

Analyzing the Impact of Musical Habits on Mental Health

Jason Lebov & Peter Loiselle

Research Motivation

- Music is a staple in society and has become part of **daily routines**
 - Listening to Favorite Artists and Songs Daily
 - Commute, Gym, School, etc.
- Music can be a **positive stimulant**
 - Used Relax or Relieve Stress

Research Question: Could the types of music a person listens to correlate with certain types of mental disorders?



Agenda

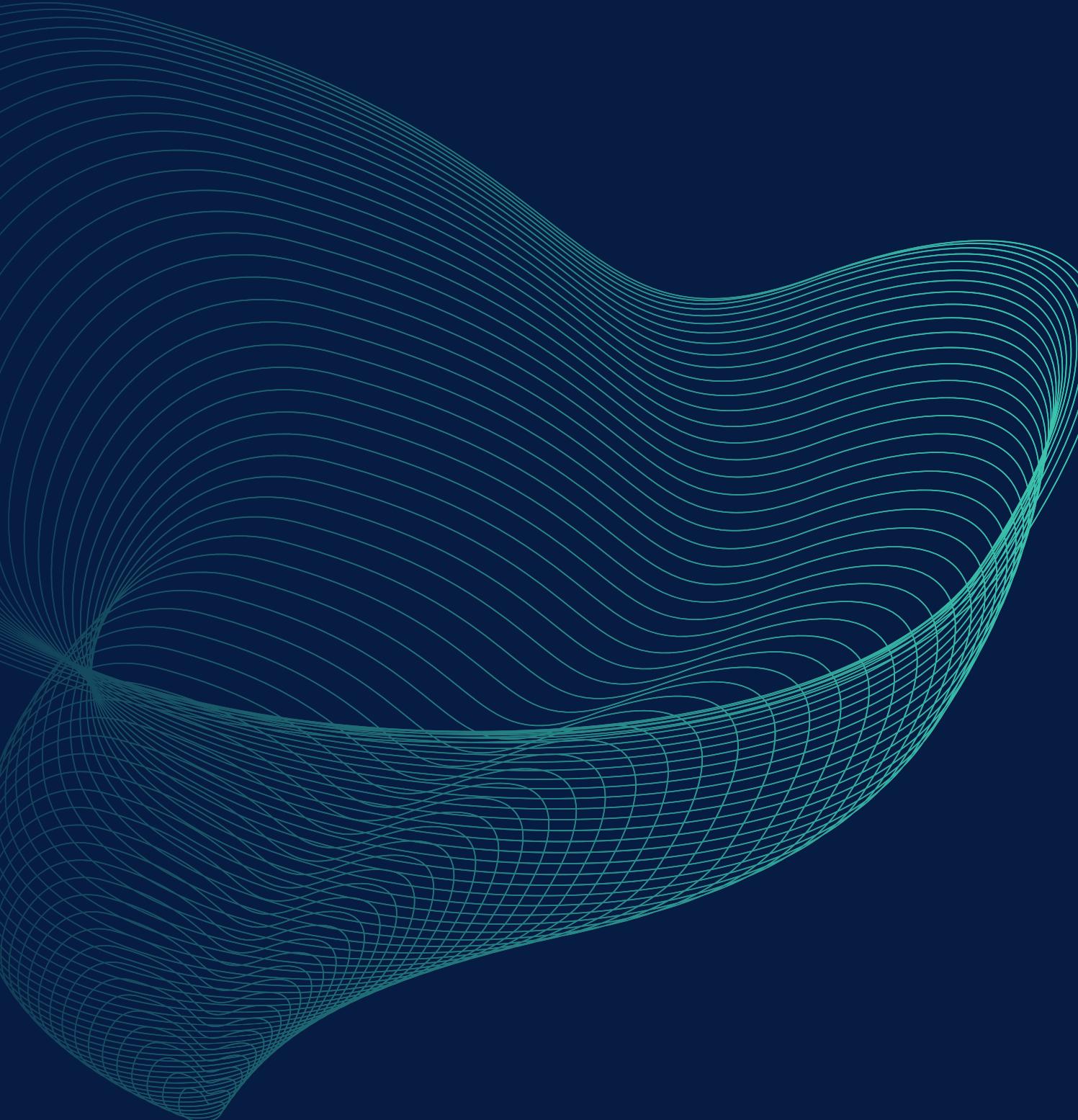
A decorative graphic in the bottom-left corner consisting of numerous thin, light blue lines that curve and overlap, creating a sense of depth and motion.

