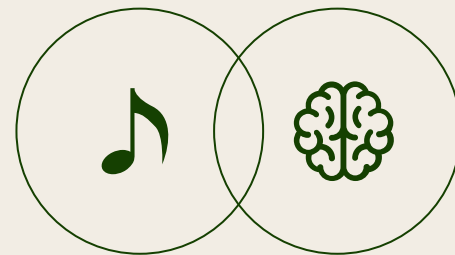


# Music and Mental Health

---

Keara Toop



# Table of Contents

01

---

Introduction

02

---

Insights

03

---

Components of  
the Dataset

04

---

Machine Learning

01

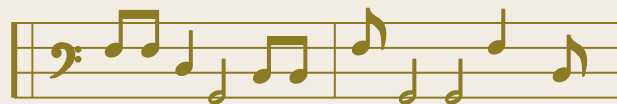
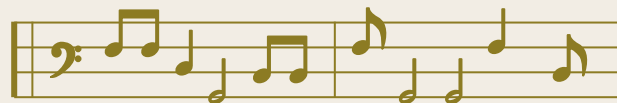
---

# Introduction



# Project Goals:

1. Deepen understanding of how music can affect mental health
2. Enhance customer experience through curated public playlists
3. Advertising to increase revenue through subscription sales



# The Overall Value

The stakeholder will gain a better understanding of their customer base and how users with mental health conditions benefit from their service. Additionally, the stakeholder will receive access to data to help create playlists curated for specific mental health conditions. From there, the stakeholder can market these playlists for Mental Health Awareness Month and possibly gain revenue from increased subscription sales.



# The Process

01

---

Understand the  
data and gain  
initial insights

02

---

Perform k-means  
clustering; a machine  
learning technique

03

---

Analyze the results  
and provide  
recommendations  
to Spotify

02

# Insights



# Understanding the Results

The purpose of this project is to give you, Spotify, a selection of genres that people who experience certain mental health illnesses listen to frequently.

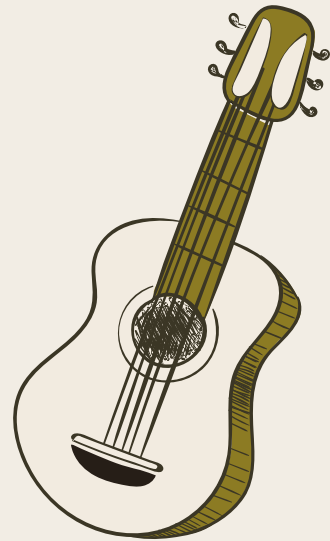
From this selection, Spotify will be able to make playlists specific to the listening habits of people with anxiety, depression, insomnia, and ocd.






# My Recommendations

The following four slides outline the recommended genres for each mental health playlist based on the results of a k-means clustering algorithm and analysis on some of the top clusters.

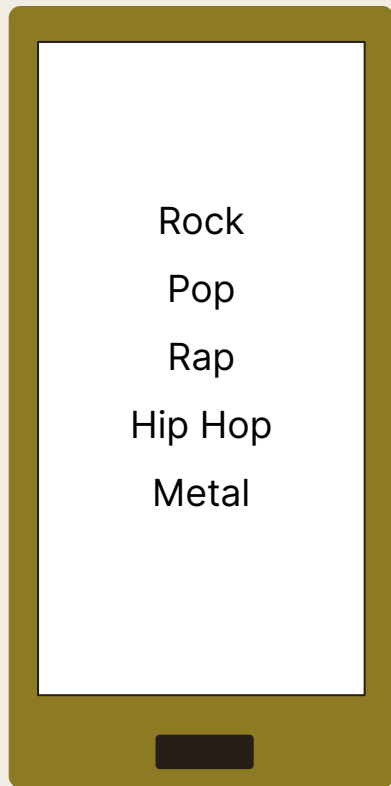




Hip Hop  
Rap  
Rock  
Pop

# Top Genres for Anxiety






# Top Genres for Depression





# Top Genres for Insomnia





Hip Hop

Rock

Pop

Rap

R&B

Latin

# Top Genres for OCD





# Lowest Ranked Genres To Avoid Including in the Playlists

## Anxiety

K-pop, Gospel, Classical,  
Video Game Music, Latin

## Depression

Gospel, K-pop, Classical,  
R&B, Latin, EDM

## Insomnia

Classical, Gospel, K-pop,  
EDM, Video Game Music

## OCD

K-pop, Gospel, Classical,  
Video Game Music

# Two Ways to Use This Information

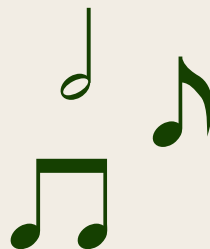
## Public Playlists

For each mental illness, create a public playlist using the recommended genres that any Spotify user can access. These playlists will appear the same for each user.



## Personalized Mixes

Create mental illness 'mixes' using songs with the recommended genres that are personalized to each Premium user based on their listening habits.



# Competition

- From the survey, Spotify is the most used platform, but the other main competitors are YouTube Music and Apple Music.
- For marketing, Spotify should target the Baby Boomers and Generation X as they are the least represented by the data.
  - These generations tended to use Pandora.
  - This would be a marketing cost to Spotify, but their marketing costs have dropped recently, so there is space for an increase.
- More consumers means more revenue in subscription sales.



