Data Handling

Here’s an account of what transformations and procedures I did to/with the data. This will help understand the R files (which themselves have comments explaining what’s happening).

# Transformations

 **Renaming Columns**: I simplified and standardized the column names to make them easier to reference and interpret during analysis. This involved shortening names and ensuring they followed a consistent format.

 **Re-coding Responses for Yes/No and Frequency Questions**:

* For the question about listening to background music, I converted "Yes" and "No" answers to numerical values, with 1 for "Yes" and 0 for "No."
* For questions on the frequency of listening to music during complex and simple tasks, I transformed textual answers (like "Never" or "Always") into a scale from 1 to 5, allowing me to quantify listening habits.

 **Transforming Personality-related Responses**:

* I recoded personality questions from textual responses (e.g., "Disagree strongly," "Agree") to numeric scores. For most questions, higher scores indicate a stronger trait (e.g., more talkative).
* Some questions, however, were reverse-coded (e.g., "I see myself as someone who is reserved"). For these, I reversed the scale so that all questions aligned, with higher values consistently representing higher levels of extraversion.

 **Calculating an Extraversion Score**: I computed an overall extraversion score for each participant by averaging the scores from multiple personality-related questions. This provided a single measure of extraversion.

 **Classifying Participants by Personality Type**: Based on the extraversion score, I classified each participant as either an "Extrovert" or "Introvert" for further analysis, using a cutoff score of 2.5.

 **Processing Age and Creating Age Brackets**:

* I extracted and converted the age values to ensure they were numeric.
* I then grouped participants into age brackets, assigning a rank to each bracket (from youngest to oldest). This allowed me to analyze data by different age groups.

 **Creating Binary Columns for Music Type Preferences**:

* For questions about the types of music listened to during different tasks (like "Vocal" or "Instrumental"), I created separate columns for each music type. In each column, a 1 indicated that the participant selected that type, while a 0 indicated they did not.
* I did this separately for complex, simple, and other tasks to capture specific preferences across task types.

 **Creating Binary Columns for Reasons for Listening to Music**:

* Similarly, for questions about why participants listened to music (e.g., "To mask other background noise" or "To help me concentrate"), I created binary columns. Each column contained a 1 if the participant selected that reason and a 0 if they did not.
* I organized these columns by task type (complex and simple).

 **Creating Binary Columns for Reasons for Avoiding Music**:

* For questions about why participants avoided listening to music (e.g., "Music distracts me" or "Music makes me anxious"), I again created binary columns to capture whether each reason was selected or not, organized by task type.

 **Creating Binary Columns for Effects of Listening to Music Over Time**:

* For questions about the effects of listening to music over extended periods (e.g., "It keeps me alert" or "It makes me tired"), I created binary columns to reflect whether participants experienced each effect.

 **Saving the Transformed Data**: Finally, I saved the modified dataset to a new Excel file to ensure all transformations were preserved and ready for statistical analysis.

# Demographic & Descriptive Statistics

 **Calculate Summary Statistics for Listening Frequency**:

* I created a function, calculate\_descriptive\_stats, to compute summary statistics—Mean, Standard Deviation (SD), Median, and Interquartile Range (IQR)—for specified columns.
* I applied this function to obtain these statistics for the listening frequency during complex and simple tasks. This provides an overview of how often participants listen to background music in different contexts.

 **Frequency Counts for Music Type Preferences**:

* I created a function, calculate\_music\_type\_counts, to calculate the frequency counts for each type of music (e.g., Vocal, Instrumental) that participants listen to during complex and simple tasks. This step helped quantify the types of music preferences across different task types.
* I used this function to generate separate tables for complex and simple tasks, allowing a comparison of music preferences based on task complexity.

 **Gender-Based Demographic Statistics**:

* I calculated demographic statistics based on gender, including the count, percentage representation, mean age, median age, SD, IQR, and the rate of extraversion for each gender group.
* To provide an overall view, I added a row for "Total" participants, summarizing the statistics across the entire dataset.
* This demographic summary gives insights into the age and personality type distribution across gender groups.

 **Reasons for Listening and Avoiding Background Music**:

* I created a function, calculate\_reasons, to calculate the percentage of participants selecting specific reasons for listening to or avoiding background music.
* Using this function, I calculated percentages for reasons associated with both complex and simple tasks. Reasons include masking background noise, aiding concentration, setting the mood, and more.
* These calculations give insight into participants' motivations and hesitations for listening to music, contextualized by task type.

 **Frequency Counts for Text-Based Categories**:

* I defined another function, calculate\_text\_categories, to calculate frequency counts and percentages for specific text-based categories. These include categories for various open-ended questions (e.g., Q3, Q5, Q7, Q9, Q11).
* I applied this function to organize and calculate frequency counts for text responses across multiple questions, helping to categorize the qualitative data in a structured format.