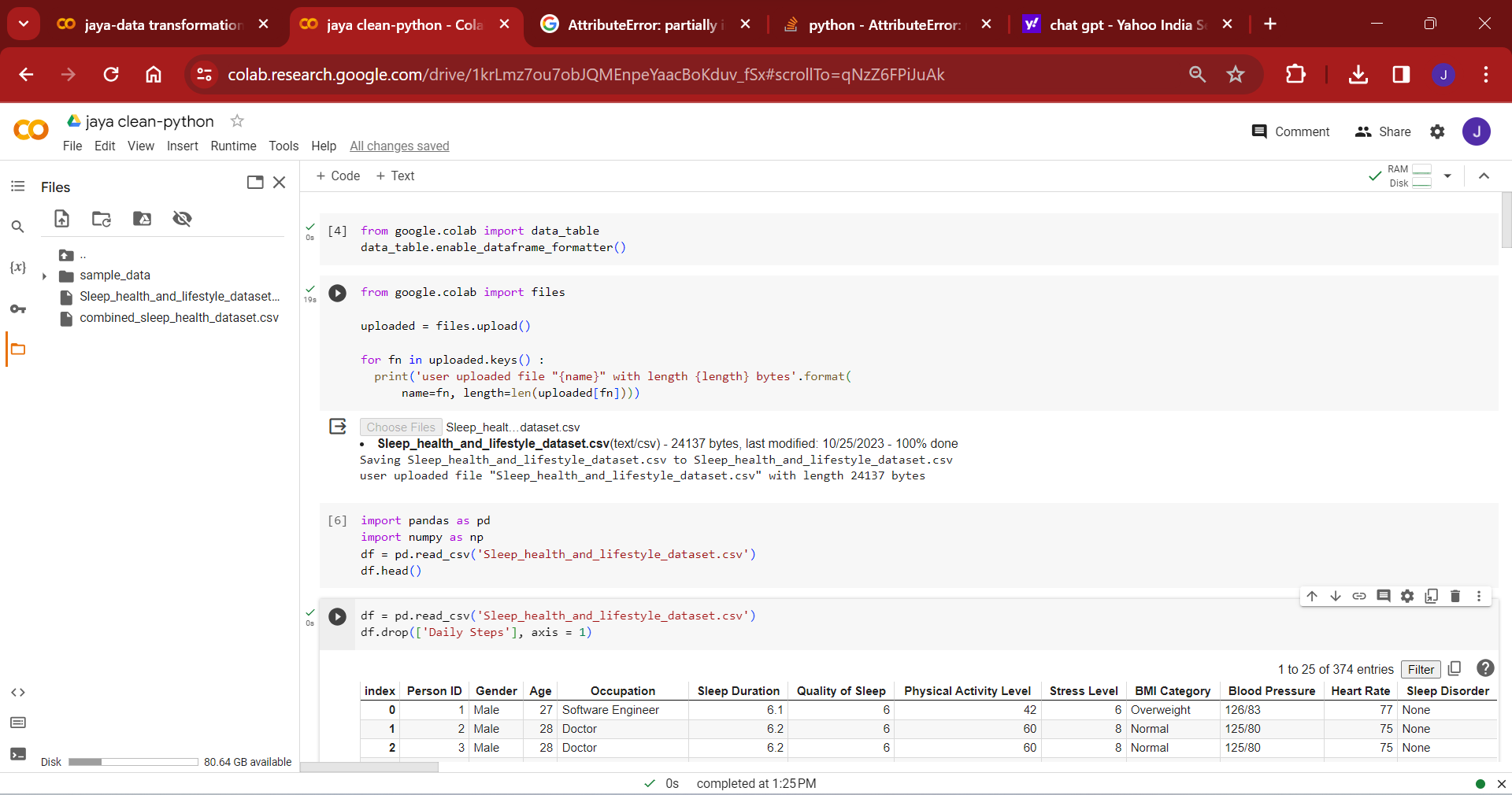
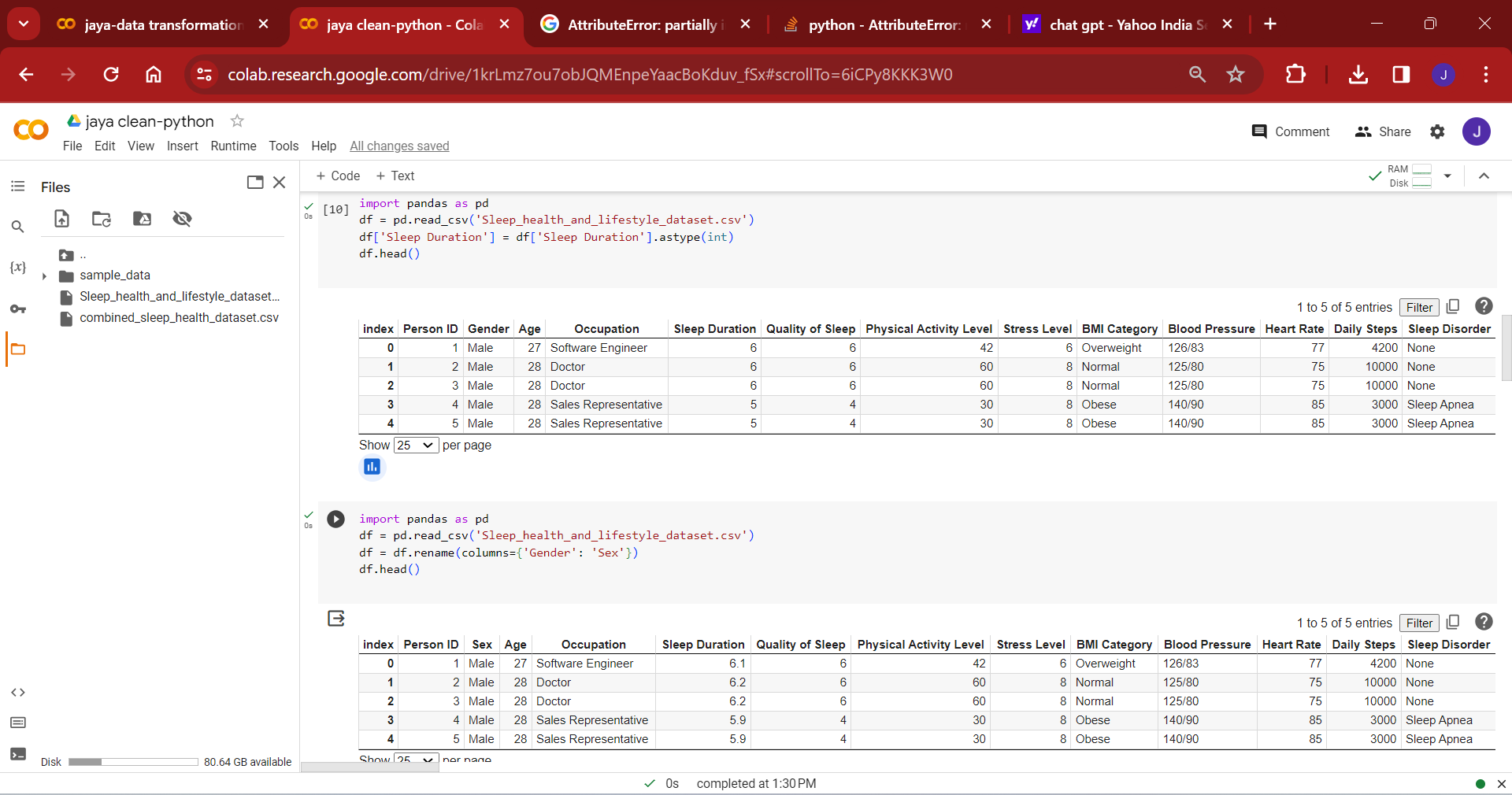
CLEANING DATA AND DATA TRANSFORMATION USING PYTHON