**Music**

pandora®



# Listeners at Work



Due to the majority of participants claiming to listen to music while they are working, Spotify should create mental health-focused playlists to be work-appropriate.

If consumers with mental illness experience the illness while at work, the content in the playlists should both aid their symptoms and should also be suitable for a work environment.



# Effects of Music on Mental Health

The majority of listeners said that music helped with the symptoms of their mental illness, but the second majority said that music had “No effect.”

With the creation of playlists that are aimed to directly aid specific illnesses, those who claimed music had no effect might find that the curated playlists will help them.

Some claimed that music “Worsened” their symptoms. Those consumers would not be in our target group for this project.



03

---

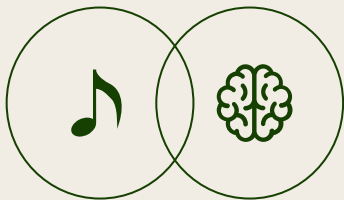
# Components of the Dataset



# The Dataset

## Mental Health and Music Survey

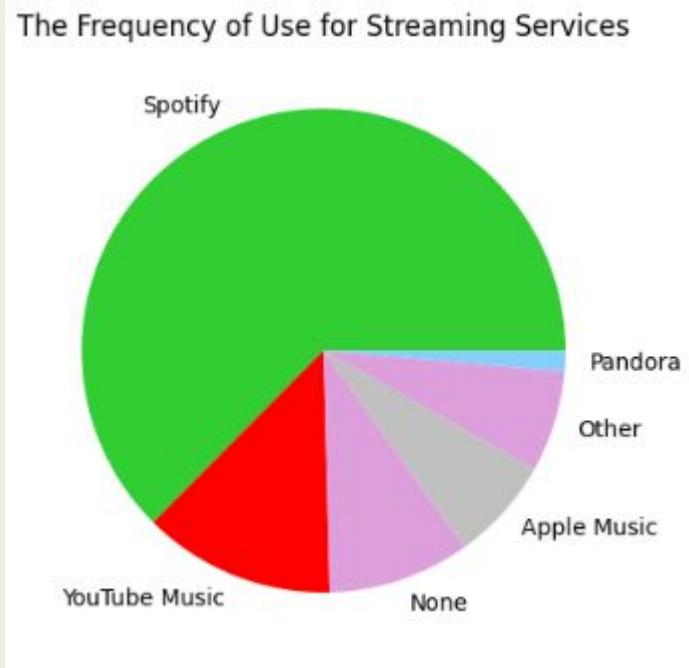
This will be the primary dataset to understand the relationship between music and mental health



## Columns of Interest

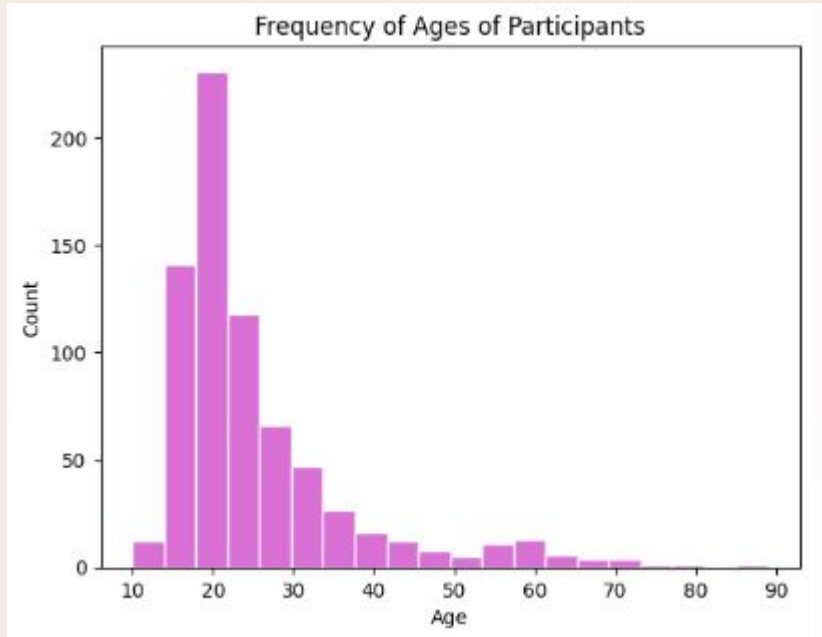
- Age
- Primary streaming service
- Hours per day
- While working
- Fav genre
- Frequency [all genres]
- Anxiety
- Depression
- Insomnia
- OCD
- Music effects