Data
Exploration

Model
Selection &
Creation

Data
Preprocessing

Model
Evaluation



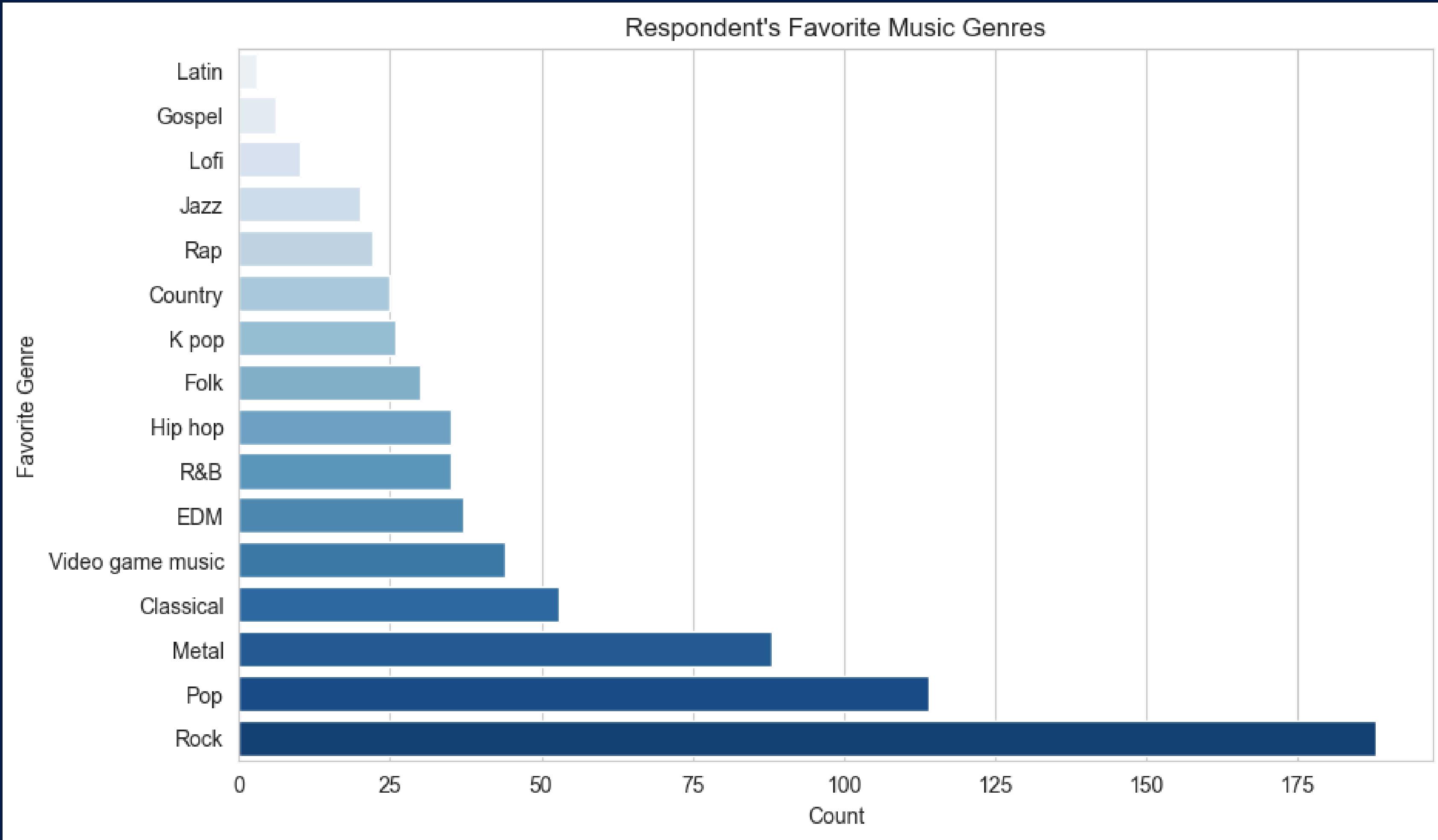
Data Exploration

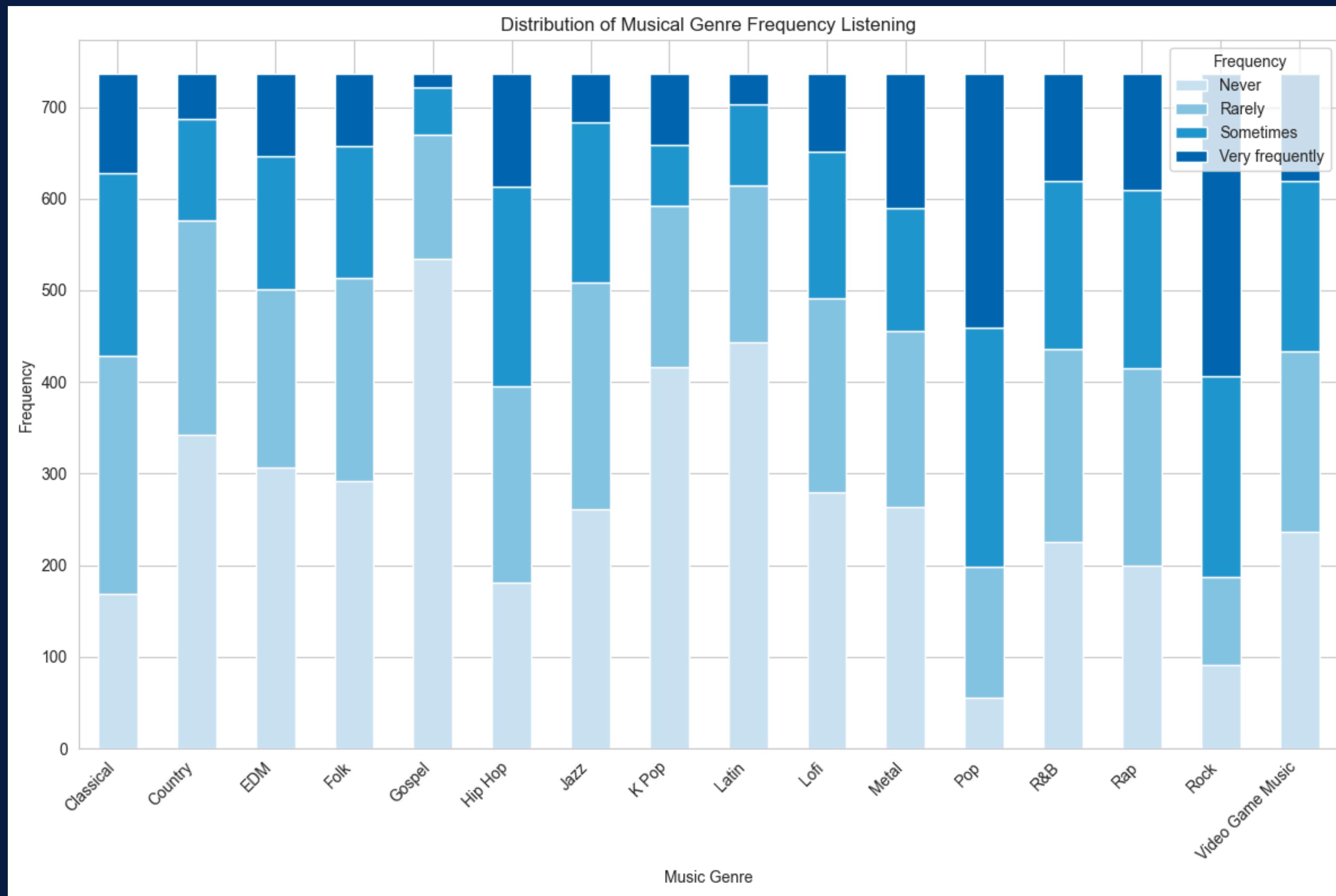
Data Exploration

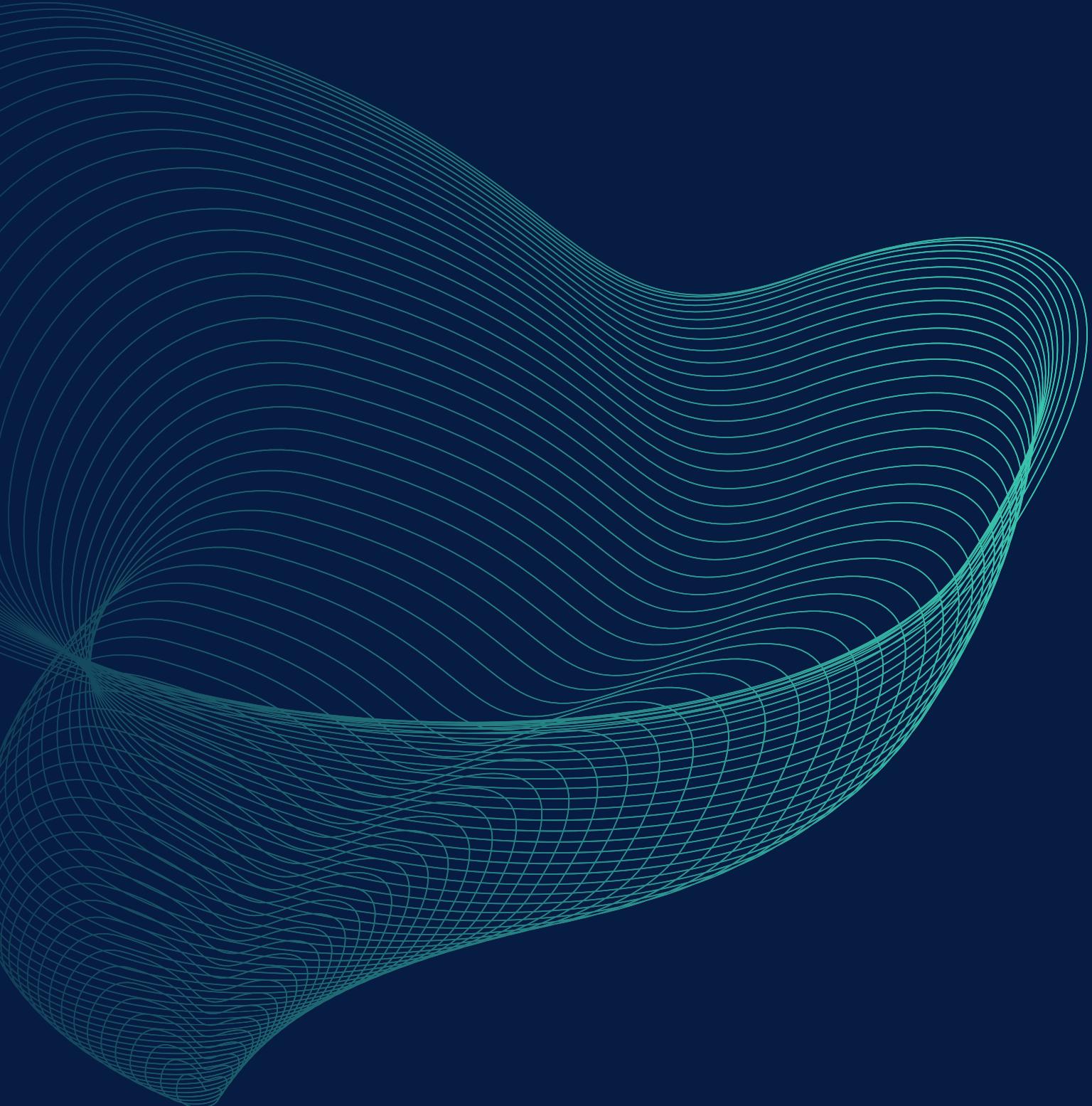
Dataset:

- Kaggle Dataset - 736 Samples, 33 Features
- Information Regarding Musical Habits...
 - Favorite Genre
 - Listening Frequency of Different Genres
 - Musically Involved? (Composer, Instrumentalist, etc.)
 - *Ordinal Metric of Mental Illness Impact (*Anxiety, Depression, OCD, Insomnia*) --> Range 1-10