SCREENSHOTS OF FUNCTIONS AND THEIR OUTPUTS:

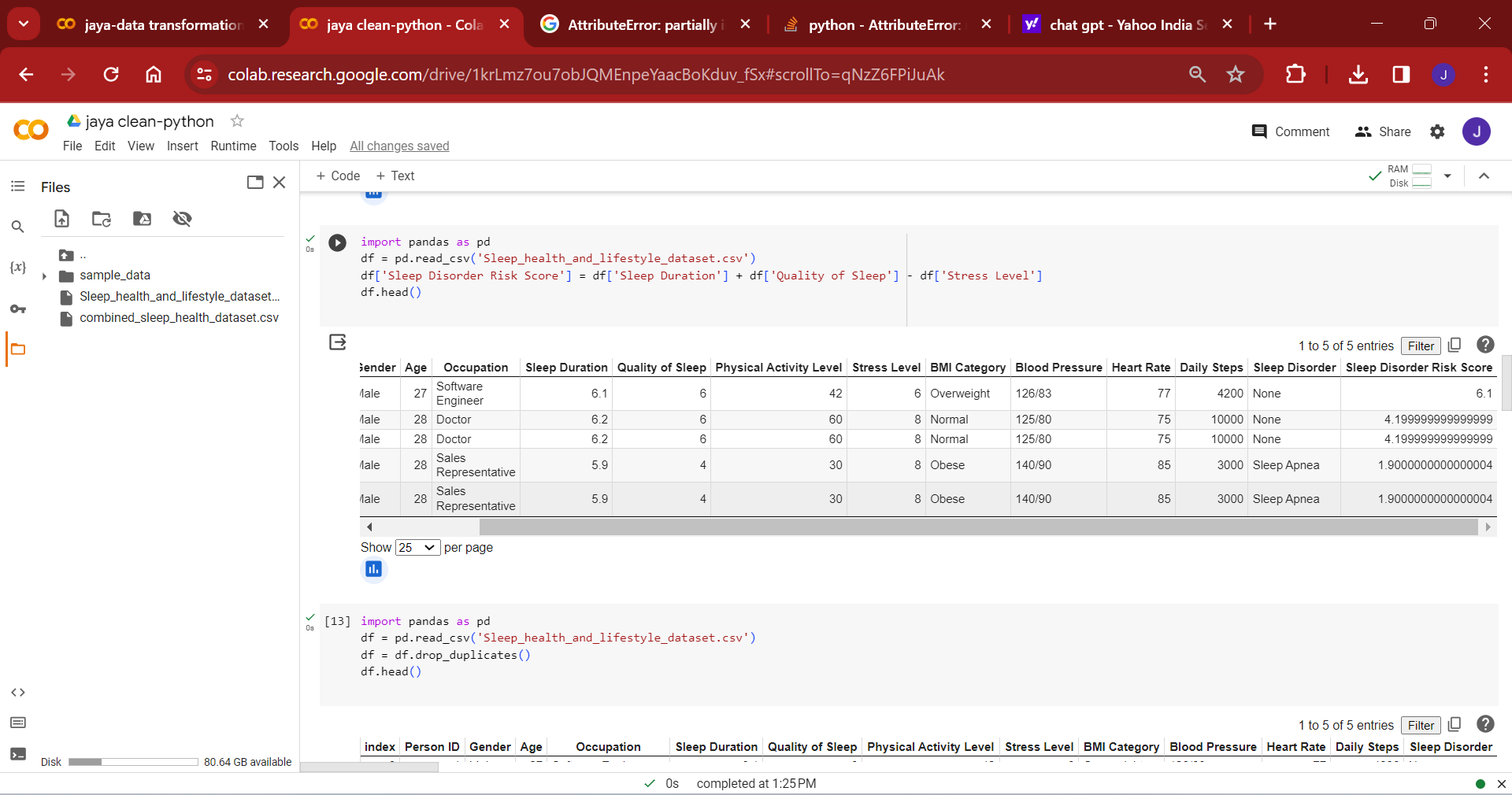
This screenshot shows that how I have loaded the dataset into the data frame. The second function in the image shows that I have eliminated the 'daily steps' column. That column, in my opinion, is unnecessary because there is a column called 'physical activity level,' which implies that daily steps are included regardless.