# Streaming Services



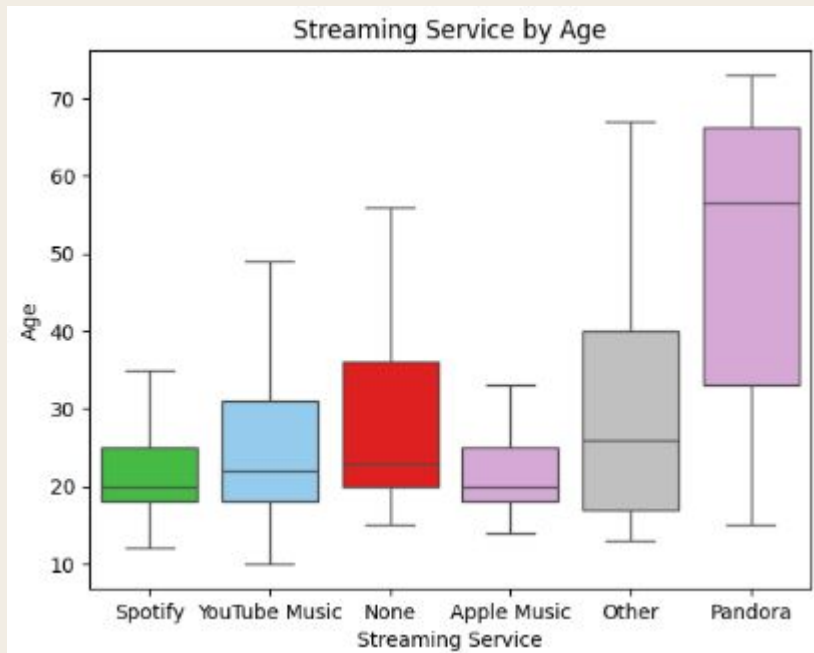
Among the participants in the survey, Spotify is the most frequently used streaming service. YouTube Music is the second most common, so it would be considered Spotify's biggest competitor based solely on this survey.

# Age Representation



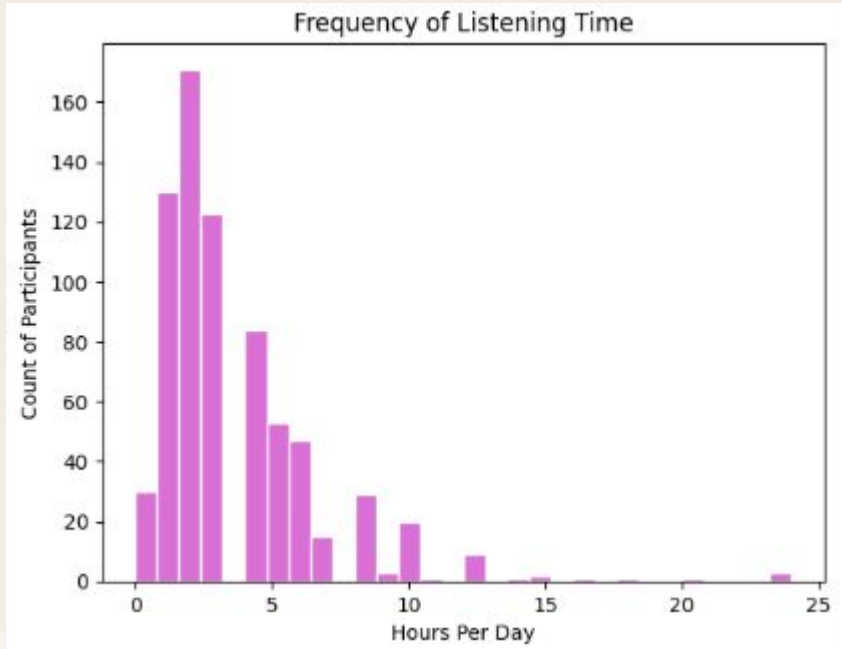
About 70% of participants are in the 15-25 year-old range. This means that there is not as much representation from the Baby Boomer and Generation X generations.

# Combining Streaming Services and Age Representation



- Starting with Spotify's data, the age group that mostly uses that platform is in the 18-26 years old age range.
- Apple Music has a very similar age demographic.
- Pandora has, by far, the largest age representation with a range from about 33-63 years old.

# Distribution of Hours of Time Spent Listening to Music Per Day



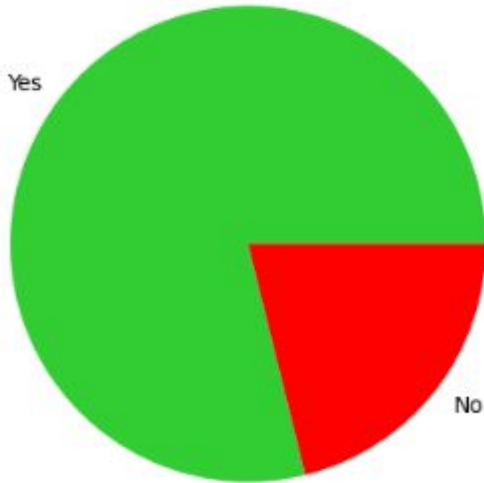
60% of users spend about 1 to 3 hours per day listening to music. There are also significantly many who listen from 4 to 5 hours per day; about 20%.

There are a few interesting outliers who claimed they listened to 24 hours of music per day.