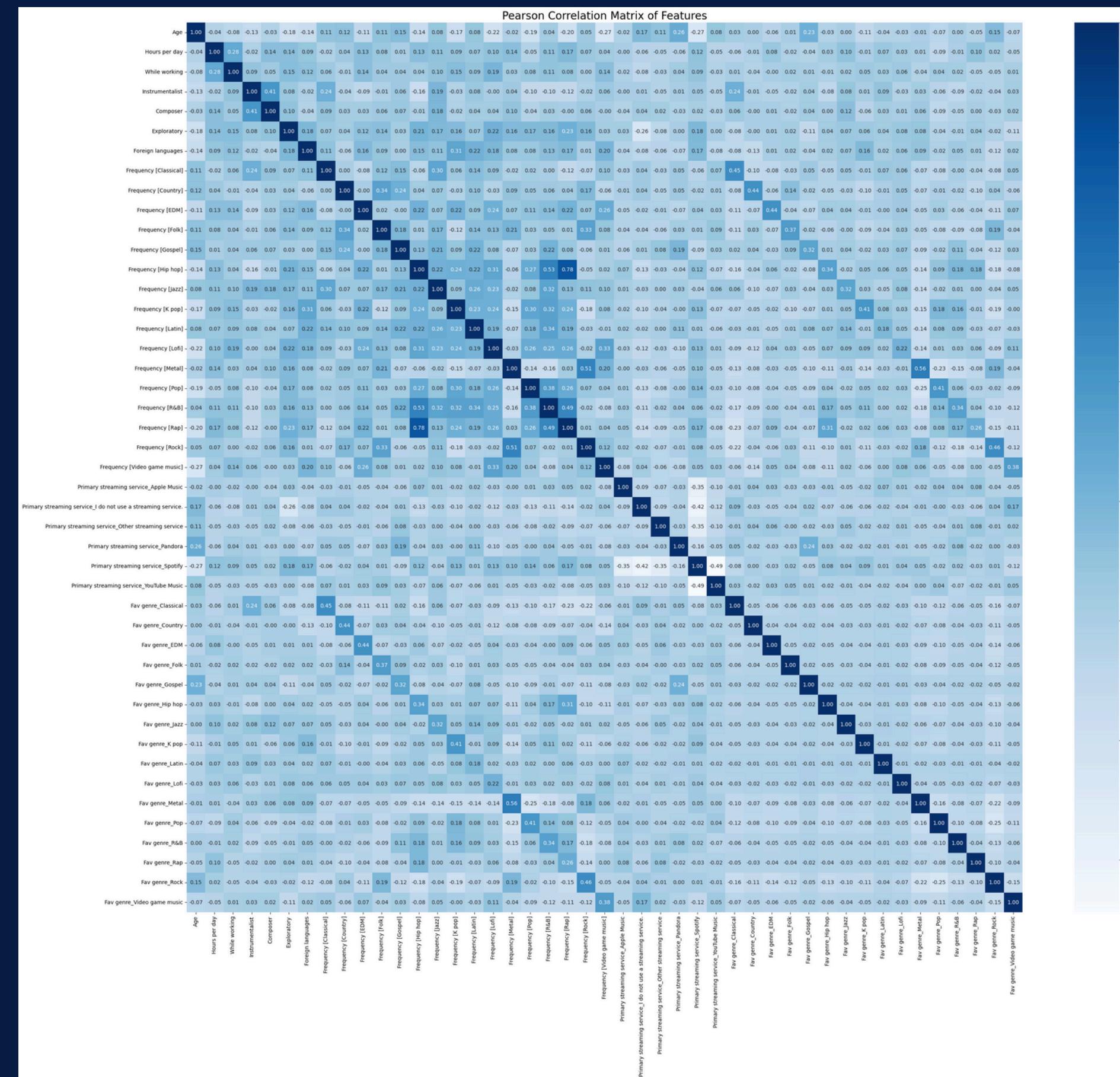
Data Preprocessing

Steps Taken

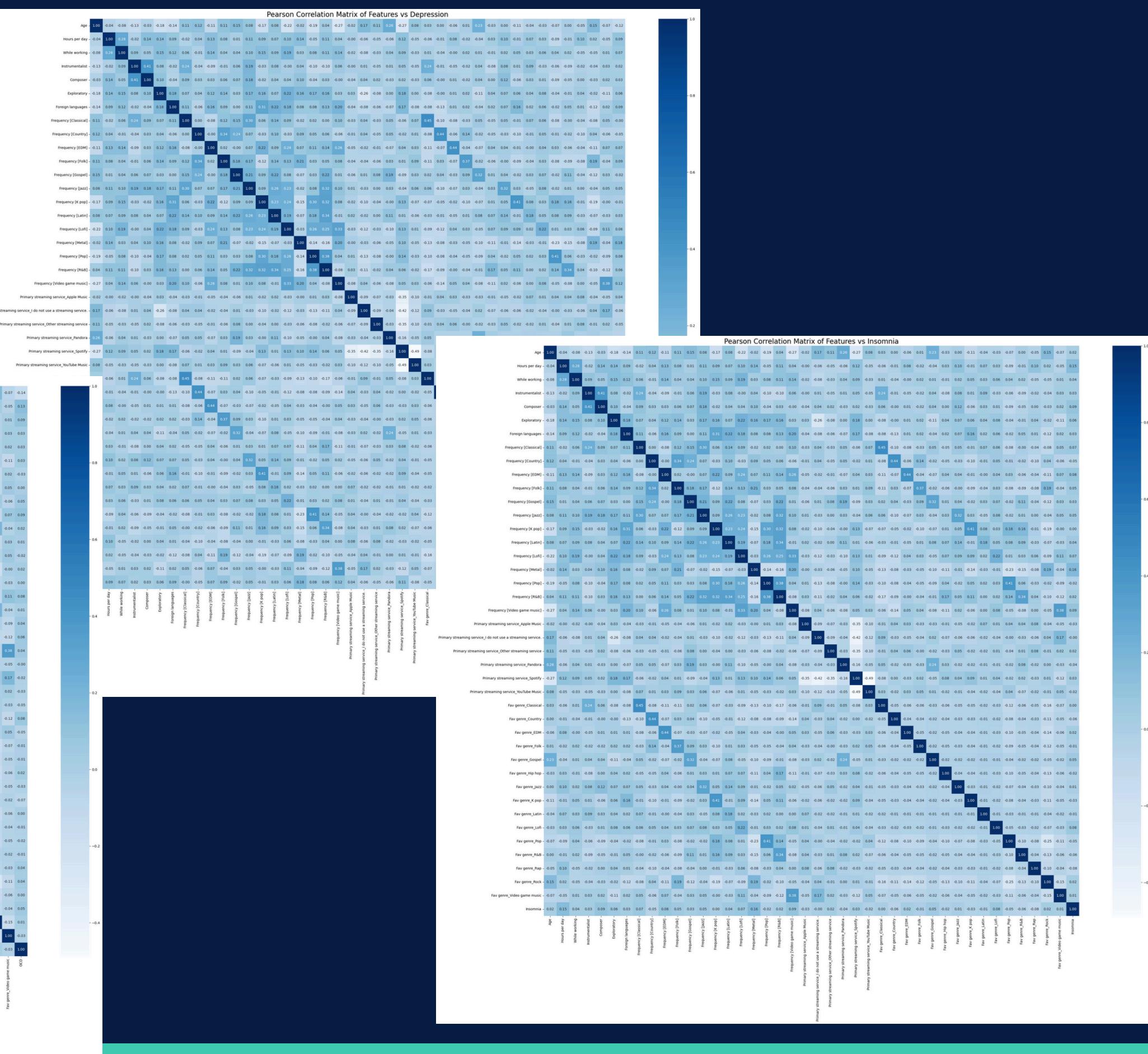
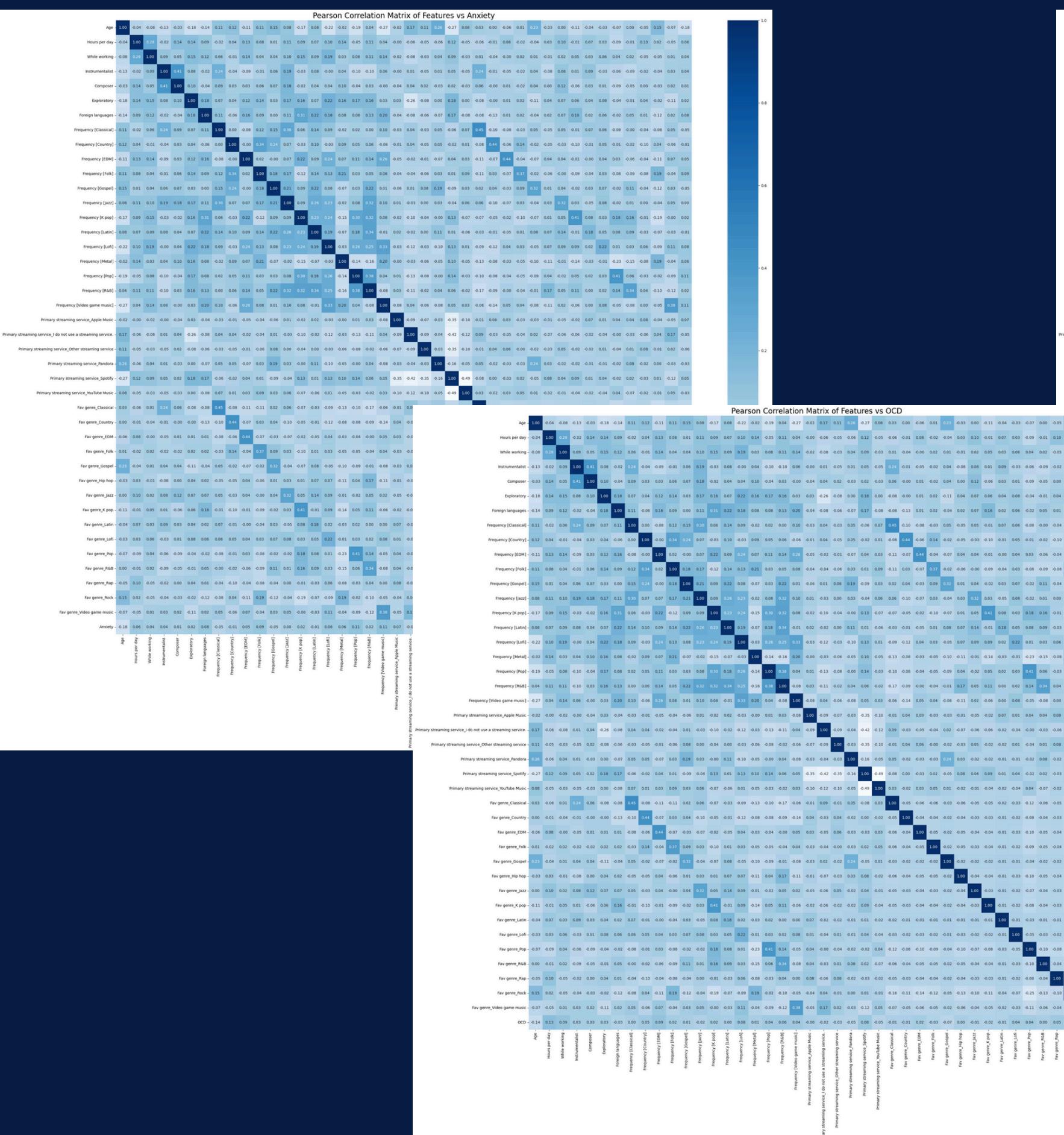
- **Preliminary Drop Features/Samples**
 - Irrelevant Features & NaN's
- **Convert Labels to Categorical Ranges**
 - Asymptomatic (0.0-1.0), Mild (1.5-4.0), Moderate (4.5-7.0), Severe (7.5-10.0)
- **One-Hot-Encode Features**
- **Label-Encode Features**
- **Pearson Correlation Matrix of Features & Remove Features**
 - Delta = 0.75, Gamma = 0.01
- **Create 4 Datasets --> Split Up by Mental Illness**
- **Label Encode Label Options**
 - 0 = *Asymptomatic*, 1 = *Mild*, 2 = *Moderate*, 3 = *Severe*
- **Train-Test Splits of 4 Datasets [80:20]**
- **Min-Max Scale Numeric Features**