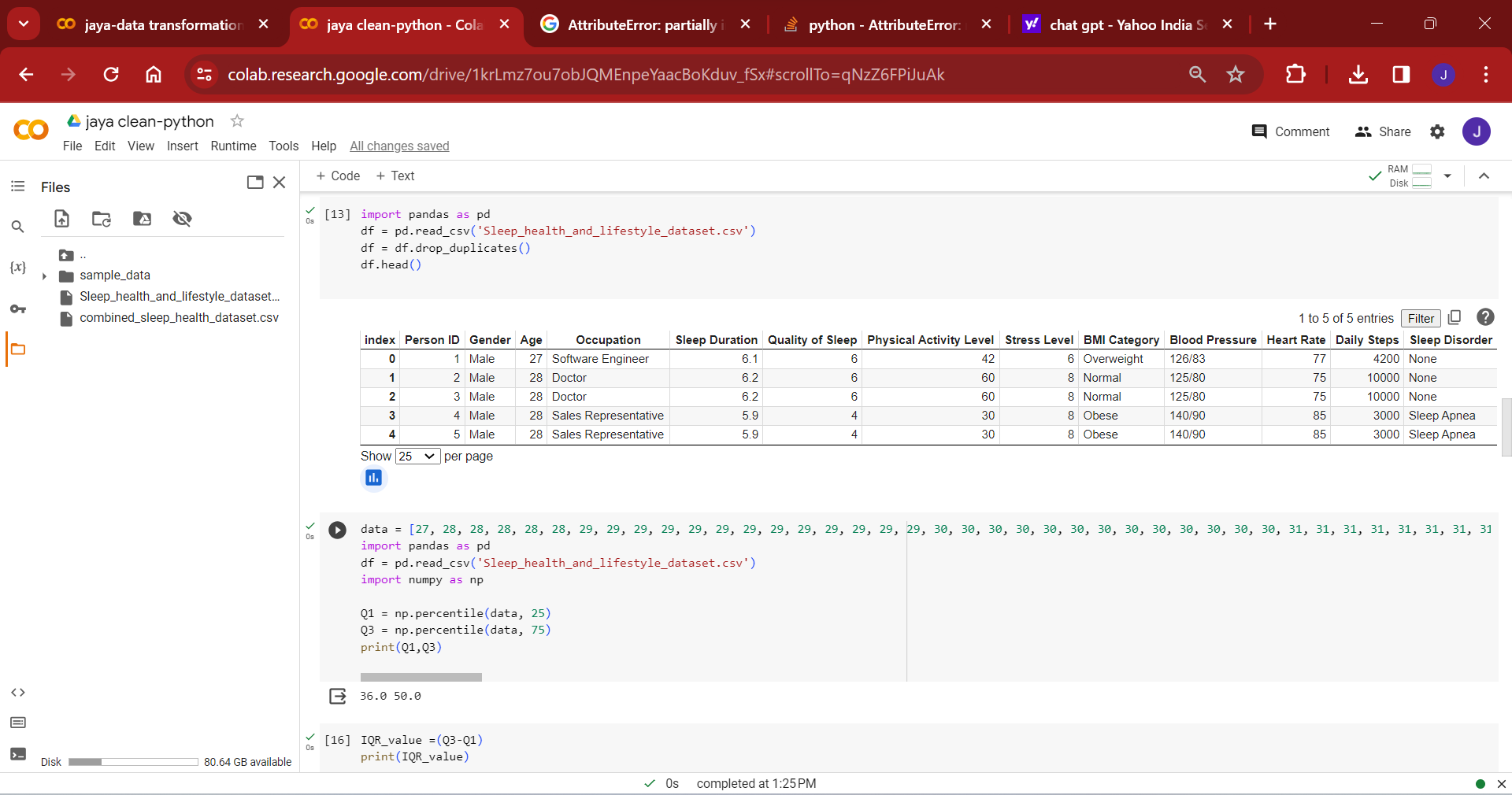
This picture shows that data transformation and renaming columns. I have changed the data type of the ‘sleep duration’ column float into integer. Next function is to rename the ‘gender’ column as ‘sex’.



