Task 1: Extracting and Analyzing CSV Data

Approach:

* Reading a Single Column (FR1): I used the open function to grab a single column from a CSV file. I used readline for the header and readlines for the data. The result is a list with both the header and column values, raising an 'out of index' error for an invalid column index.
* Reading CSV Data into Memory (FR2): I created a loop using the read\_a\_column function, continuing until an IndexError showed the end of columns. This produced a dictionary with column headers as keys and corresponding values as lists.
* Calculating Kendall Tau Correlation (FR3): I made a robust function handling exceptions for lists of different lengths. Nested loops compared pairs of elements, calculating concordant and discordant pairs.
* Generating Kendall Tau Correlations (FR4): I introduced a function for computing Kendall Tau Correlation Coefficients across all possible column pairs, minimizing redundant calculations and storing results in a list.
* Printing a Custom Table (FR5): Creating a custom table was challenging. Nested loops structured rows and columns, and string-building techniques enhanced formatting with careful attention to borders and lines.

Reflection:

* *Thought Process:* It started challenging but evolved into a streamlined process with efficient functions.
* *Strengths:* Efficient loop-based solutions; comprehensive error handling.
* *Weaknesses:* Manual data processing may lack efficiency for larger datasets.
* *Improvements:* Investigate alternative libraries for optimized CSV handling.

Task 2: Exploring Mental Health Data

Approach:

* Merging Data (FR6): I seamlessly merged data from two CSV files using the pandas library. This involved reading and merging based on specified columns, highlighting the efficiency of using appropriate libraries.
* Exploring the Dataset (FR7): Thorough exploration unfolded using the describe and info methods. I found intriguing patterns in a correlation table for all countries, with a focus on high correlations between 'eating disorders' and 'bipolar disorder.' I conducted a granular investigation into individual countries over time, enriching the analysis.
* Detecting and Removing Outliers (FR8): I navigated the intricacies of outlier detection using the Z-score method. Prudent grouping based on correlation criteria effectively removed potential outliers, contributing to a more robust analysis.
* Defining a Hypothesis (FR9): I formulated a hypothesis expecting a high correlation between 'eating disorders' and 'bipolar disorder,' grounded in observed patterns and data preprocessing steps.
* Testing the Hypothesis (FR10): I executed a meticulous one-sample t-test to scrutinize if the correlation was at least greater than 0.4. The findings cautioned against making overarching statements about correlations due to the inherent diversity within the dataset.

Reflection:

* *Thought Process:* Streamlined with the use of pandas; correlation analysis provided valuable insights.
* *Strengths:* Efficient use of libraries; detailed outlier analysis.
* *Weaknesses:* Hypothesis formulation could benefit from more nuanced considerations.
* *Improvements:* Explore additional hypothesis testing methods for a more comprehensive analysis.

Conclusion: This exploration emphasized methodical approaches in data science, highlighting the importance of evidence-backed insights, judicious library use, and the continuous quest for nuanced understanding.