Step 5: Pearson Correlation Matrices



Step 5: Pearson Correlation Matrices

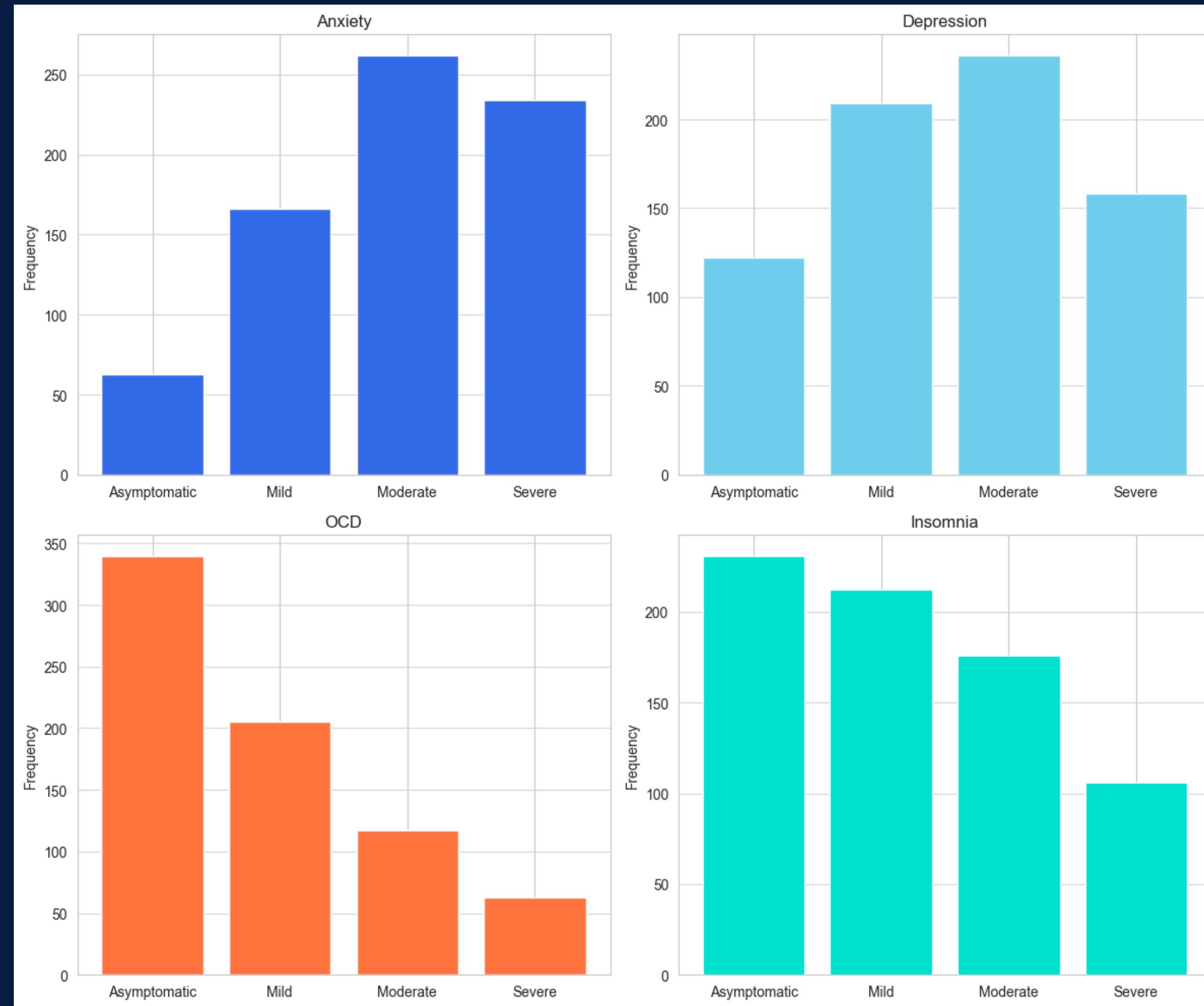


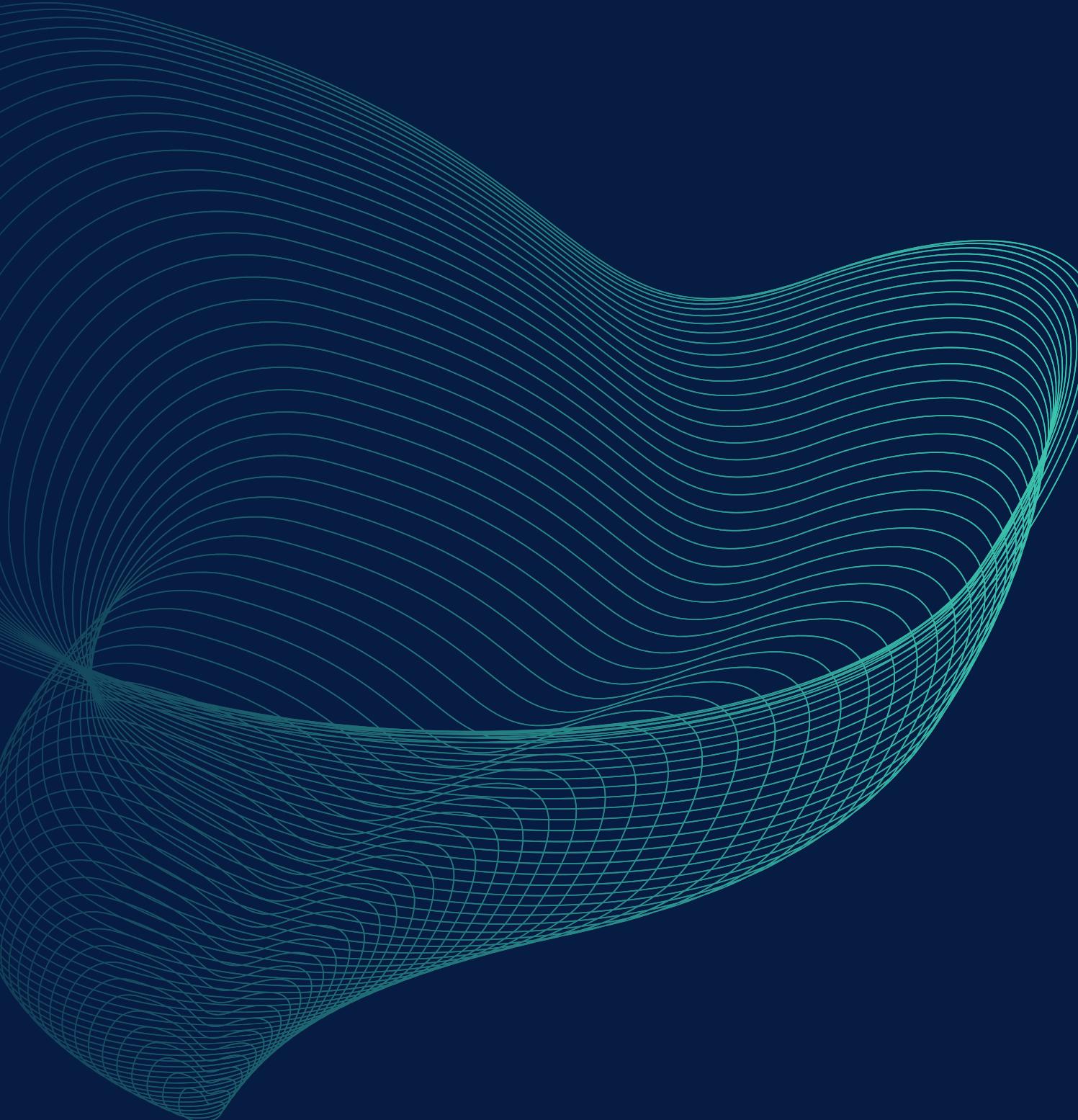
Post Preprocessing

Datasets:

- 4 Datasets [*Anxiety, Depression, OCD, Insomnia*]
 - **Shape:** 725 Samples, 45 Features
- Restructured Labels
 - **Asymptomatic** = 0.0-1.0 --> 0
 - **Mild** = 1.5-4.0 --> 1
 - **Moderate** = 4.5-7.0 --> 2
 - **Severe** = 7.5-10.0 --> 3







Model Selection & Creation

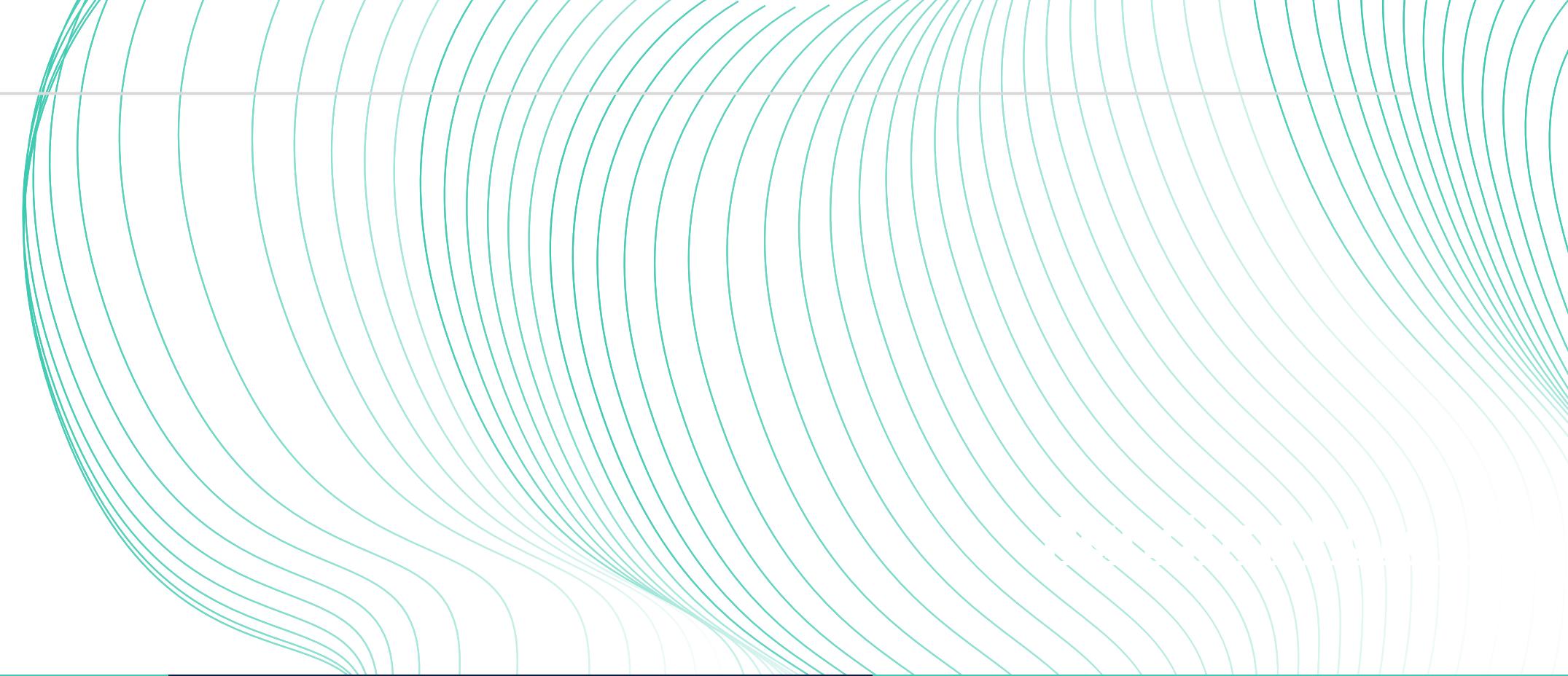
Selected Models

KNN

Decision
Tree

Random
Forest

Neural
Network



Model Creations

GridSearchCV

- **KNN**
 - param grid = # Neighbors (1-150) & Weights
- **Decision Tree**
 - param grid = Criterion, Max Depth, & Min Samples Split & Leaf Size
- **Random Forest**
 - param grid = # Estimators, Max Depth, & Min Samples Split & Leaf Size
- **Neural Network**
 - param grid = Hidden Layer Sizes, Activation Function, Solver, Alpha & Learning Rate



KNN Model – Parameters

Anxiety

- # Neighbors = 30, Weight = ‘Uniform’

Depression

- # Neighbors = 140, Weight = ‘Distance’