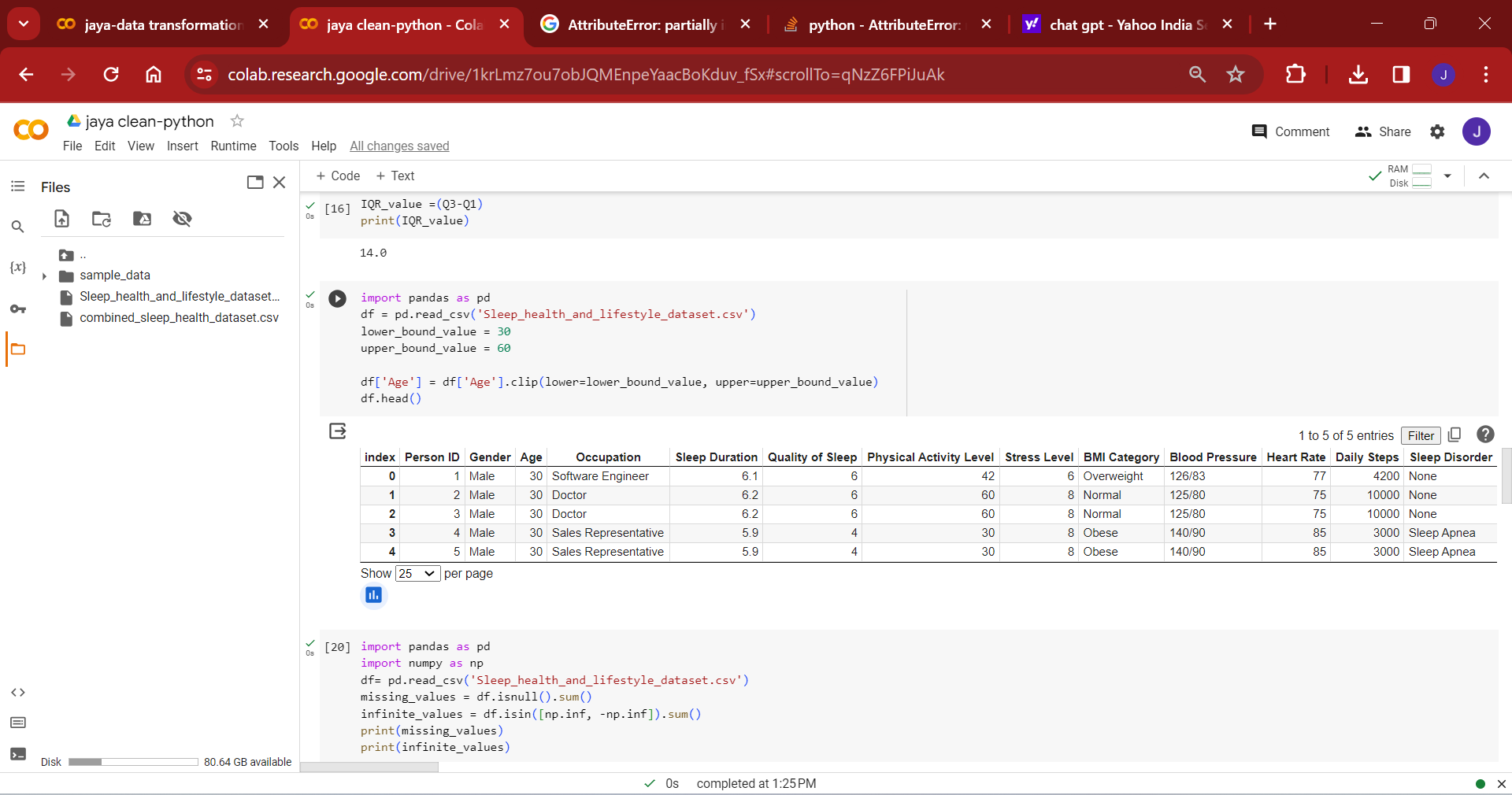
This screenshot shows creation of calculation field. By calculating the ‘sleep duration’, ‘quality of sleep’, and ‘stress level’, I created a field called ‘sleep disorder risk score’.