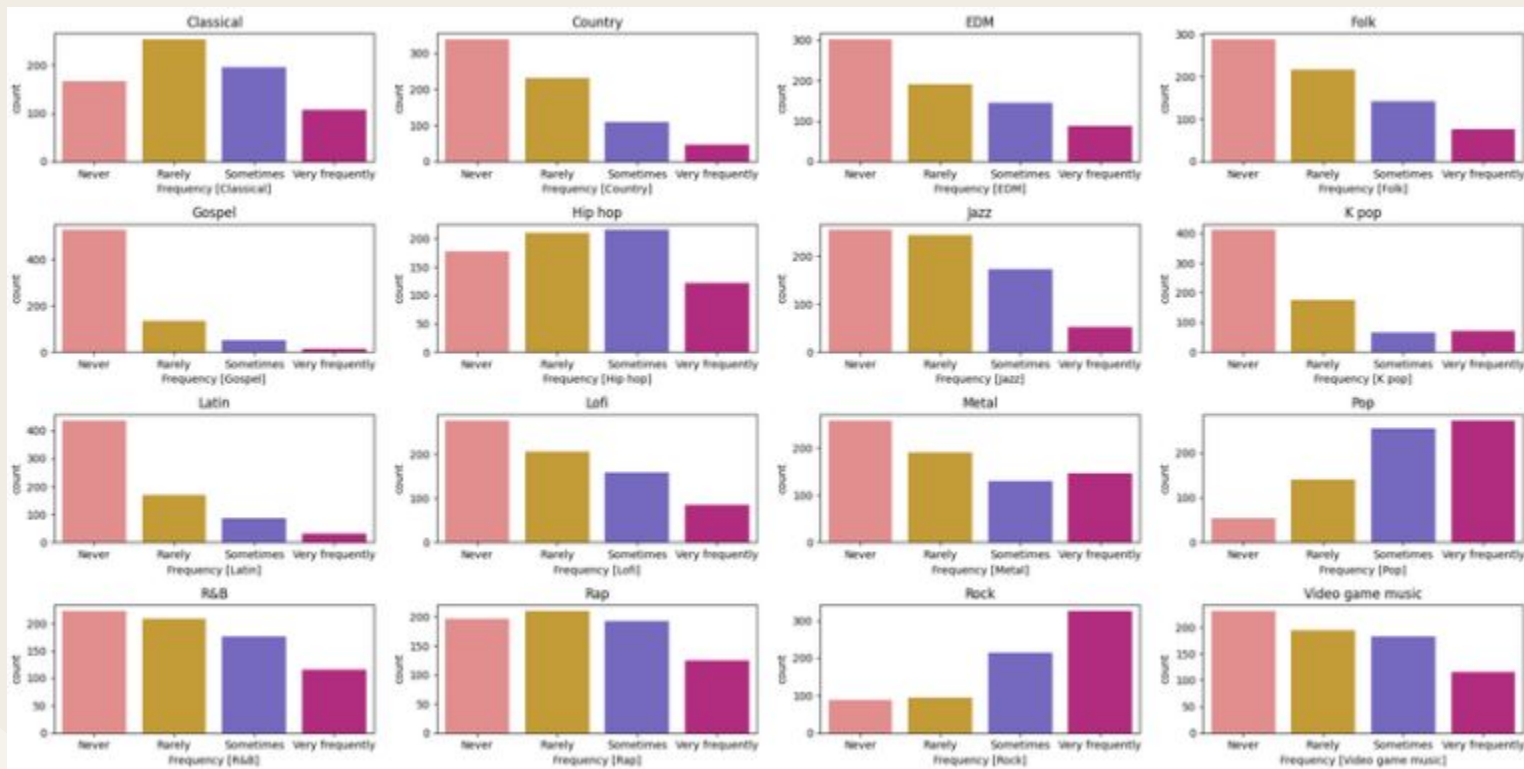
# Correlation of Participants Who Do or Do Not Listen To Music While Working

Do the Participants Listen to Music While at Work?



The majority of participants, about 80%, said they listen to music while they are working.

# Frequency of Listening to Different Genres - Part 1



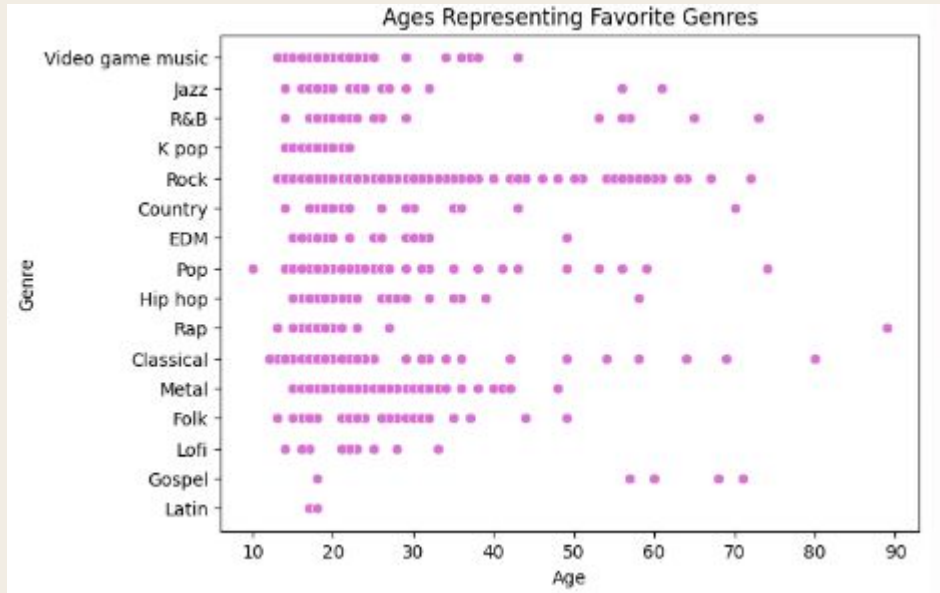
# Frequency of Listening to Different Genres - Part 2

- **The most frequently listened-to genres compared to the “Very frequently” ratings of other genres:** Pop, Rock
- **Also frequently listened to genres:** Hip Hop, Metal, R&B, Rap, Video Game Music
- **The least listened to genres:** Country, EDM, Folk, Gospel, Jazz, K pop, Latin, Lofi, Metal, R&B, Video Game Music

Analysis: each participant ranks every genre of music, so the most popular genres for some, might be the least popular for others. The most important information from this graph is the most frequently listened to genres because we can expect our playlists to mostly be made from those genres.