OCD

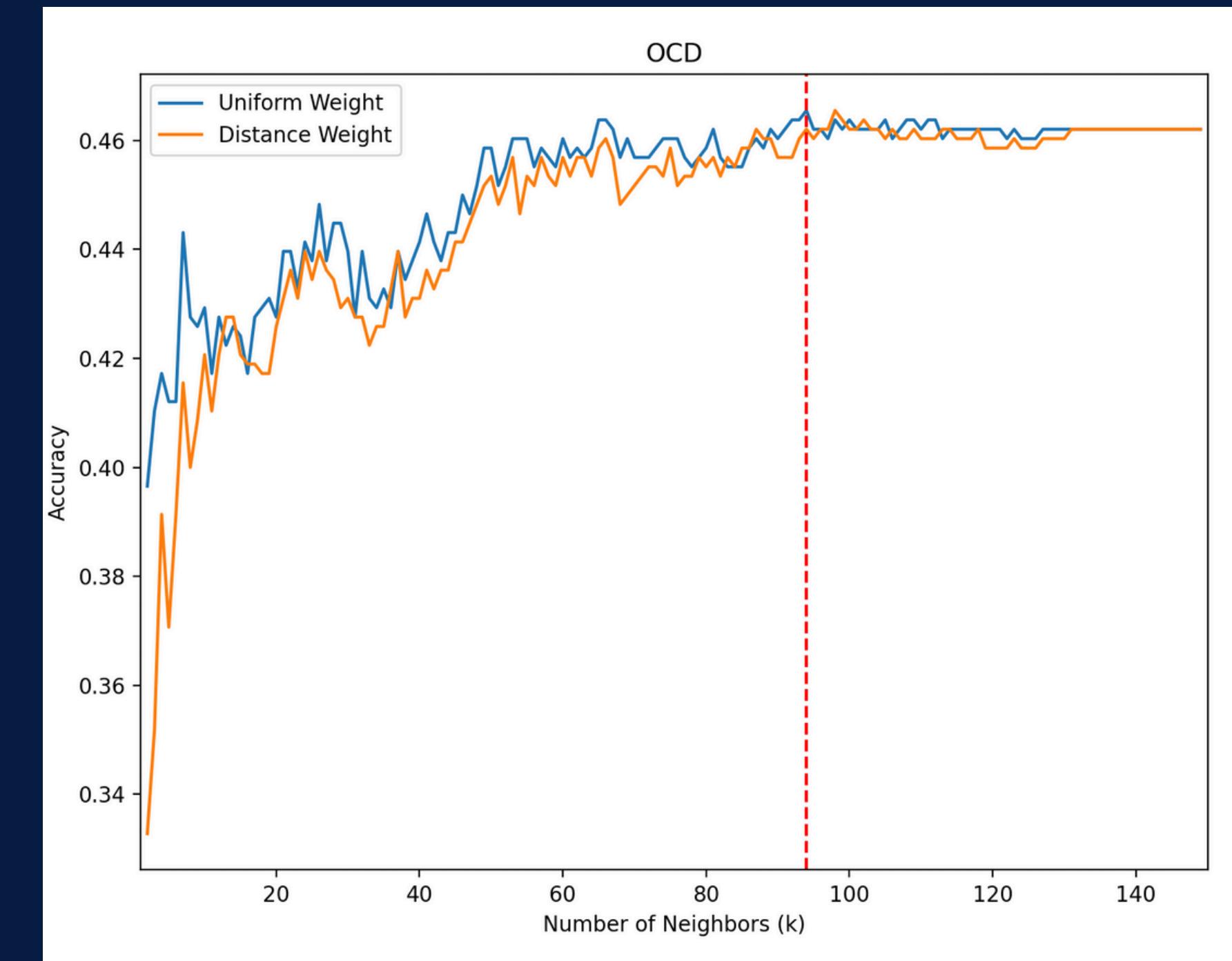
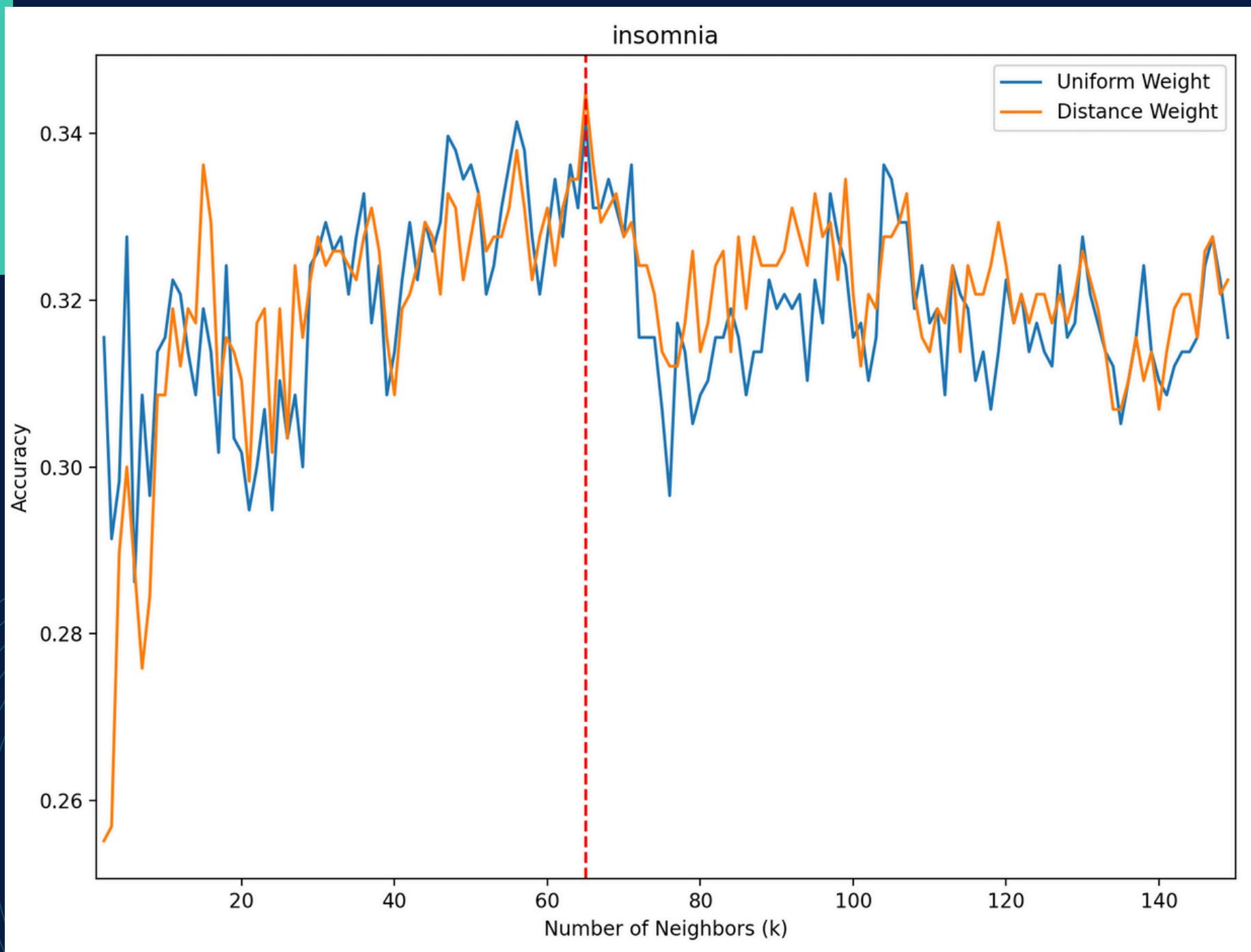
- # Neighbors = 94, Weight = ‘Uniform’

