space. There are no assumptions about linearity or axis-aligned splits which can help the performance potentially.

For the Pipeline:

- Medium Imputer → Scaler → KNeighborsClassifier
- Grid-search over neighbors [3-11] and weights [uniform, distance]

Observations:

- **Depression:** AUC is around 0.55 but recall is around 18%, missing 82% of cases
- **Anxiety:** AUC = 0.53. This symptom also has balanced precision/recall of around 50%, catching half of high-anxiety cases
- **Insomnia & OCD:** The model fails to predict any “high” cases, defaulting to the majority class

KNN’s reliance on local proximity in the mixed feature space is insufficient. Only anxiety displays a weak, yet consistent pattern. There’s severe class imbalance in this dataset which prevents the model from performing well for rarer conditions. Few “High’” labels means most neighbors are “Not-high”, so KNN almost never votes positive.

Neural Net (Multi-Layer Perceptron)

Why?: To capture potential nonlinear feature interactions without super complex feature engineering.

For the Pipeline:

- Imputer → Scaler → MLPClassifier (early_stopping = True)
- Tuning: hidden layer sizes, α (regularization), learning rate by CV ROC AUC

Observations:

- CV AUC: 0.54 → Test AUC: 0.59
- Threshold of 0.3 → Recall = 95%, Precision = 51%

From the metrics above, Depression, Insomnia, and OCD have an AUC that hovers anywhere from 0.51 to 0.58. The default threshold rarely predicts positives or floods with false alarms at low thresholds.

The Neural Net only meaningfully outperforms random for Anxiety, and allows me to tune for very high sensitivity. The symptoms associated with Anxiety map more consistently onto listening intensity and genre breadth. This may be the case because anxious individuals use music in more predictable coping patterns (like cycling between energetic and chill tracks).

Depression, Insomnia, and OCD relationships to music are more heterogeneous, requiring better features beyond a static survey snapshot.

Conclusion

I set out to answer the question:

What listening patterns - in terms of genre, frequency, or streaming habits - signal elevated mental health risk?

Across both exploratory analysis and predictive models, a consistent story emerged:

- **Listening Intensity & Genre Breadth**
  - Hours per day and genre engagement (how often someone listens across multiple genres) rose steadily with Depression, Anxiety, and Insomnia severity.
  - Those with more severe symptoms tune into one to two additional genres "sometimes or more," and their median daily listening time climbs by 1-2 hours versus low-severity respondents.
- **Genre Preferences**
  - Energetic styles like Rock, Pop, Metal, EDM and to a lesser extent Lofi show the strongest positive correlations (Pearson r = 0.10–0.19) with Depression and Anxiety.
  - Insomnia aligns with both intense (Metal, Rock) and soothing (Classical, Lofi) styles, showcasing mixed coping strategies.
  - OCD exhibits minimal genre-level shifts, indicating that obsessive-compulsive traits do not map cleanly onto broad genre habits.
- **Streaming Platforms**
  - Spotify users are disproportionately represented in high-severity groups (over 50–60% of its audience), whereas Pandora and "None" skew toward lower symptom levels.
  - However, platform choice is a weaker predictor than individual listening behaviors.
- **Modeling Takeaways**
  - ***Logistic Regression and Decision Trees*** performed at or near chance (AUC = 0.50–0.52), unable to develop a linear or rule-based boundary in the feature space.
  - ***The Random Forest*** lifted Depression screening to AUC = 0.55, identifying Age, Hours/day, Genre_Diversity, Avg_Genre_Frequency, and BPM as top predictors.
  - ***KNN*** struggled under class imbalance in the dataset, never capturing rare "High" OCD or Insomnia cases and only mildly helping Anxiety (AUC = 0.53).
  - ***The Neural Net (MLP)*** achieved its best results on Anxiety (AUC = 0.59, Recall = 95% at a 0.3 threshold), offering a tunable trade-off for high-sensitivity screening. For Depression, Insomnia, and OCD it again hovered around chance.

While more time-consuming, broader, and faster listening correlates with higher symptom severity, the static survey features of the dataset alone yield only weak screening performance (peak AUC = 0.59). To build a reliable digital "music-based" screening tool, I'd require more richer, dynamic aspects of listening behavior.

Next Steps

Richer dataset:

- Time-stamped listening events, such as session lengths, skipping rates, and part-of-day usage which can tell us more about coping rituals (late-night playlists for insomnia for example).
- Genre switching patterns can allow me to flag mood-regulation attempts such as transitioning between upbeat and calm tracks.
- Incorporating Audio and Lyrical Features in the dataset would permit me to extract tempo, loudness, and lyric sentiment to go beyond broad genre labels. This would be interesting to handle, given it's a different data type.
- Richer demographics and context would make predictive modeling more efficient and accurate. Occupation, existing diagnoses by doctors, and survey motivations would allow me to personalize risk estimates and differentiate healthy vs problematic listening.
- Augmenting "High" cases via targeted recruitment or data synthesis to stabilize KNN and tree-based models would provide better results. Through all of this, ensuring user consent data anonymization would ensure any screening tool respects privacy and avoids stigmatization.
- Having a listening source (mobile vs desktop vs smart speaker) and location can add behavioral context.

Analytical and modeling avenues:

- Multi-task learning: train a single model to predict all four symptoms simultaneously, allowing it to iterate and improve upon signal extraction for rarer outcomes (such as in Insomnia or OCD).
- Pilot "music prescriptions" for small test cohorts to test if modifying listening behavior has measurable mental-health benefits.
- Developing a prototype or dashboard for real-time monitoring where models are automatically ran based on new data that's embedded into the dataset would be interesting.

- Applying Natural Language Processing to determine if people gravitate toward particular lyrical themes when distressed, especially if the dataset contains top songs that individuals have listened to.
- Applying Transformers on listening sequences to detect mood-regulation routines (from "energy-boost" to "calm-down")

The above analysis conveys that while everyday listening behaviors (how much people listen, how many genres they explore, and which platforms they favor) carry a lot of signals of mental-health risk. The static survey in this analysis provides snapshots and hints but lacks the fidelity for reliable screening. **By enriching the data** based on the above suggestions, along with **adding advanced modeling techniques**, it's possible to move closer to a personalized, music-informed monitoring system.