Task 1:

In Task 1, I was challenged to develop a set of functions to work with CSV data using only Python built-in functions and data structures. The task consisted of five functional requirements (FRs), each with a different objective and difficulty level.

**FR1:** The first requirement was to develop a function to read a single specified column of data from a CSV file and return the column name and a list of data values. This was the most challenging part for me, as I was used to using the pandas library for this task. I had to review the lectures and learn how to use the open, readline, and split functions to read and parse the CSV data from a file. I also had to use the len and range functions to iterate over the lines and columns of the data. I used the append function to add the data values to a list. This task took me a lot of time to complete, but it also helped me refresh my Python basics and learn how to work with files without external libraries.

**FR2**: The second requirement was to develop a function to read all columns of data from a CSV file and store them in a dictionary, where each key-value pair corresponds to a column name and a list of data values. This was a relatively easy task, as I could reuse and modify the code from FR1. I used a nested for loop to iterate over the lines and columns of the data, excluding the first line that contained the column names. I used an if statement to check if the current column name was already in the dictionary, and if not, I created a new key with the column name and an empty list as the value. I then used the append function to add the data values to the corresponding list in the dictionary.

**FR3**: The third requirement was to develop a function to calculate the Kendall Tau Rank Correlation Coefficient for two lists of data, which measures the degree of similarity between the rankings of the data values in the lists. This was a completely new concept for me, so I had to do some research on the internet to understand what it was and how to calculate it. I learned that the Kendall Tau Rank Correlation Coefficient is a non-parametric statistic that ranges from -1 to 1, where -1 indicates a perfect inverse relationship, 0 indicates no relationship, and 1 indicates a perfect positive relationship. The formula to calculate the coefficient is:

where C represents the number of concordant pairs and D represents the number of discordant pairs. A pair of data values is concordant if they have the same relative order in both lists, and discordant if they have the opposite relative order. To implement this formula, I used a nested for loop to iterate over all possible pairs of numbers in the lists. I used a count variable to keep track of the number of concordant and discordant pairs. I also used a try and except statement to handle the case where the lists were of unequal length and raise an error. I then returned the coefficient value as the output of the function.

**FR4**: The fourth requirement was to develop a function to generate a list of tuples containing the Kendall Tau Rank Correlation Coefficients for every pair of columns in the CSV data file. This was a straightforward task, as I could use the function from FR3 as a helper function. I used the dictionary from FR2 as the input parameter for this function. I used a nested for loop to iterate over all possible pairs of keys in the dictionary, which represented the column names. I then used the function from FR3 to calculate the coefficient for the corresponding lists of data values in the dictionary. I used the append function to add a tuple with the column names and the coefficient value to a list. I then returned the list as the output of the function.

**FR5**: The fifth and final requirement was to develop a function to print a custom table showing the Kendall Tau Rank Correlation Coefficients for a subset of the column pairs generated in FR4. This was a new and interesting task for me, as I had to learn how to use the format and join functions to create and print a formatted table. I used a function to create the table, and I came up with a unique approach for the design. I decided to use the asterisk symbol to create the table borders, with the length determined by the maximum length of the column names, multiplied by three to accommodate the three output columns. I used the format function to centre the output values within the table cells, and the join function to concatenate the table rows with newline characters. I also used the print function to display the table on the screen. This task allowed me to present the data in a clear and organized manner.

This project was a great learning experience for me, as it challenged me to use only Python built-in functions and data structures to perform various operations on CSV data. I learned how to read and parse CSV data from a file, how to store and manipulate data in lists and dictionaries, how to calculate and interpret the Kendall Tau Rank Correlation Coefficient, and how to create and print a formatted table. I also improved my Python skills and my problem-solving abilities.