Insomnia

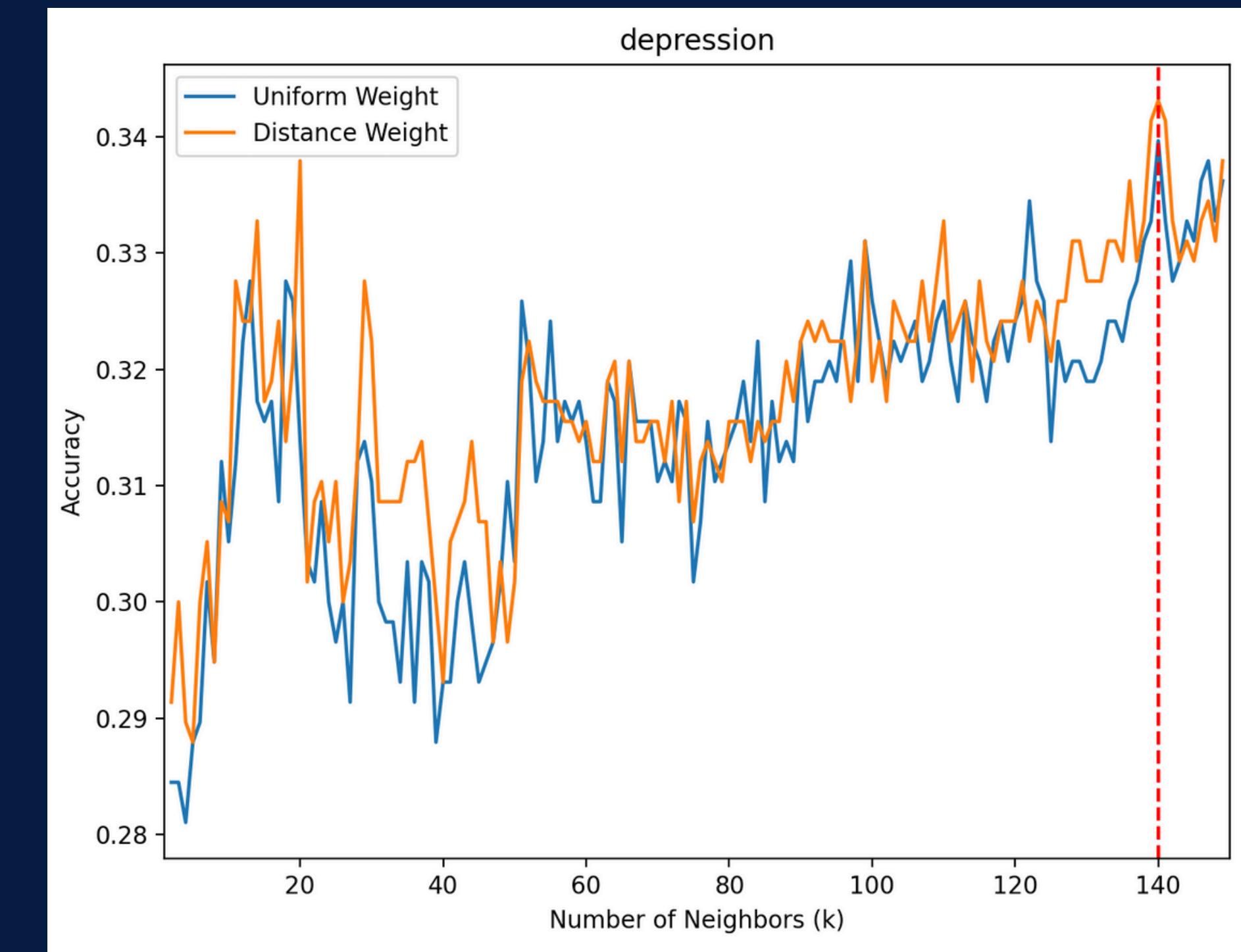
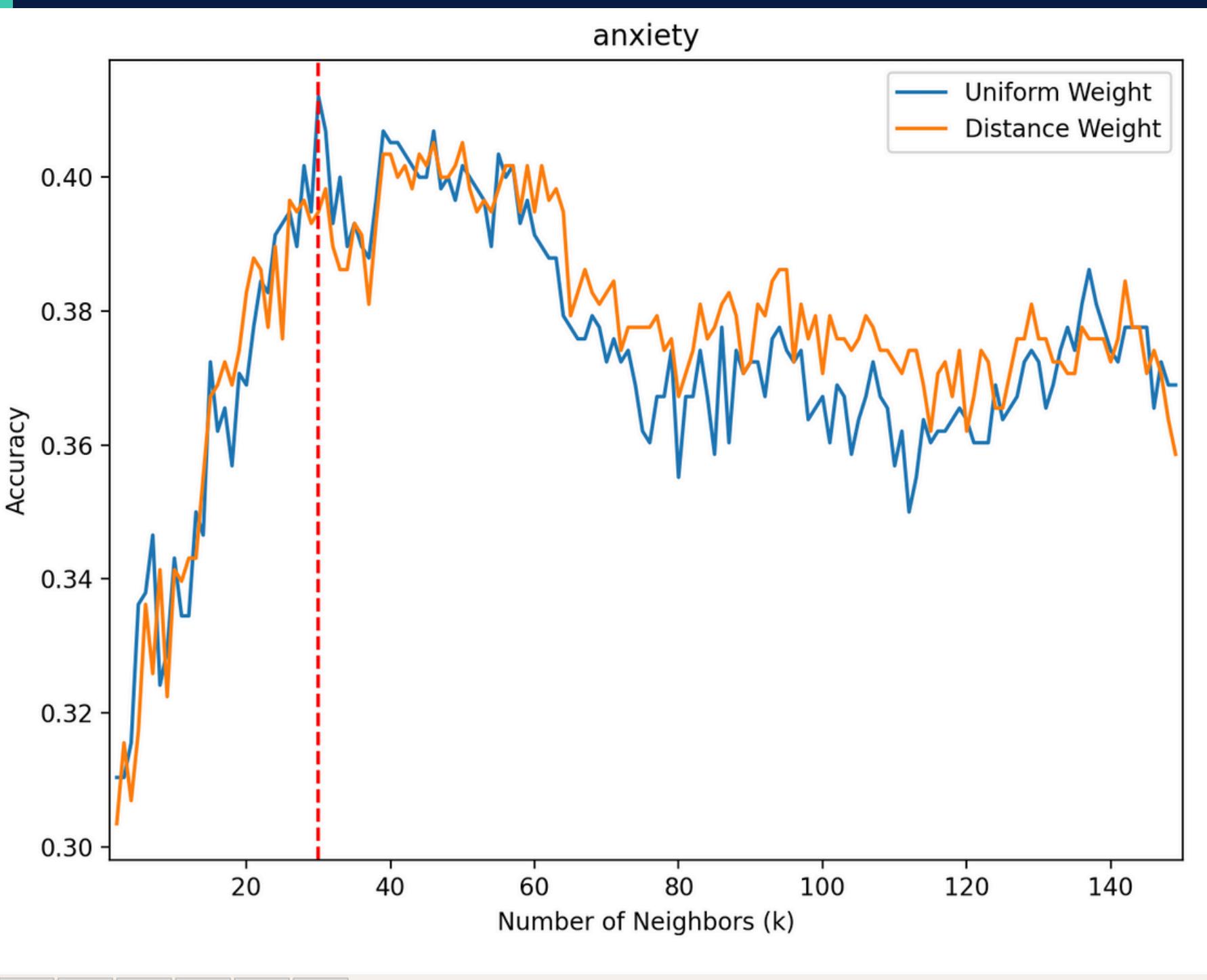
- # Neighbors = 65, Weight = ‘Distance’



Optimal Number of Neighbors



Optimal Number of Neighbors



Decision Tree Model – Parameters

Anxiety

- Criterion = 'gini', Max Depth = None, Max Features = 'log2', Min Samples Split = 5, Leaf Size = 4

Depression

- Criterion = 'entropy', Max Depth = 20, Max Features = 'sqrt', Min Samples Split = 5, Leaf Size = 4

OCD

- Criterion = 'entropy', Max Depth = 40, Max Features = 'log2', Min Samples Split = 5, Leaf Size = 4

Insomnia

- Criterion = 'gini', Max Depth = None, Max Features = 'log2', Min Samples Split = 5, Leaf Size = 4

Random Forest Model – Parameters

Anxiety

- # Estimators = 35, Max Depth = 7, Min Samples Split = 2, Leaf Size = 2

Depression

- # Estimators = 39, Max Depth = 31, Min Samples Split = 4, Leaf Size = 6

OCD

- # Estimators = 39, Max Depth = 5, Min Samples Split = 7, Leaf Size = 7

Insomnia

- # Estimators = 53, Max Depth = 23, Min Samples Split = 8, Leaf Size = 2

Neural Network - Parameters

Anxiety

- Hidden Layer Sizes = (50, 50), Activation Function = 'relu', Solver = 'sgd', Alpha = 0.0001, Learning Rate = 'constant'

Depression

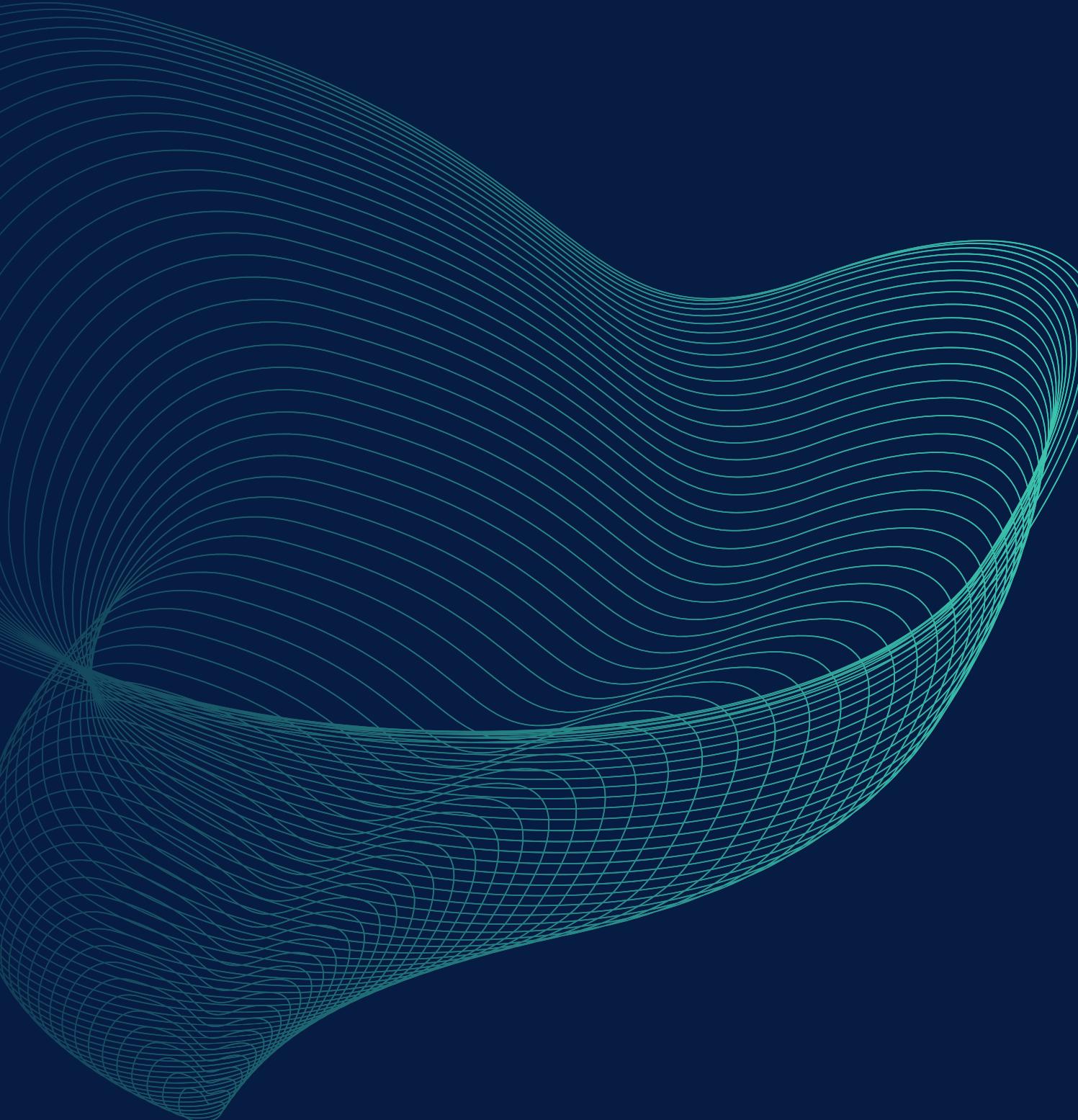
- Hidden Layer Sizes = (50, 50), Activation Function = 'relu', Solver = 'adam', Alpha = 0.001, Learning Rate = 'constant'

OCD

- Hidden Layer Sizes = (25, 25, 25), Activation Function = 'relu', Solver = 'sgd', Alpha = 0.0001, Learning Rate = 'adaptive'