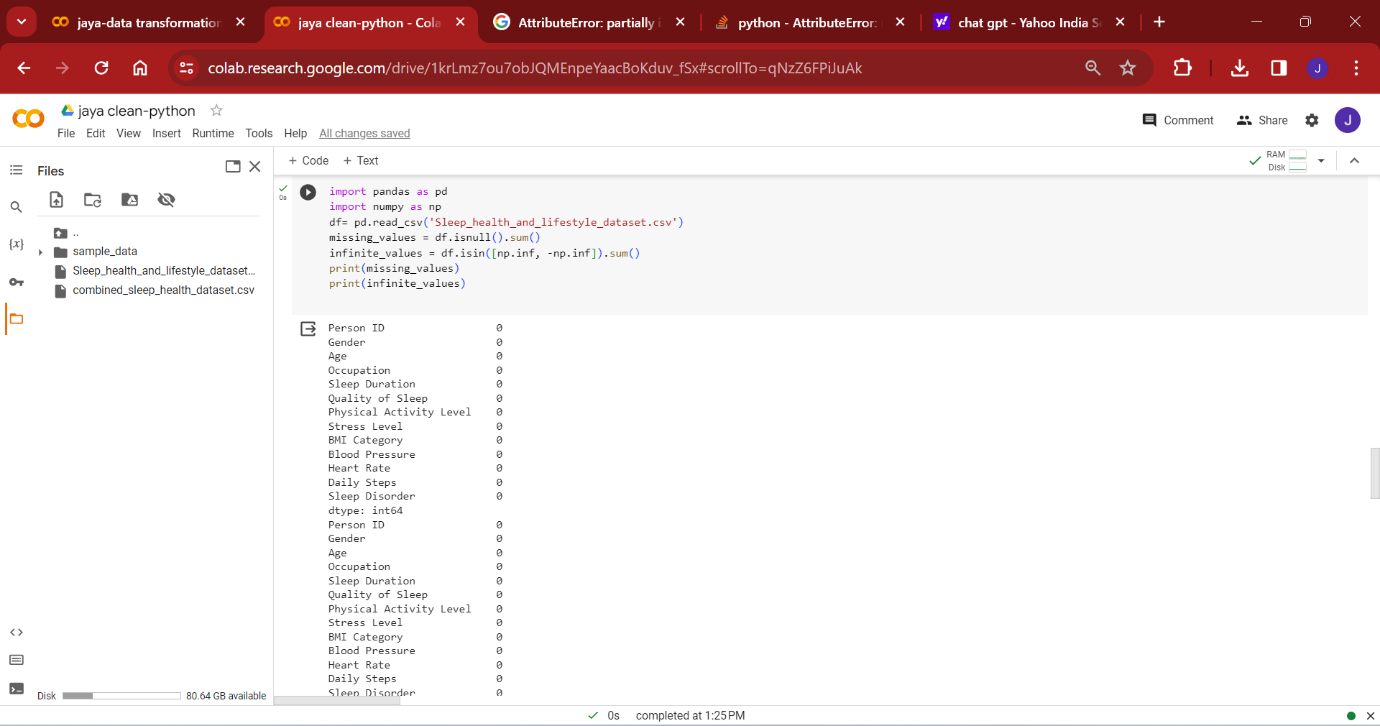
I performed dropping of duplicates. Because there are no duplicates in my dataset, the outcome is unchanged when I run this code. Next function in the picture identifies the outlier and IQR value in the ‘age’ column.



