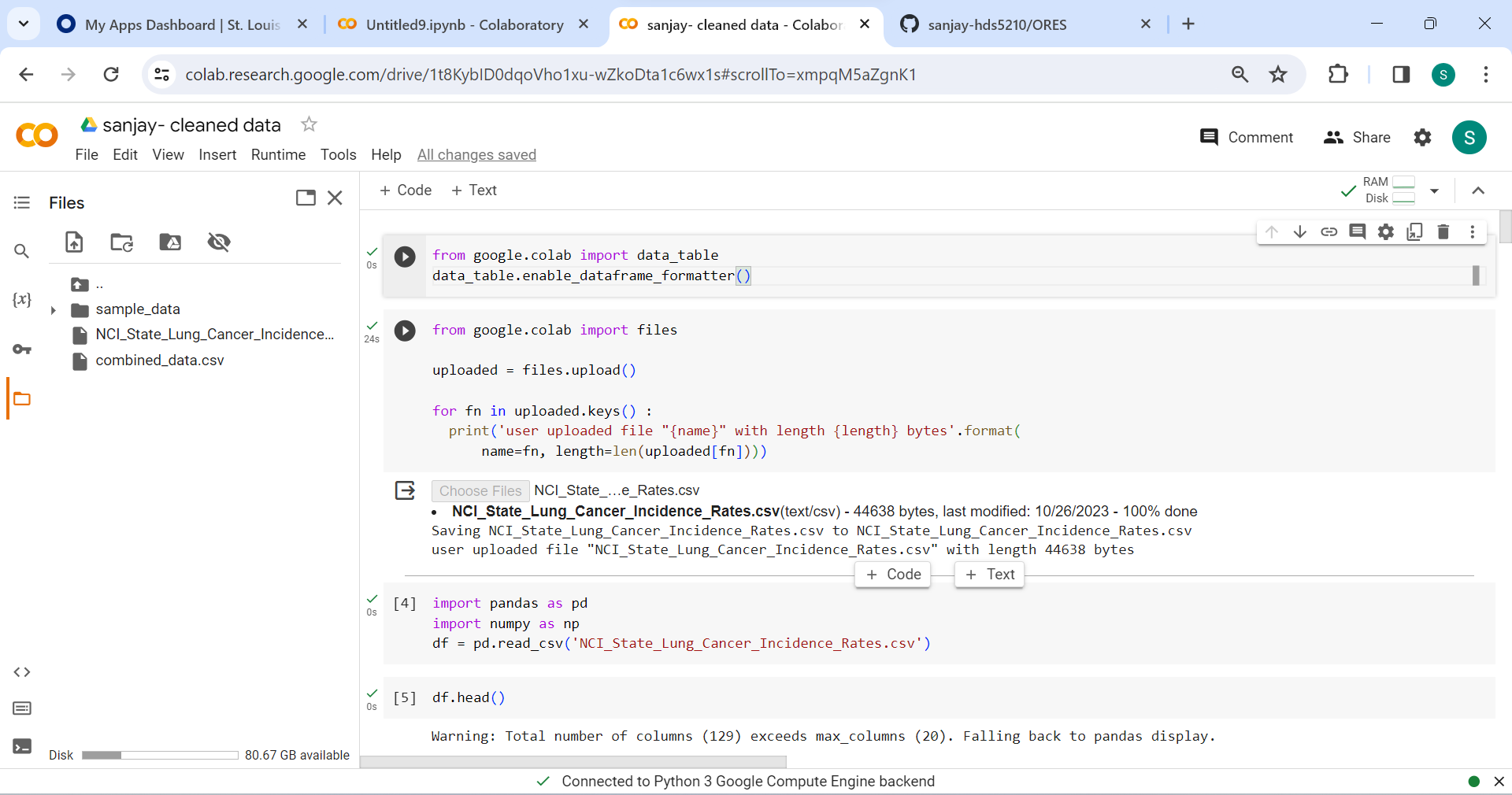
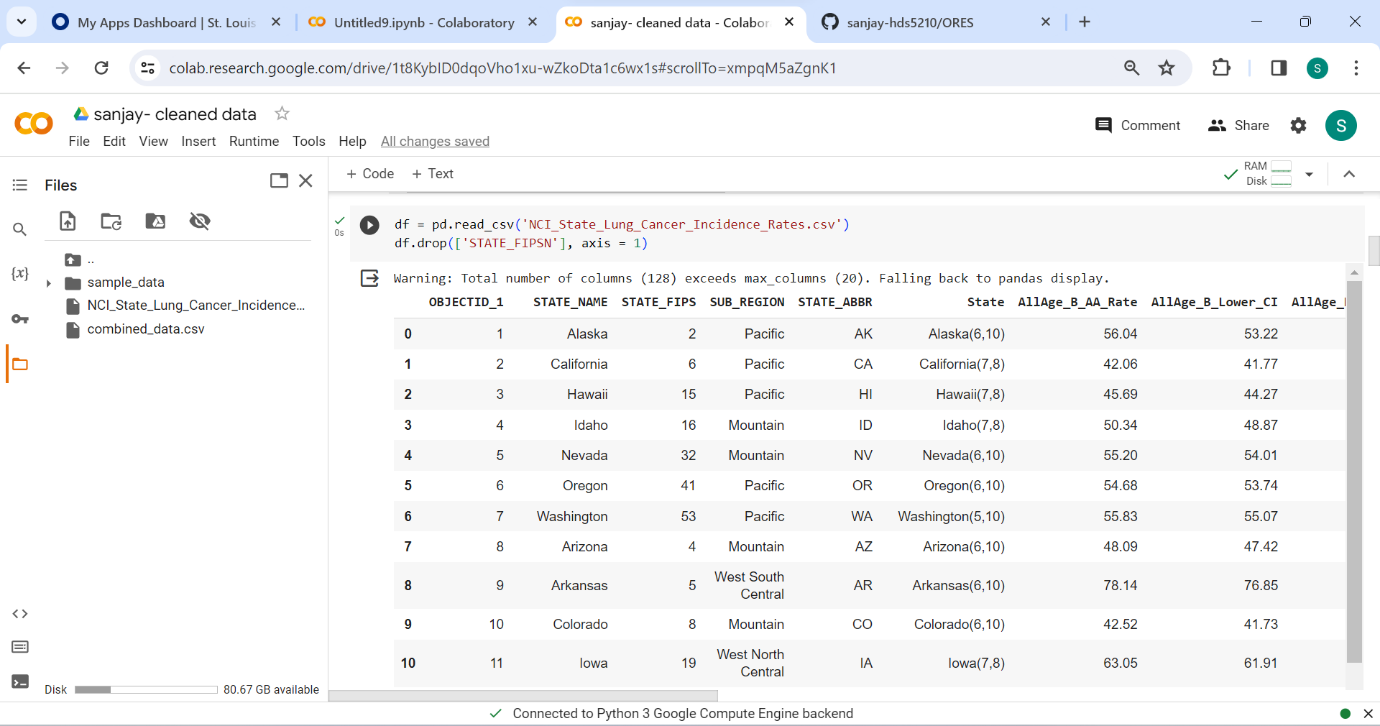
DATA TRANSFORMATION WITH PYTHON- CLEANING DATA

STEPWISE SCREENSHOT [FUNCTIONS & OUTPUTS]