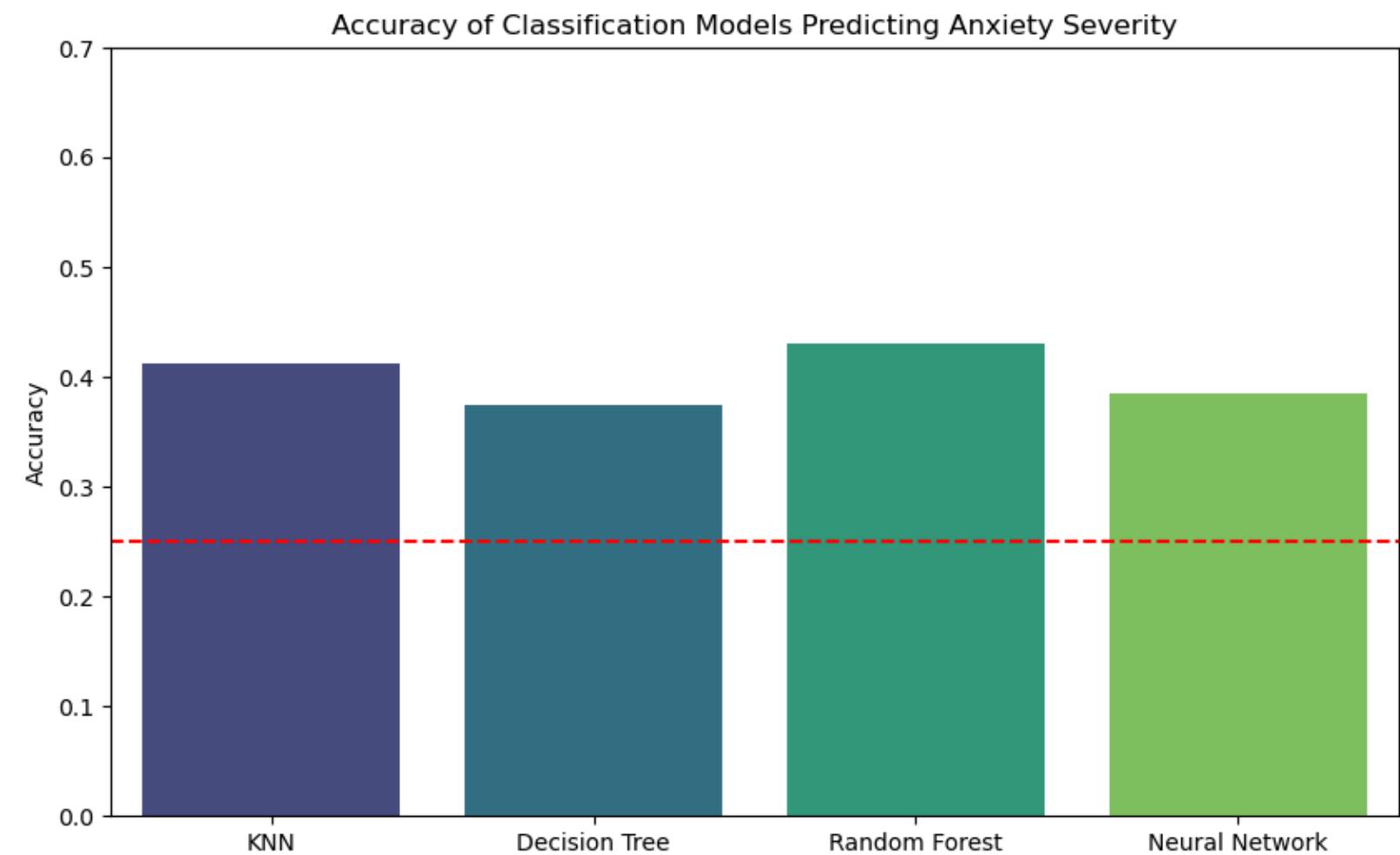
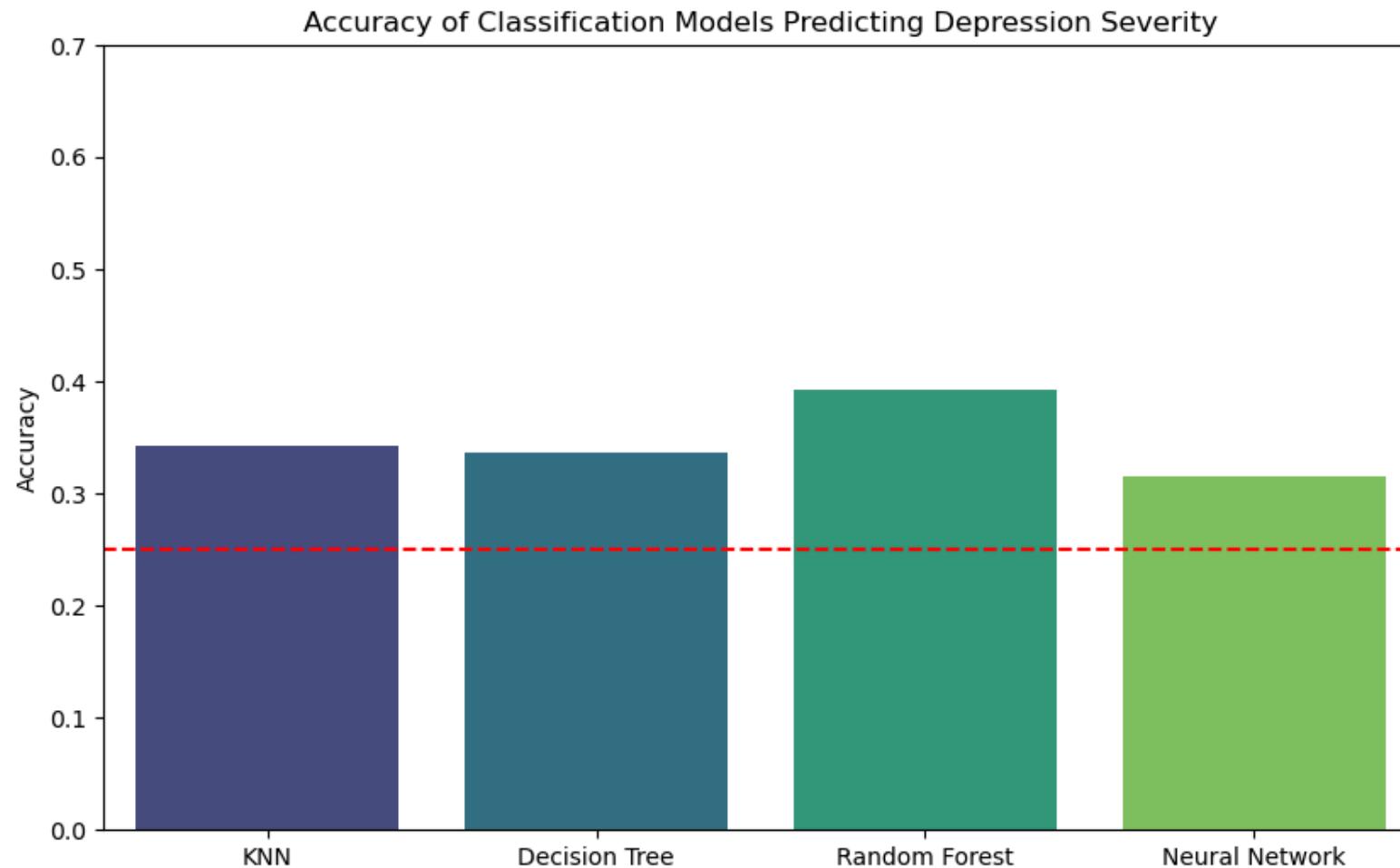
Insomnia

- Hidden Layer Sizes = (50, 50), Activation Function = 'tanh', Solver = 'sgd', Alpha = 0.0001, Learning Rate = 'constant'

