Date: 2025/8/27

**Summary**

This project investigates the potential relationship between music listening habits and mental health. The analysis was conducted using a dataset sourced from Kaggle, leveraging **MySQL** for data preparation and exploratory analysis, **Python** for statistical correlation, and **Power BI** for visualization. The findings indicate either no correlation or a very weak correlation between music-related factors and the mental health variables.

**Methodology**

**1.1. Data Source & Preparation**

The dataset, sourced from Kaggle, contains survey responses about mental health status and music habits. including indicators for depression, anxiety, OCD, and insomnia.

**1.2. Date Cleaning**

1. Data Inspection: The first step was to inspect the dataset's structure using DESCRIBE to understand the available columns, data types, and get a preliminary view of the data.
2. Outlier Identification: A significant data quality issue was identified in the BPM (Beats Per Minute) column, where an erroneous value exceeding 100 million was present. This issue was handled by:
   * Disabling safe update mode (sql\_safe\_updates = 0).
   * Updating the erroneous record to a value of 0.
   * Re-enabling safe update mode (sql\_safe\_updates = 1).
3. Validation Checks: Logic checks were performed on the Hours per day column to identify any invalid values (e.g., hours greater than 24 or less than 0).
4. Data Type Standardization: The Age and BPM columns were identified as being stored in an incorrect data type. The ALTER TABLE command was used to convert and standardize these columns to the INTEGER (INT) type to facilitate correct numerical analysis.

* Misspelling Standardization: The distinct values in key categorical columns were enumerated to standardize the data. This included:
  + Music genres (Fav genre)
  + Streaming services (Primary streaming service)
  + Listening habits (While working, Instrumentalist, Composer)

2.1. Following the cleaning process, an Exploratory Data Analysis (EDA) was preformed to uncover initial insights getting a deeper analysis.

* Descriptive Statistics:  Statistics like averages, minimums, and maximums were calculated for:
  + Demographic data: Age
  + Listening habits: Hours per day
  + Mental health variables: Anxiety, Depression, Insomnia, and OCD to understand their distribution.
* Aggregation and Grouping: The data were aggregated to answer specific business questions:
  + The popularity of music genres was ranked by counting occurrences.
  + Listening preferences while working were analyzed against favorite genres to identify trends.
  + The average mental health scores were calculated and grouped by music genre to surface potential correlations for further investigation.

**3.Correlations**

Importing Libraries: First step was importing PANDAS library for importing the data, followed up by importing SPEARMANR library to aggregate the correlations.

The results:

* 1. **Listening Hours Per Day Vs. Mental Health Variables:**
* Hours per day and anxiety: 0.09
* Hours per day and depression: 0.13
* Hours per day and OCD: 0.12
* Hours per day and Insomnia: 0.14
  1. **Listening While Working Vs. Mental Health Variables:**

All correlations showed either a weak correlation or invalid P-value

* While working and anxiety: 0.02 (P-value 0.5)
* While working and depression: 0.05 (P-value 0.12)
* While working and OCD: 0.09 (P-value 0.01)
* While working and Insomnia: 0.01 (P-value 0.6)

Note: In the column of “While working” the values were replaced by 1 for “Yes” and 0 for “NO”.

* 1. **BPM (Beat per Minute) Vs. Mental Health Variables:**
* BPM and anxiety: 0.03 (P-value 0.2)
* BPM and depression: 0.06 (P-value 0.8)
* BPM and OCD: 0.01 (P-value 0.6)
* BPM and Insomnia: 0.07 (P-value 0.05)

Note: Null values were replaced by BPM’s average.

* 1. **Age Vs. Mental Health Variables:**
* Ageand anxiety: 0.06 (P-value 0.06)
* Age genreand depression: 0.008 (P-value 0.8)
* Ageand OCD: 0.08 (P-value 0.01)
* Ageand Insomnia: 0.02 (P-value 0.5)

**4. Data visualization**

The aggregated data from MySQL and Python was imported into **Power BI** to create an interactive dashboard. The dashboard includes:

* **Area chart:** showing the effect different music genres on different metal health variables.
* **Line chart:** showing the effect of long listening hours on different mental health variables.
* **pie chart:** representing the percentage of people that are observing mental health effect due to the music effect.
* **Slicer:** allowing breakdown by **age** and **the usage of instruments**.
* **Cards:** Showing the total number of the **survey takers**, the **average of the age**, the **average of the listening hours per day** among the sample, and the average of the mental health status (on a scale from 0 to 10).
* **Refresh Button:** For resting all slicers.

**Tools**

MYSQL was used for data cleaning and exploratory data analysis (EDA).

Python was used for aggregating the correlation.

Power BI was used for visualization and dashboard creation.

Power Point was used for creating the visualization backgroung.

**Insights**

* **Lofi** music has the highest average in the genres effect on mental health variables by **5.6** for **Insomnia**, **6.6** for **Depression**, and **3.4** for OCD.
* NO strong correlation was found between music and any mental health variables and most of them were not significant. The strongest correlation was between **listening hours per day** and **Insomnia** by 0.14.
* Listeners of Gospel music showed a more healthy mental health compared to other genres. The averages of mental health variables were **0.3** for **OCD**, **4.8** for **anxiety**, **2.6** for **depression**, and **5.3** for **Insomnia**. And these results could be attributed to Gospel music’s spiritual load.

**Conclusion**

This analysis investigated the correlation between music listening habits and mental health. While the survey data captured showed reported effects on mood, the quantitative analysis revealed **no statistically significant strong correlations.** The strongest relationship identified was a very weak positive correlation between listening hours and insomnia (r=0.14). Consequently, we conclude that while music may exert a noticeable short-term influence on mood, its long-term effect on mental health conditions appears to be weak to non-existent.