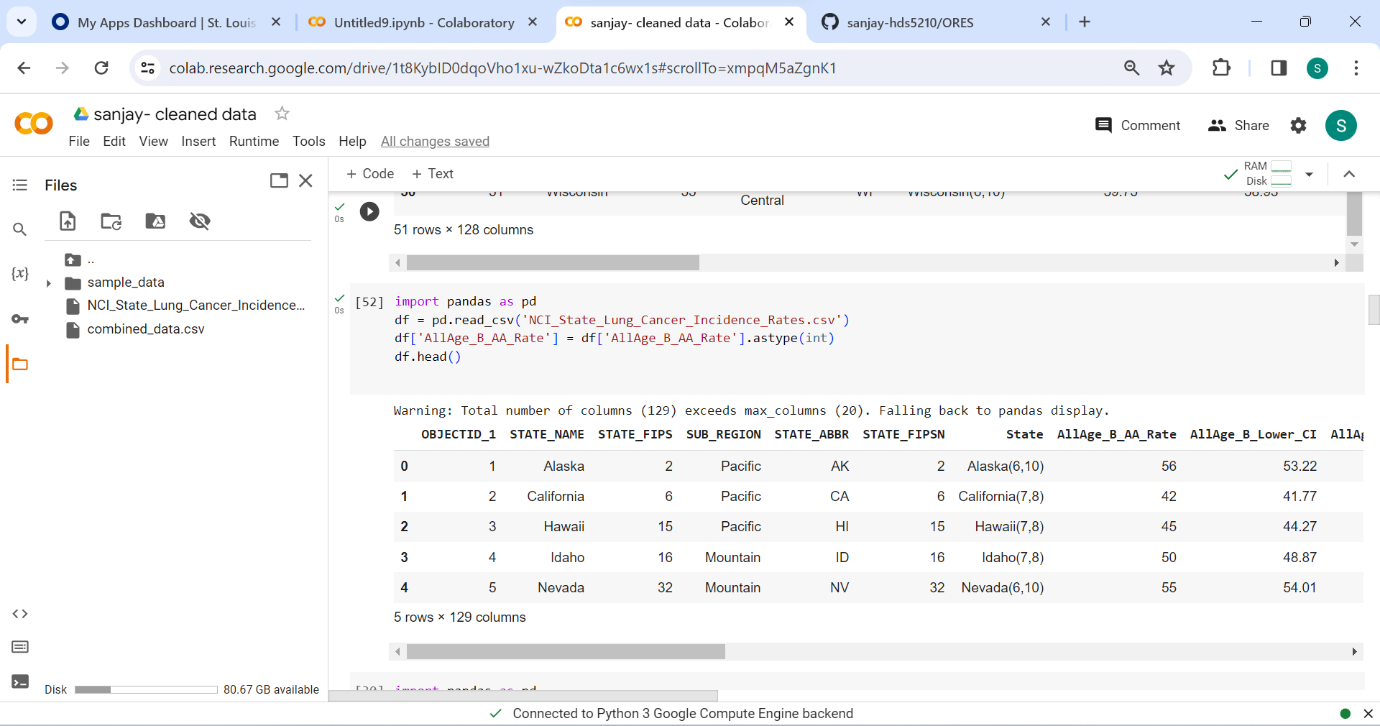
Loaded dataset file using this function and created a data frame.



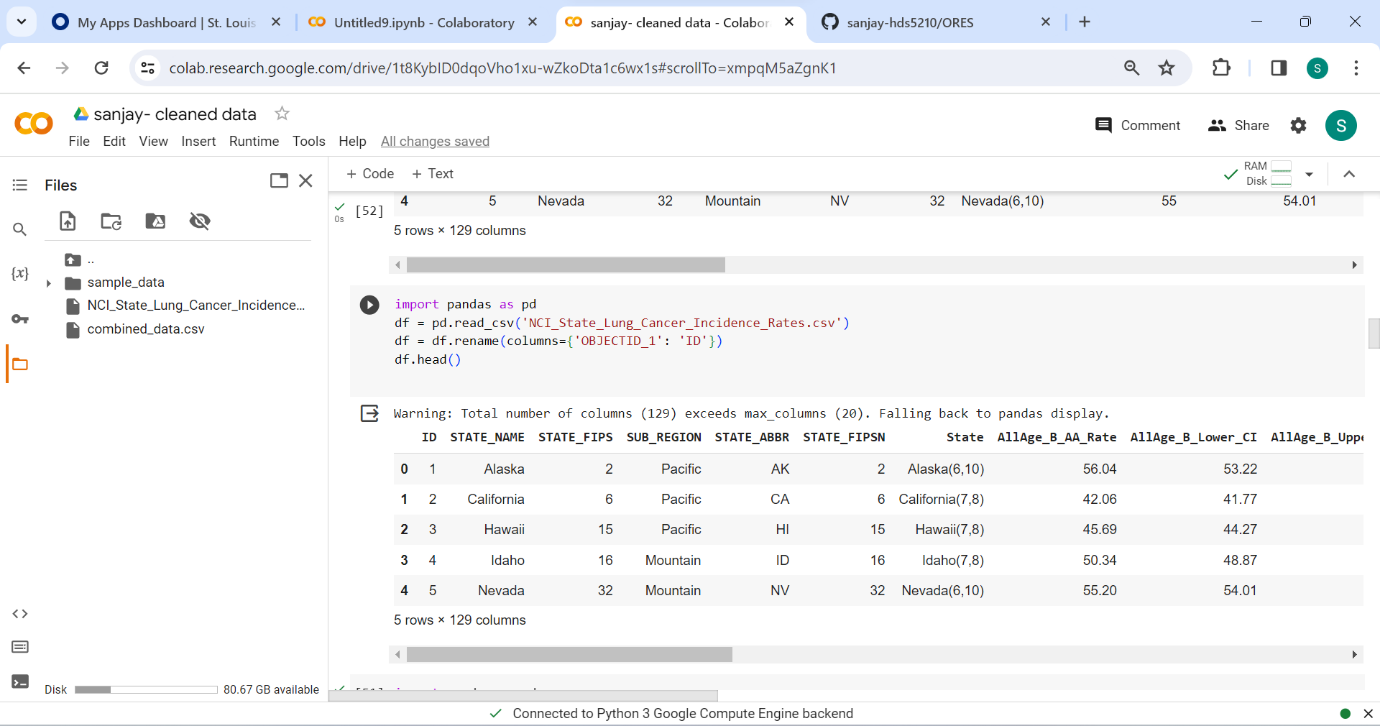
