Statistical/Hypothetical Question

* Does Music help in improving Mental health?
* Hypothesis: Yes, in general, many people say that Music is soothing and calming to the brain and it improves the mental health
* Null Hypothesis: Music improves mental health by more than 10%

The outcome of the Exploratory Data Analysis (EDA)

* From the EDA, we can conclude that the average age of the individuals who listen to music is 25 years. Mode is 18 years, which means many people of age 18 listen to music. The variance or the spread is around 136. The above tails show how quickly the probability drops off as we move away from the mode. Also, we can conclude that on average, individuals listen to music for 3.7 hours per day. The mode is 2 hours. Variance or the spread is around 3 standard deviations.
* Probability mass Function among teenagers is more when compared to the pmf of others.
* With CDF, it looks like about 80% of people are around 25 years. Here the mode is 18. There are few values below 13 years. So, the CDF in this range is flat. From this, we can conclude that people who are around 13 to 25 are listening to more music.
* When we generated a normal probability plot with Age, we can see clearly that the data deviate from the model systematically.
* Most individuals from ages 13 to 25 listen to music around 2-3 hours a day
* Individuals from 13 to 25 years are listening to music for 2-3 hours and the depression rate among them is more.
* The correlation between age and hours per day is negative, it is inversely proportional. We cannot say that as age increases, it is the cause of the decrease in 'Hours per day'. They may have different factors causing it. Since the correlation between Anxiety and depression is positive. They are directly proportional. It makes sense as they both are mental health issues.
* From the heatmap on the correlation between music listening and mental health, we cannot conclude what features contribute to improving mental health (this is evaluated by the listeners). But it appears that there may be intriguing connections between various music genres. For instance, individuals who enjoy hip-hop music may have a greater inclination towards listening to rap and R&B, and metal listeners are more likely to want to listen to rock. Also, it looks like depression and anxiety are correlated which makes sense as they both are mental health issues.
* From the pie chart on music effects Vs mental health, we can infer that 75.5% of individuals that listening to music improves Mental Health. 22.1% think music has no effect on mental health and only a negligible 2.4% of people think that music worsens mental health.
* From the plot on Music’s effects on Mental health by favorite genre, it looks like most of the listeners think music improves their mental health.
* From the hypothesis test on the correlation between ‘hours per day’ and age, since the p-value is greater than 0.05, It is said to "failed to reject the null hypothesis". That means the null hypothesis is true i.e., mental health can be improved by listening to more music.
* Based on the linear regression we can conclude that, for the individual age 22, who listens to music: 10 hours per day, Anxiety: 5, Depression: 5, Insomnia:5, OCD: 5, and if the music listening is increased from an average of 2 hours per day to 10 hours per day, mental health improves

What do you feel was missed during the analysis?

I feel the data that I took is less (736 records). I would have taken much more data to analyze the data correctly.

Were there any variables you felt could have helped in the analysis?

Also, I did not take the variable bpm (beats per minute) into consideration as I saw many values were missed in the data. I would have cleaned the bpm column and performed the analysis.

Were there any assumptions you felt were incorrect?

I assumed that listening to music for more than 18 hours is not feasible. Maybe my assumption is wrong. I would have taken 24 hours into consideration as there are 24 hours in a day.

What challenges did you face, what did you not fully understand?

Faced some issues while performing the hypothesis testing and Linear regression. But I was able to resolve them. I did not delve more into Logistic regression.

*Reference for data:*

[*Music and Mental Health EDA | Kaggle*](https://www.kaggle.com/code/yannansu/music-and-mental-health-eda/input?select=mxmh_survey_results.csv)