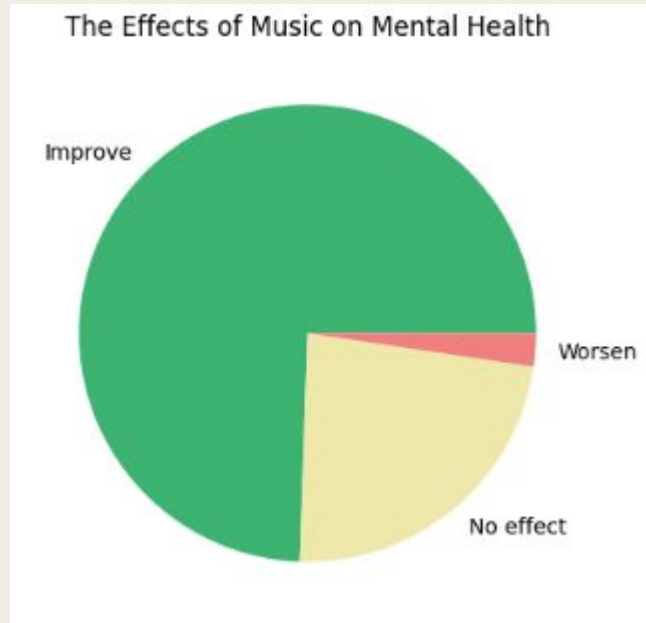
# Age and Favorite Genres



“Rock” is the most popular favorite genre, and there is also a wide age distribution.

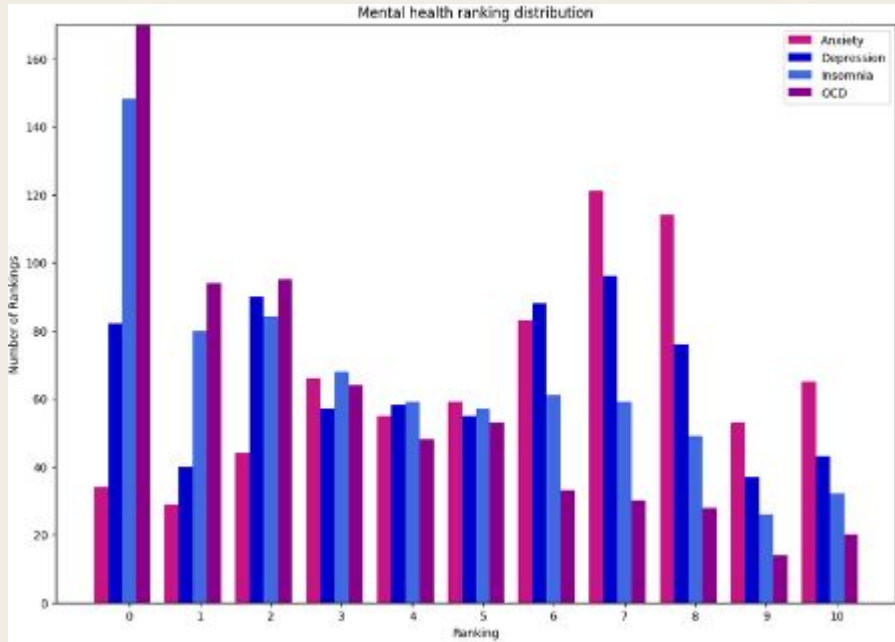
This genre has the most points in the Gen X and Baby Boomer generations, so targeting this genre could be a significant marketing factor to attract those age groups.

# Effects of Music on Mental Health



Of the participants, about 75% claimed that listening to music improves their mental health. Only a small percentage claimed it worsened their mental health.

# Distribution of Rankings for Illnesses



- **X-axis:** participants ranked how often or how extreme they experienced each illness on a scale from 0 to 10.
- **Y-axis:** the number of participants who chose that ranking.
- **Bars for each ranking:** there are separate columns for each illness.
- Anxiety has the most high ranking whereas OCD has the most low ranking.

# Limitation - Survey Data

The "**Music and Mental Health**" dataset is solely based on survey data. This means that the data is based on peoples' opinions about themselves. This lends space to biases that we cannot understand without knowing the participants.

There is also the possibility of internet trolls who claim to listen to 24 hours of music per day, for example.



# Limitation - Age Distribution

The age representation is the highest for 15-25 years old. This skews the averages for data to be aligned with the trends for these ages.

This will especially affect the genres (popularity changes based on generation) and the ratings for mental health illnesses. Mental health is a newer topic, so some older participants might think of it differently and therefore rate themselves differently. Ideally, this dataset would have a wider distribution of ages.

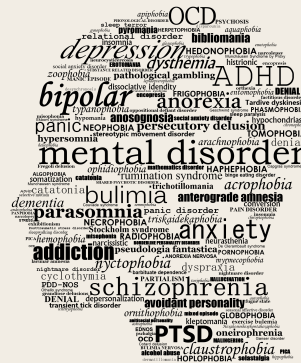
There is not as much representation of Generation X and the Baby Boomers as there is of millennials and Gen Z.





## Limitation - Mental Illnesses

There are only four mental illnesses included; anxiety, depression, insomnia, and OCD. While these are the most common, there are so many others. Because this survey includes only these four, we cannot examine how music affects other mental illnesses.