Dropped the unnecessary column. I dropped the column STATE\_FIPSN because already a column named STATE\_FIPS depicts the same value in the column STATE\_FIPSN.



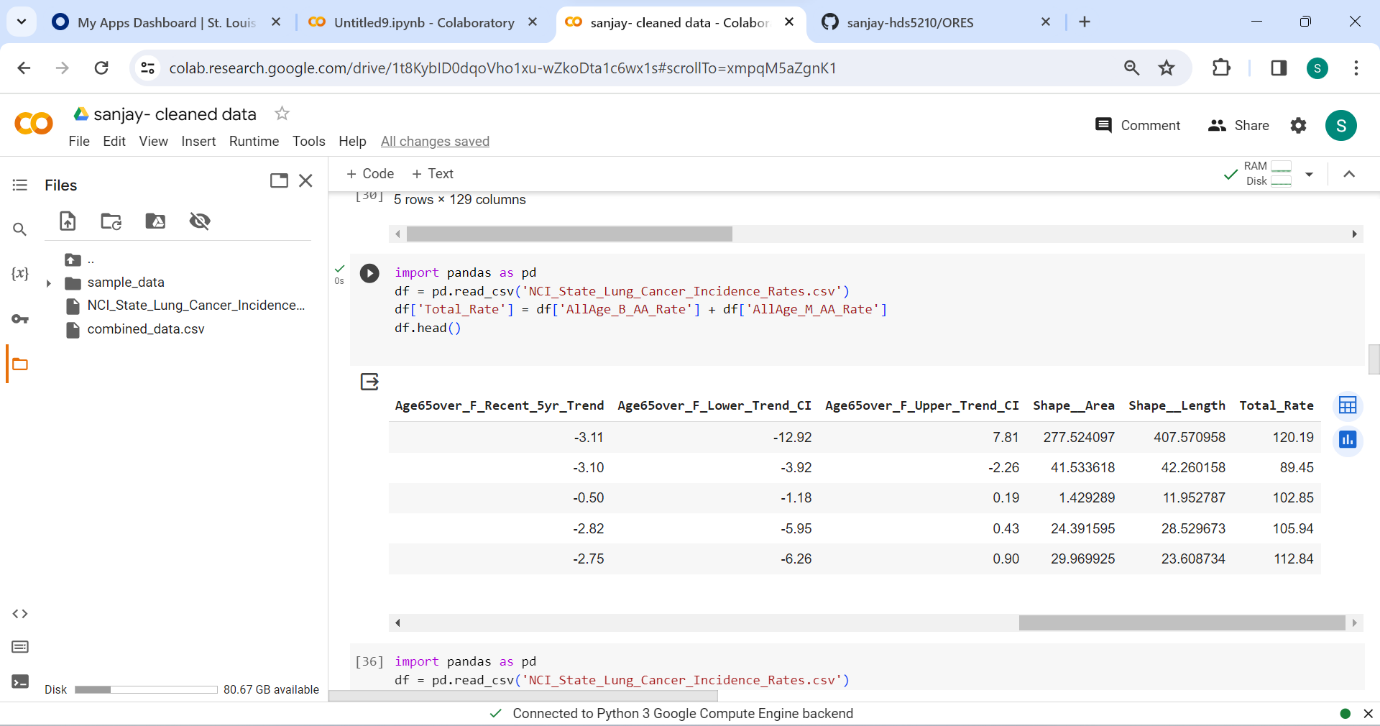
This function represents the data transformation. Here, I changed data type of the column ALLAGE\_B\_AA\_RATE from float to integer.