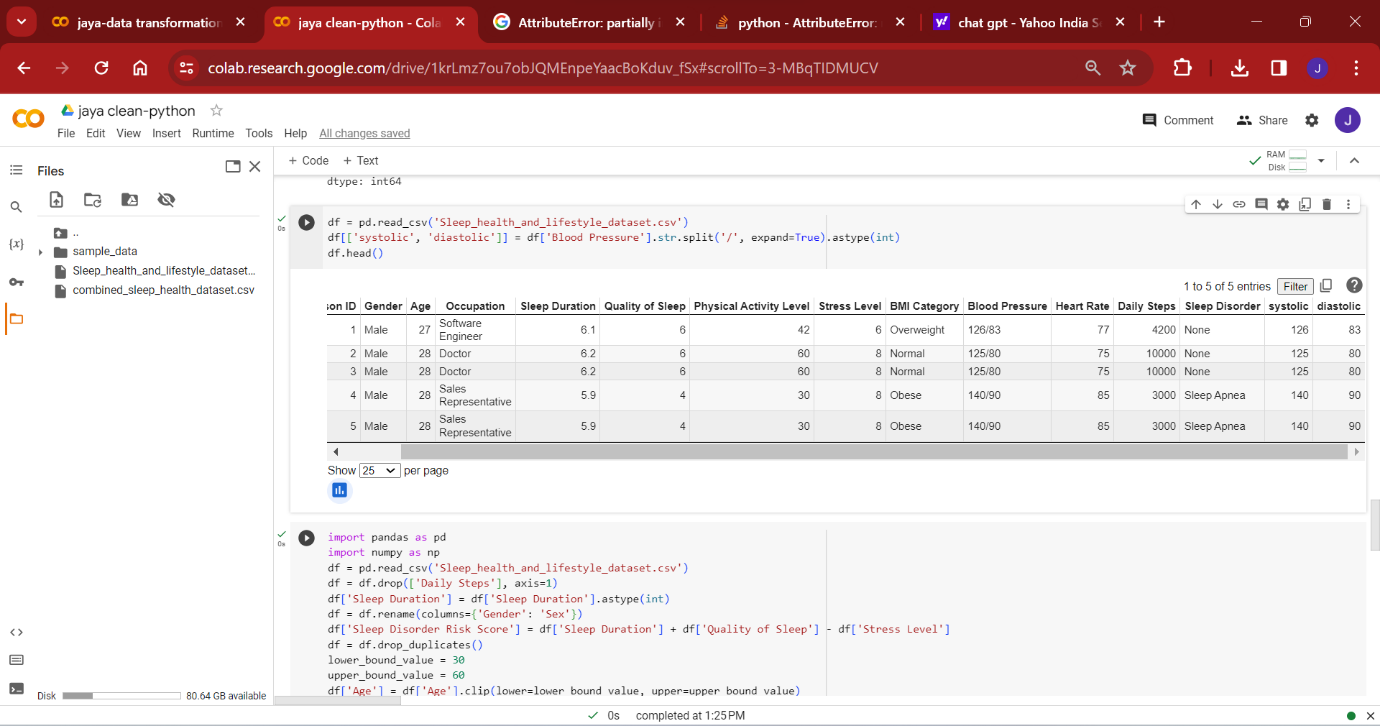
In this picture, I have performed clipping approach in the ‘age’ column. For this, I want only mid-aged persons from 30 to 60. So, I have set the lower and upper bound value as 30 and 60 respectively. Accordingly, I got the output that holds only given limit.