# Limitation - Music Genres

There is a possibility that their actual favorite genre was not an option on the survey.

Each participant can only say if music in general affected their mental health. They did not answer specifically for each genre, so there are some genres where this information is missing.



04

---

# Machine Learning



# K-means Clustering

- Unsupervised machine learning
- **Goal:** partition a dataset into a predetermined number of clusters, where each data point belongs to the cluster with the nearest mean (centroid)
- **Uses:** pattern recognition, data exploration



After creating the clusters, we can examine how the clusters of participants ranked their experience with each mental health illness.

Next, we can take the clusters that experience the illness the most and examine which genres, on average, are the cluster members listening to the most.

These results will give us the best estimates for genres to put in playlists because we have clustered like participants together.

# Three Steps to Prepare for K-means



01

---

Prepare  
the Data

02


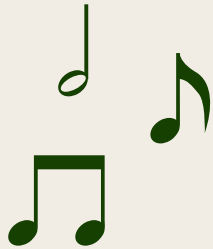
---

Principal  
Component  
Analysis

03

---

Cluster  
Determination



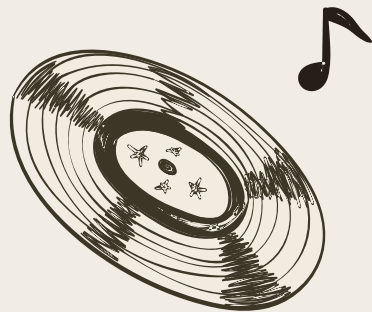
# 01. Data Preparation

## Columns of Interest

- Selecting the Frequency columns of each genre (16 total)

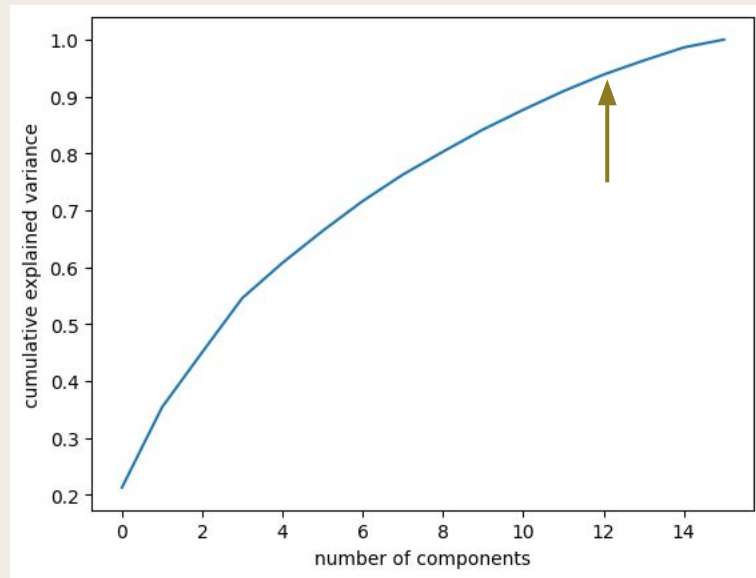
## Alter the Data

- Assign values to the frequencies to make them numeric
  - 1 - Never
  - 2 - Rarely
  - 3 - Sometimes
  - 4 - Very frequently
- This scales the data, so no other scaling is necessary

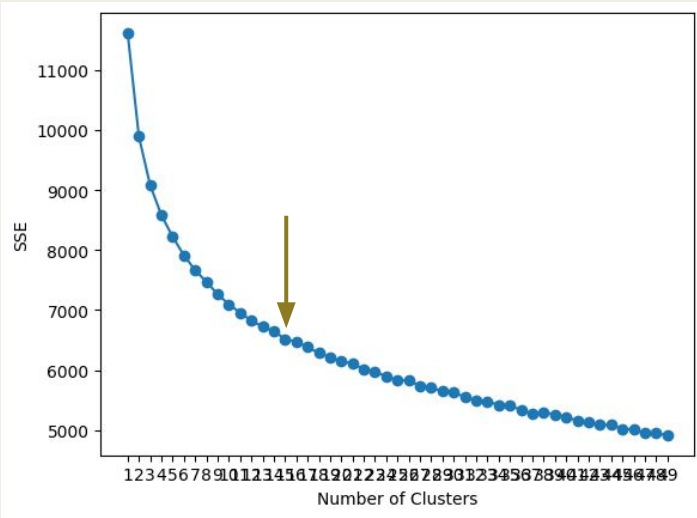


## 02. Principal Component Analysis

As explained by the graph, 12 components covers over 90% of the cumulative explained variance, so the k-means model will be completed with 12 components.



# 03. Cluster Determination



According to the graph, the “elbow” dictates the number of clusters to use for k-means.

