Errors

1. Data for SPB and BMI are both recorded in the same column, although they represent different information. The SPB and BMI data is formatted as characters instead of numerics.

2. In the ‘Age’ column, the data is represented as characters instead of numerics and includes extraneous characters such as ‘years’ which are irrelevant. there are inconsistencies in the format of the data in the cells. Due to the characters included in some cells, this column is recognized as ‘For example, ‘47 years’ must be changed to ‘47’.

3. In column number 7, The ‘United States’ is not the right column heading for this data. The information is also unnecessary.

4. In column numbers 8 and 9, all cells in these columns are empty. These columns are also unnecessary.

5. In column numbers 10 to 16, most of the cells in these columns are empty. These columns are also unnecessary.

The strategy used in this lab (Python)

1. I separated the ‘SBP & BMI” column into two seperate columns ‘SBP’ and ‘BMI ’so that their data is kept separately using the *str.split* function. Using the *pd.to\_numeric* function*,*  I changed the type of both columns to numeric. The original ‘SBP & BMI’ column was deleted using *dataframe.drop* function after performing this command.

2. I separated the ‘Age” column into two separate columns ‘Age’’and “No need” columns using the str.split function. Since “No need” columns only included blank or ‘years’ which is not necessary, I deleted that column using dataframe.drop function. Using the *pd.to\_numeric* function*,*  I changed the type of ‘Age’ column to numeric. Since the new ‘Age’ column replaced the old ‘Age’ column, there is no need to run the command to delete the column.

3. I checked the missing value of columns using *dataframe.info()* which returns the number of missing values in each column. If the return value is ‘0 non-null’ or ‘2 non-null’, I excluded the column. The column numbers 8 to 16 (original column numbers before performing wrangling) were deleted.

4. Since the column ‘United States’ contains unnecessary information, I deleted that column using the *dataframe.drop* function. By setting *inplace=True*, I didn’t have to assign a new column to the original column.