Report reflective

**FQ1**：

At first,I tried to use function “def read\_single\_column” to read a single column directly. But I got something wrong to read my dataset. So I import the csv file first then define my dataset as ‘fn’. The rest of the code is written in ‘fn’. Then I used function ‘ read\_single\_column’ from ‘ to finish this task.

**FQ2**:

Once I read this question,I realized that I can use a function that reads all the data in a CSV file and stores them as a list. Then I used function ’csv.DictReader’ to read the csv file and storing each line of list as a dictionary.

**FQ3**:

Without using pandas or numpy to caculate Kendall Tau Correlation Coefficient. I have to know the formula of it is τ=(C-D)/^(C+D+T)\*(C+D+U). Then I can use the formula to finish this task step by step.

**FQ4**:

The ‘calculate\_kendall\_taus’ function makes a data dictionary, calculates the Kendall Tau correlation coefficient between all variable pairs in this ‘fn’, and returns the result stored in a dictionary. The keys of this dictionary are pairs of variables and the values are the corresponding correlation coefficients. I used the’ kendall\_tau function’ to calculate the correlation coefficient between every two lists.

**FQ5**:

In order to print a custom table to show the selected data in the data structure, I define a function that accepts a data dictionary and a list of variables to select as input parameters. The function will then print out a table containing the data for these selected variables. In this custom tale the Mean (Mean), Median (Median), maximum (Max), and minimum (Min) values are calculated and showed for each selected variable.

**FQ6**:

By using ‘panda’, it is much easier, just need ‘read’ and ‘merge’ function can deal with that.

**FQ7**:

When I saw dataset 2a, I find that I am interested in exploring the relationship with Schizophrenia and Bipolar disorder. Because one of my friends has trouble in this two mental problems. So I choose using linear regression and pearson correlation coefficient to figure out their relationship. I used machine learning to finish this task and show the visualization.

**FQ8**:

To finish this task, I need to detect and remove outliers from the dataset before performing linear regression analysis. These are the steps and Verification procedure: First, use a scatter visually examine outliers in the data. Then calculating statistical indicators such as mean, median and standard difference of dataset. Thirdly,define outliers :by using the IQR (interquartile) method, which considers data points below the first quartile (Q1) minus 1.5 times the IQR or above the third quartile (Q3) plus 1.5 times the IQR as outliers. Then Remove outliers from the dataset according to defined outliers.finally draw the chart again to confirm that the outliers in the data have been cleared.

**FQ9**:

To test pattern or trend, I decided to use t-test. I used Null hypothesis (H0) and Alternative hypothesis (H1) which are influence by slope to judge whether these two variables has relationship. When slope goes to 0,means that the independent variable has no effect on the dependent variable. When slope doesn’t goes to 0,means that the independent variable has a significant effect on the dependent variable.

**FQ10**:

The significance level (α) is usually set to 0.05, which means that willing to accept a 5% risk of the null hypothesis (H0). then compare the P-value to the significance level: If the P-value is less than the significance level (p < α), the H0 is rejected. That means it has enough evidence that there is a significant relationship between the variables. If the P-value is greater than or equal to the significance level (p ≥ α), the H0 cannot be rejected, which means that there is not enough evidence for a significant relationship between the variables. Then I print a summary of this model, which contains details of t statistics, standard errors, p-values, and so on for each coefficient. after traverse each coefficient, compare the p-value to the significance level (0.05), and print whether to reject the null hypothesis based on the comparison results.