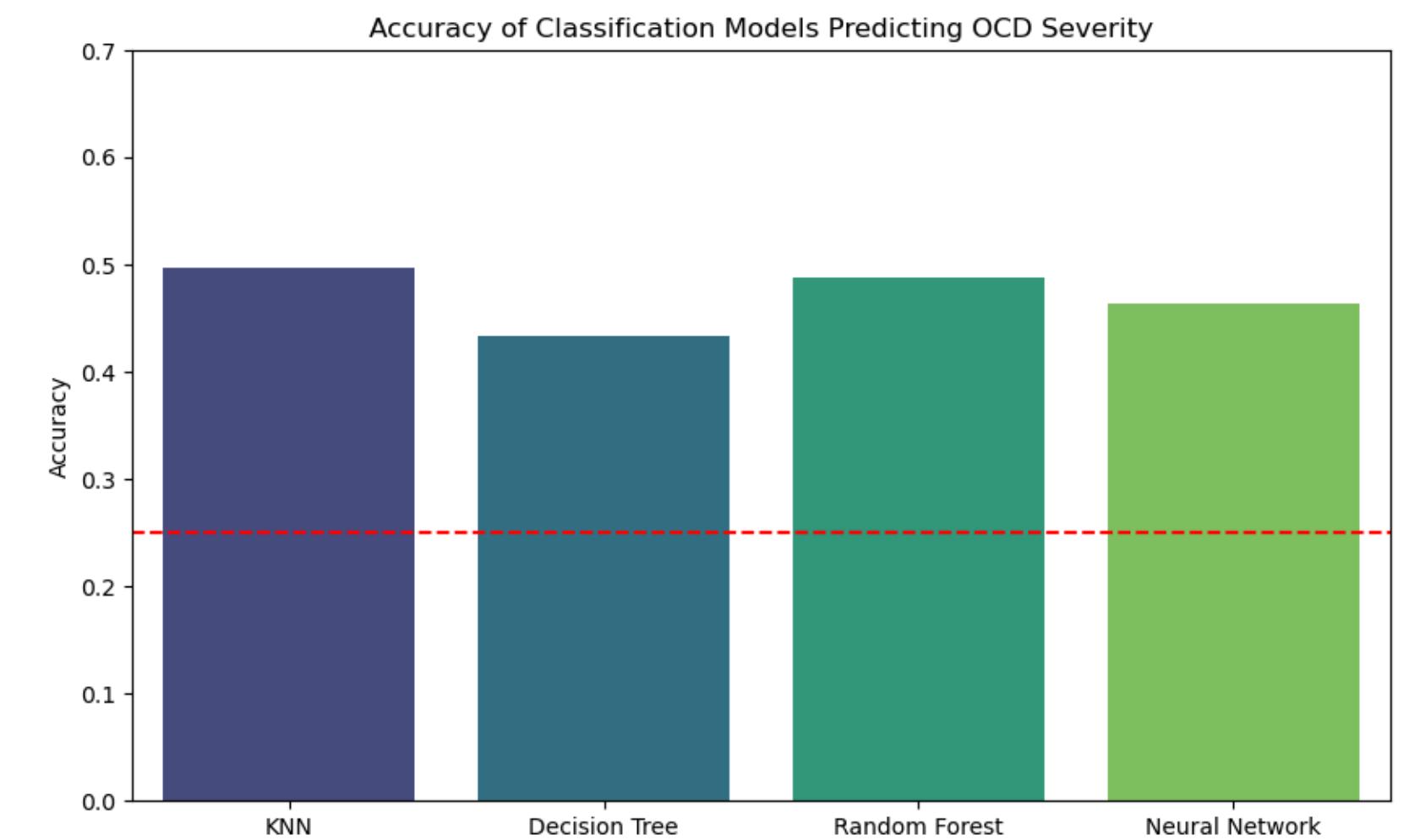
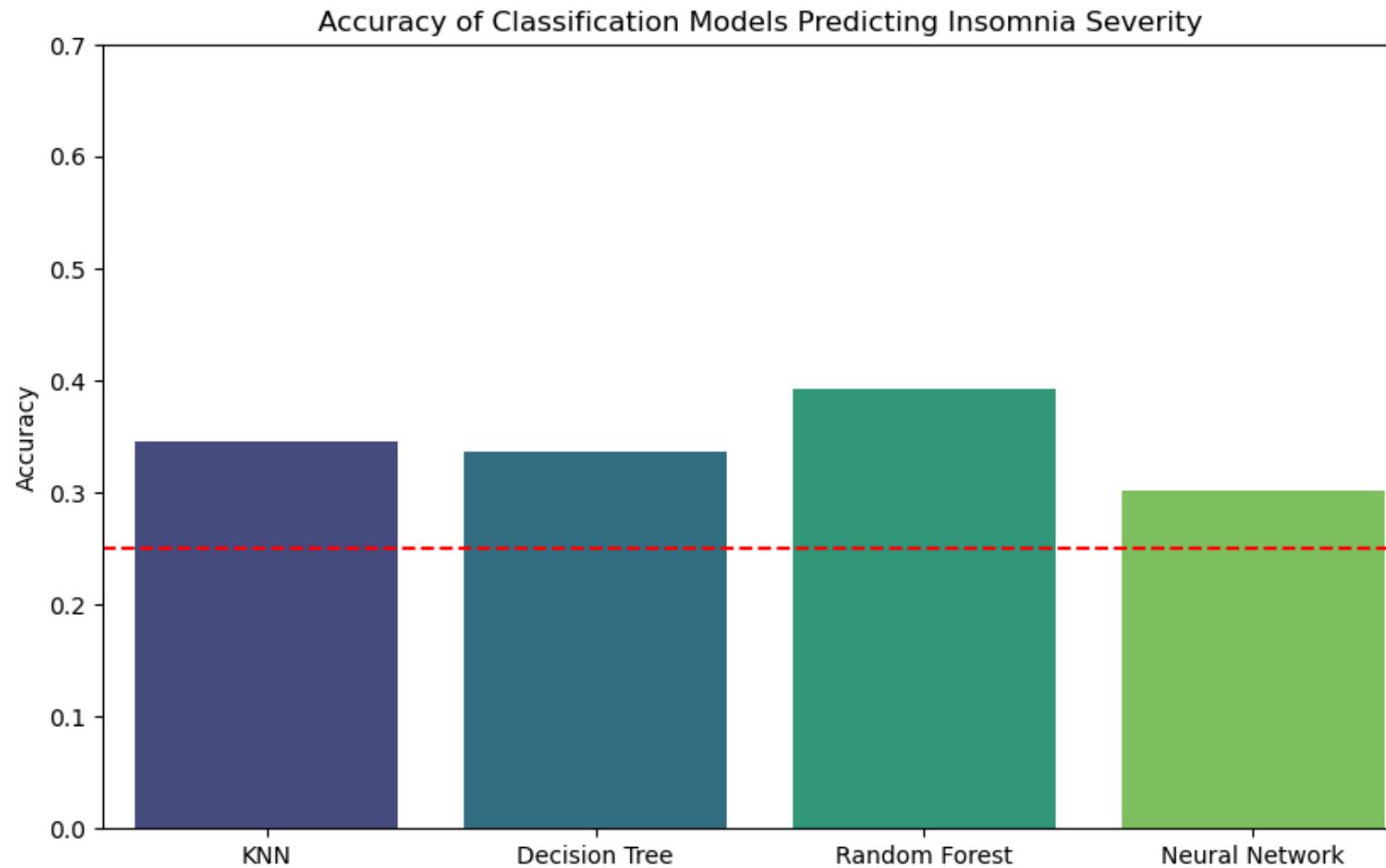


Model Evaluation

Overall Results - Accuracy



Overall Results - Accuracy



Overall Results - Accuracy

	Depression	Anxiety	OCD	Insomnia
CHANCE		-- 0.25 --		
KNN	0.290	0.393	0.500	0.310
Decision Tree	0.343	0.369	0.424	0.345
Random Forest	0.303	0.352	0.510	0.269
Neural Network	0.324	0.269	0.331	0.262

Best Model

Current Findings

- Accuracies are **higher than random chance (25%) for all 4 mental disorders**
 - Showing signs of a possible correlation between music preferences/habits & mental health.
- Accuracies were slightly higher for **Depression (34%), Insomnia (35%), Anxiety (39%)**
 - **Overall:** Any correlation between music preferences/habits and these three disorders is likely very small
- Higher Accuracy for **OCD (51% Accuracy)**
 - Our model could accurately predict **over half** of the respondents' experiences with OCD based on their musical characteristics/habits

Data Preprocessing

- Convert Problem to Binary Classification
- Extract More Features from the Dataset
- Decrease/Increase Pearson Thresholds
- Fix Class Imbalance Problem (SMOTE) + Add more Sample

Model Creation

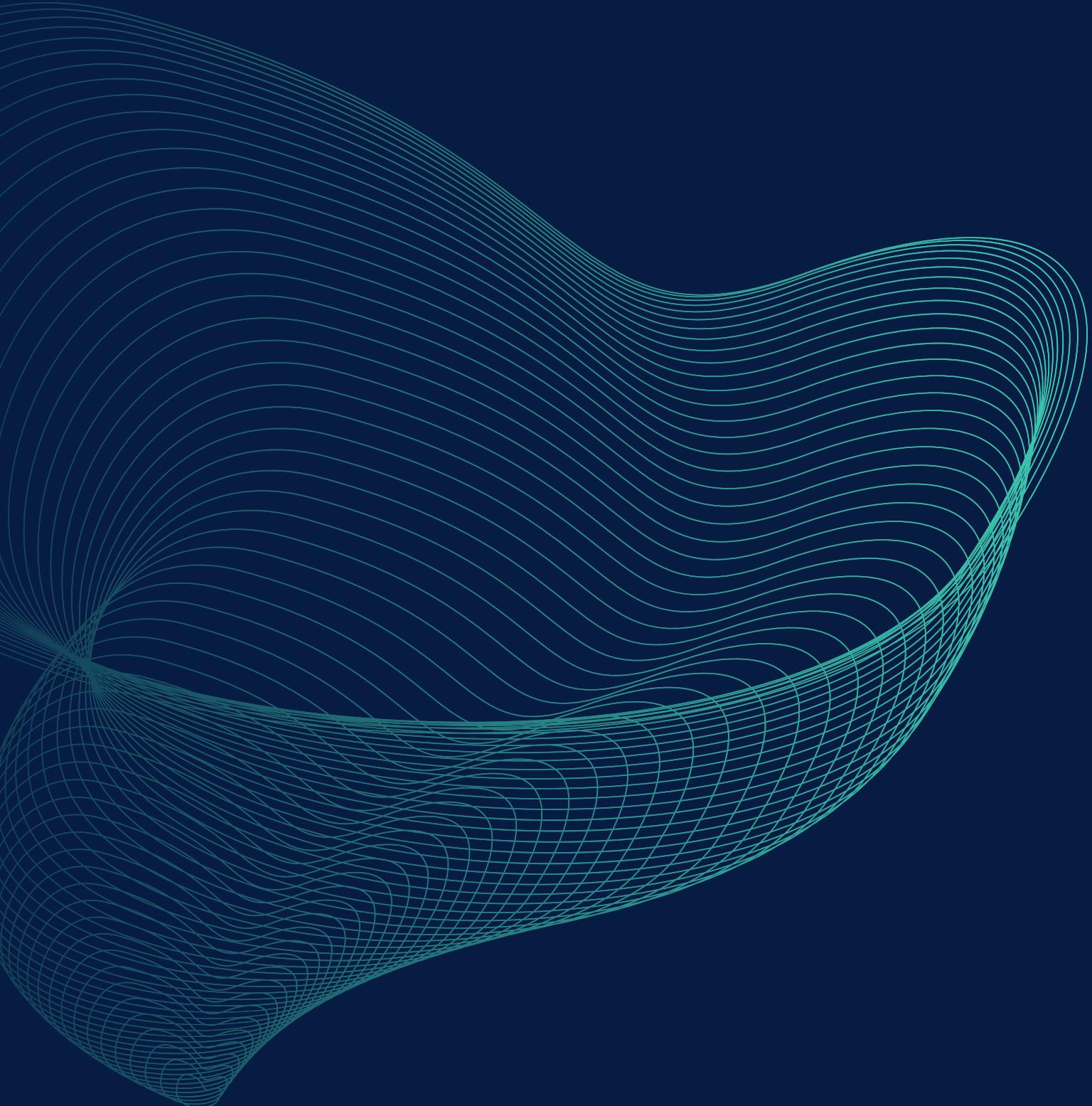
- Add More Parameters to GridSearchCV
- Attempt Different/More Complex Models
- Feature Importance

Next Steps

Data & Model Changes to Improve Results

Evaluation Metrics

- Look Deeper into Random Chance vs. Model Accuracy
- Plot ROC Curves for Model
- Look into More Metrics (MSE, F1, etc.)



Thank You