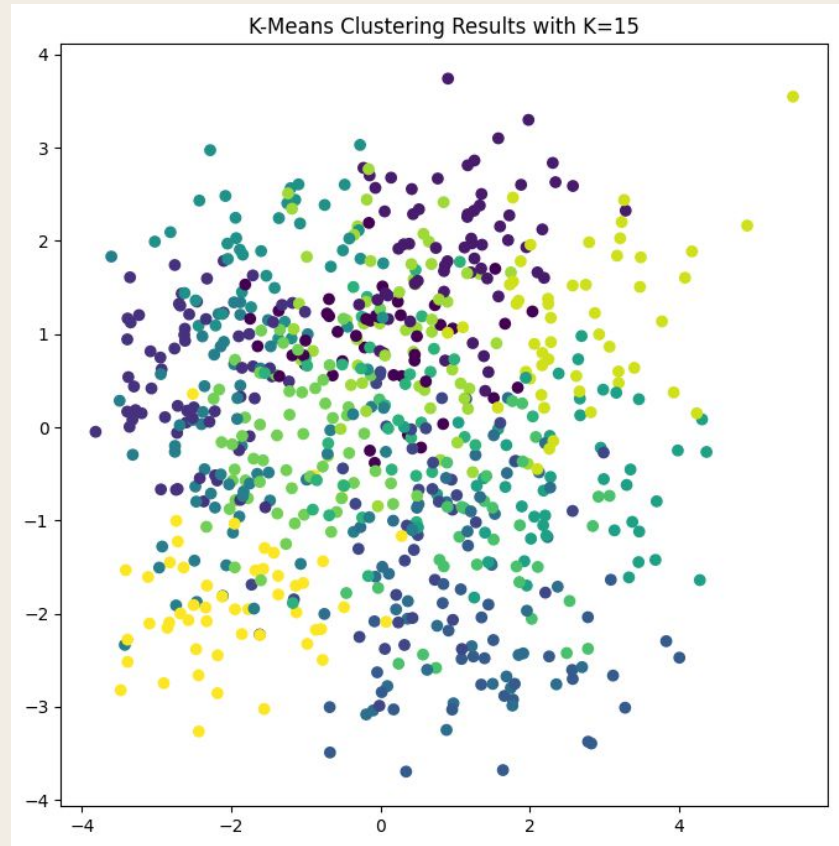
I chose 15 clusters because there is a lot of data to explain, so I wanted more opportunity for specific clusters without there being too many.



# The Actual Model

This model was made with:

- 12 components
- 15 clusters



# Examining the Levels of Mental Illness for Each Cluster from K-means

For each mental illness, group the data from the column where participants ranked the level they experienced the mental illness with the column that outlines which cluster each record was in.

Compute the mean mental illness level associated with each cluster.

Note: the graphical representations are bar graphs because there was too much data for swarm plots to be of use visually.

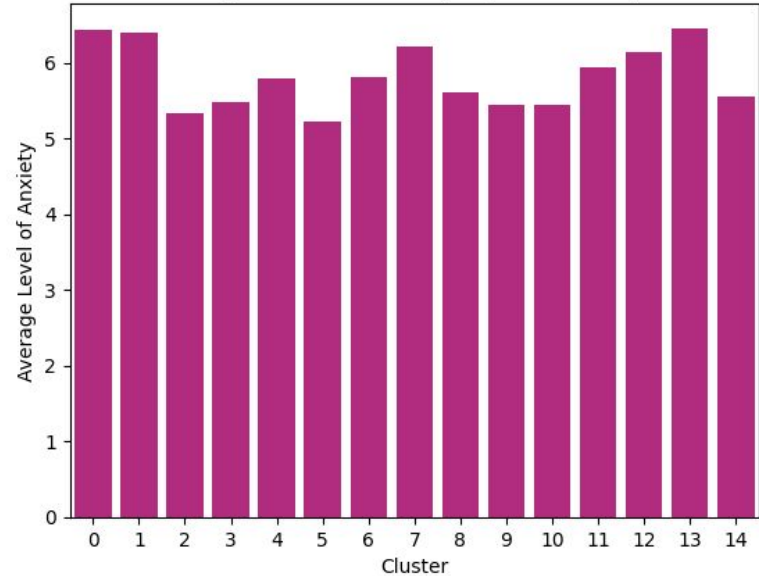


# Anxiety

The top three clusters that experience anxiety the most on average are **clusters 13, 0, and 1.**

These three clusters will be used for analysis of which genres participants with higher levels of anxiety listen to the most.

Clusters vs the Average Levels of Anxiety Experienced by the Cluster Group

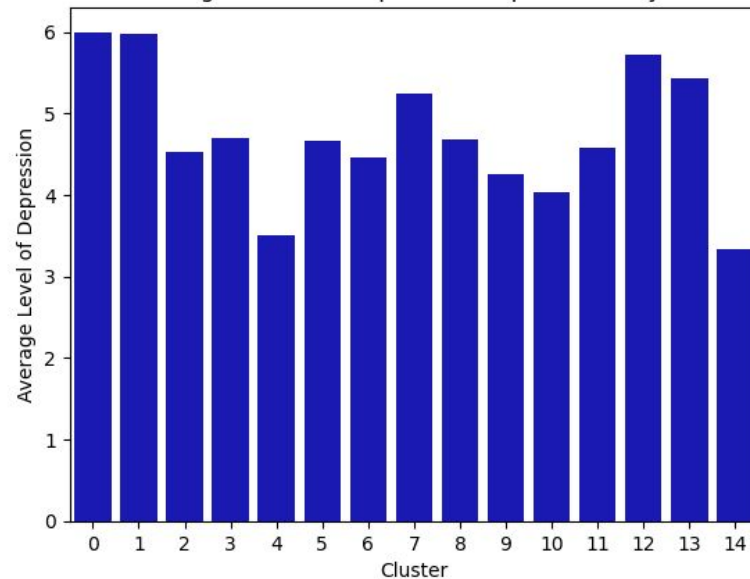


# Depression

The top three clusters that experience depression the most on average are **clusters 0, 1, and 12.**

These three clusters will be used for analysis of which genres participants with higher levels of depression listen to the most.