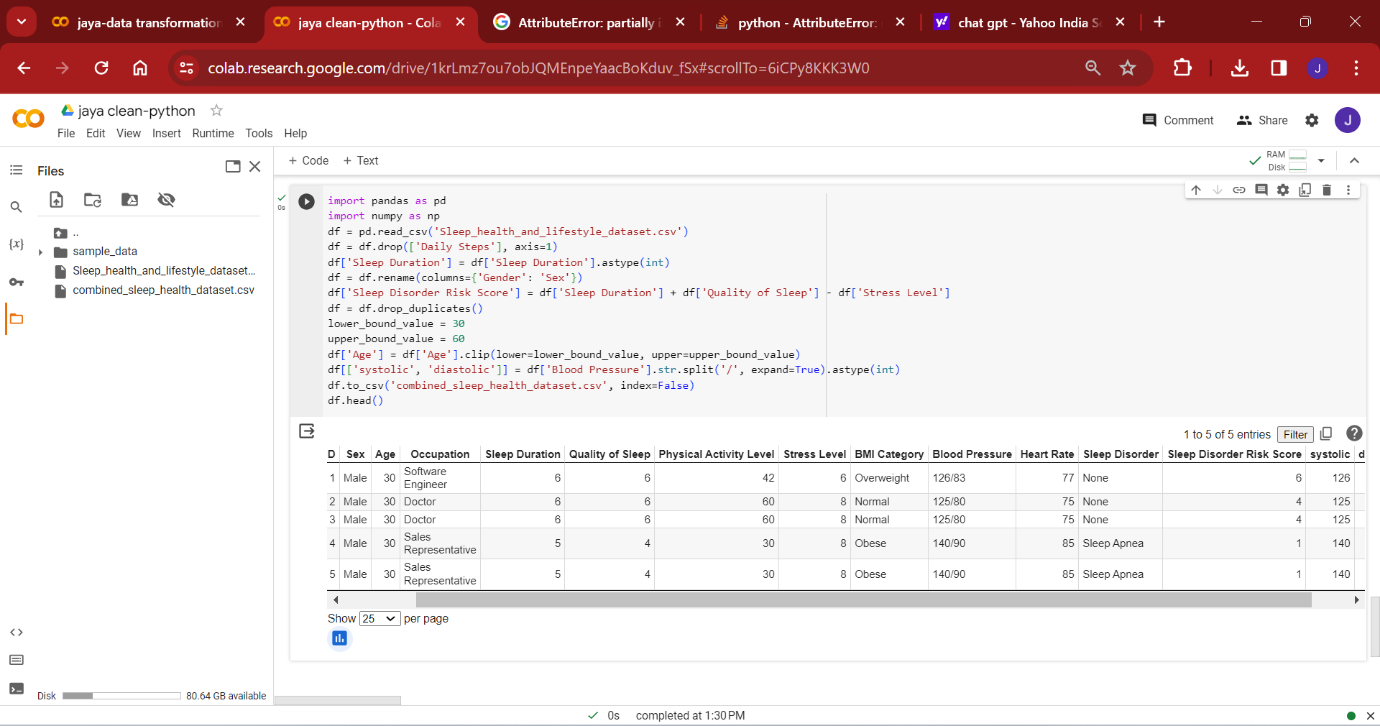
This function I performed to deal with null or infinite values in my dataset. When I run the code, I got zeros in the output as there are no null or infinite value in my dataset.



This function I provided to show other transformation like splitting. This code split the column ‘blood pressure’ into two different columns named ‘systolic’ and ‘diastolic’.



This function I gave to combine all the outputs that have been generated separately and save the csv file.



APPENDIX:

Rename columns- <https://youtu.be/6eJlRVygEMk>

Drop columns-<https://youtu.be/oeGk7hrXMlY>

Change column data type-<https://youtu.be/evKYySLSzyk>

Drop duplicates-<https://youtu.be/JU1EhK89ThE>

Data cleaning in pandas-<https://youtu.be/bDhvCp3_lYw>

Chat GPT

Prompt:

1. Appropriate code to combine all the codes to get ultimate cleaned dataset.

2. Code to save the final dataset.