Clusters vs the Average Levels of Depression Experienced by the Cluster Group

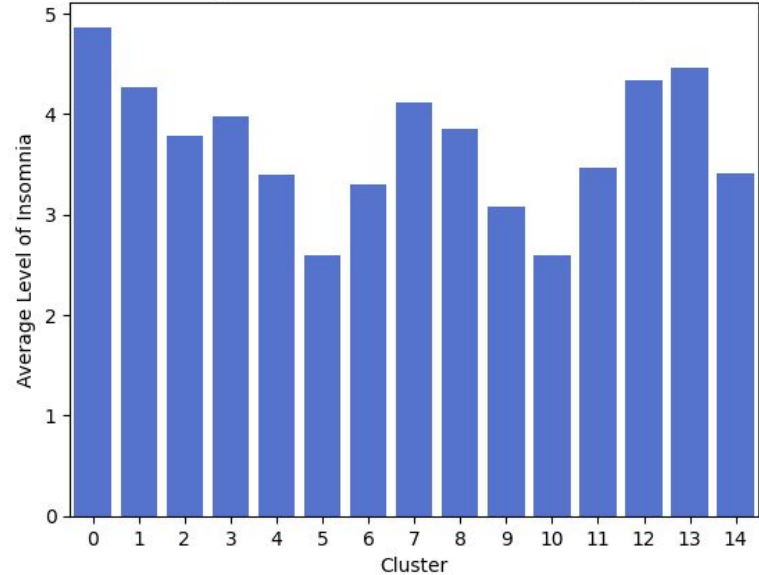


# Insomnia

The top three clusters that experience insomnia the most on average are **clusters 0, 13, and 12.**

These three clusters will be used for analysis of which genres participants with higher levels of insomnia listen to the most.

Clusters vs the Average Levels of Insomnia Experienced by the Cluster Group

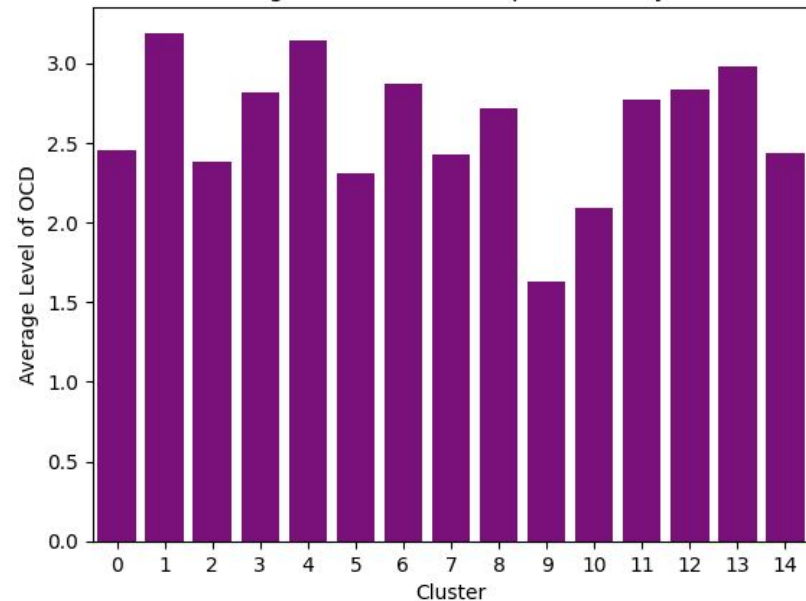


# OCD

The top three clusters that experience ocd the most on average are **clusters 1, 4, and 13.**

These three clusters will be used for analysis of which genres participants with higher levels of ocd listen to the most.

Clusters vs the Average Levels of OCD Experienced by the Cluster Group



# Genre Rankings per Cluster

Based on the top clusters for each mental illness, determine the highest and lowest ranked genres. This was done by computing the percentage: the number of times the frequency appears in a genre divided by the total number of values in the cluster. This allows us to compare the ratios of the frequencies of the genres versus just the count.

The highest ranking genres should be included in the playlists, and the lowest should be avoided for that mental illness.

# Breakdown of Top Genres For High (3&4) Ratings and Low (1&2) Ratings

## Anxiety:

High: Hip Hop (3), Rap (3), Rock (3), Pop (2)

Low: K pop (4), Gospel (3), Classical (2), Video Game Music (2), Latin (2)

## Depression:

High: Rock (3), Pop (3), Rap (3), Hip Hop (3), Metal (2)

Low: Gospel (3), K Pop (3), Classical (2), R&B (2), Latin (2), EDM (2)

## Insomnia:

High: Rock (3), Rap (3), Hip Hop (3), Metal (2), Pop (2)

Low: Classical (3), Gospel (3), K pop (3), EDM (2), Video Game Music (2)

## OCD:

High: Hip Hop (3), Rock (2), Pop (2), Rap (2), R&B (2), Latin (2)

Low: K pop (4), Gospel (3), Classical (2), Video Game Music (2)



# A Note About the Results

The recommendations from the insights section are from only a small portion of music listeners. To gain a better understanding of the listening preferences of Spotify users, Spotify should expand the dataset to better represent their 600 million users.

Additionally, this data is based on a survey, so there is space for human error and bias.





Thank you!