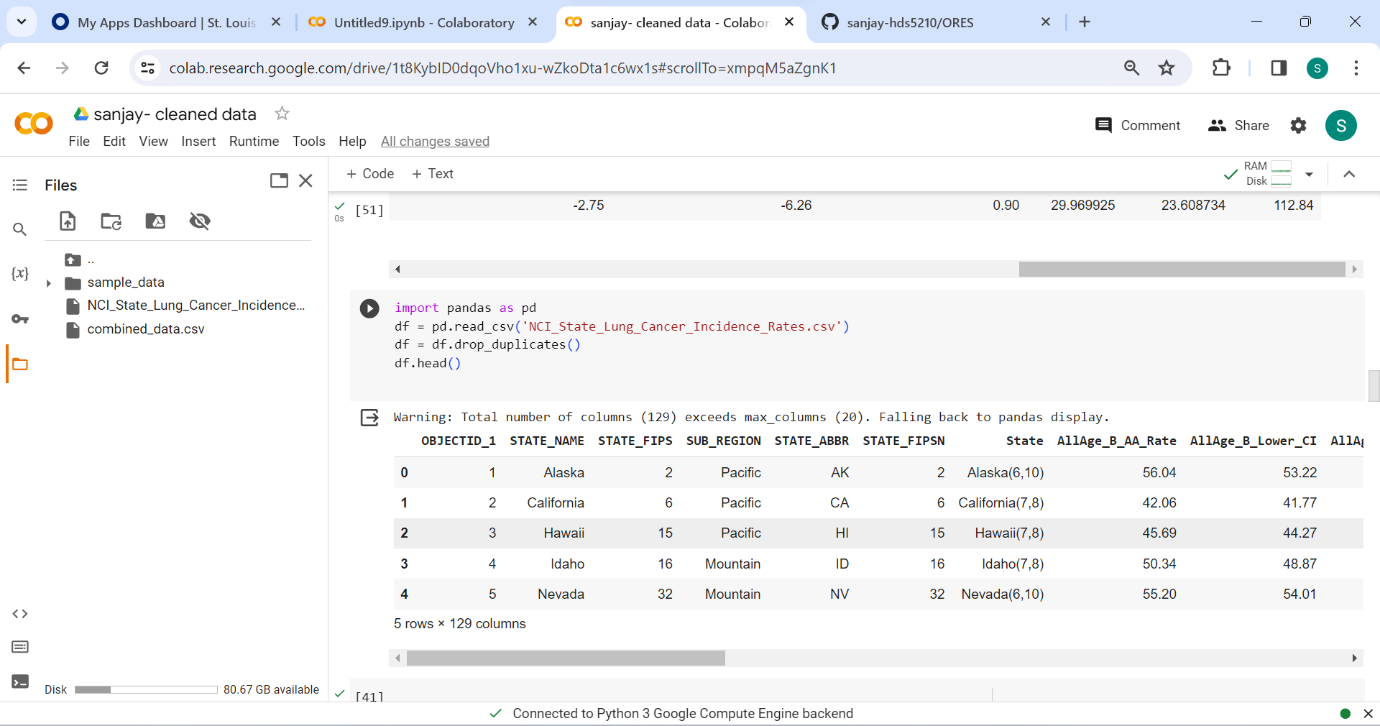
In this picture, I gave function to rename the column from OBJECTID\_1 to ID



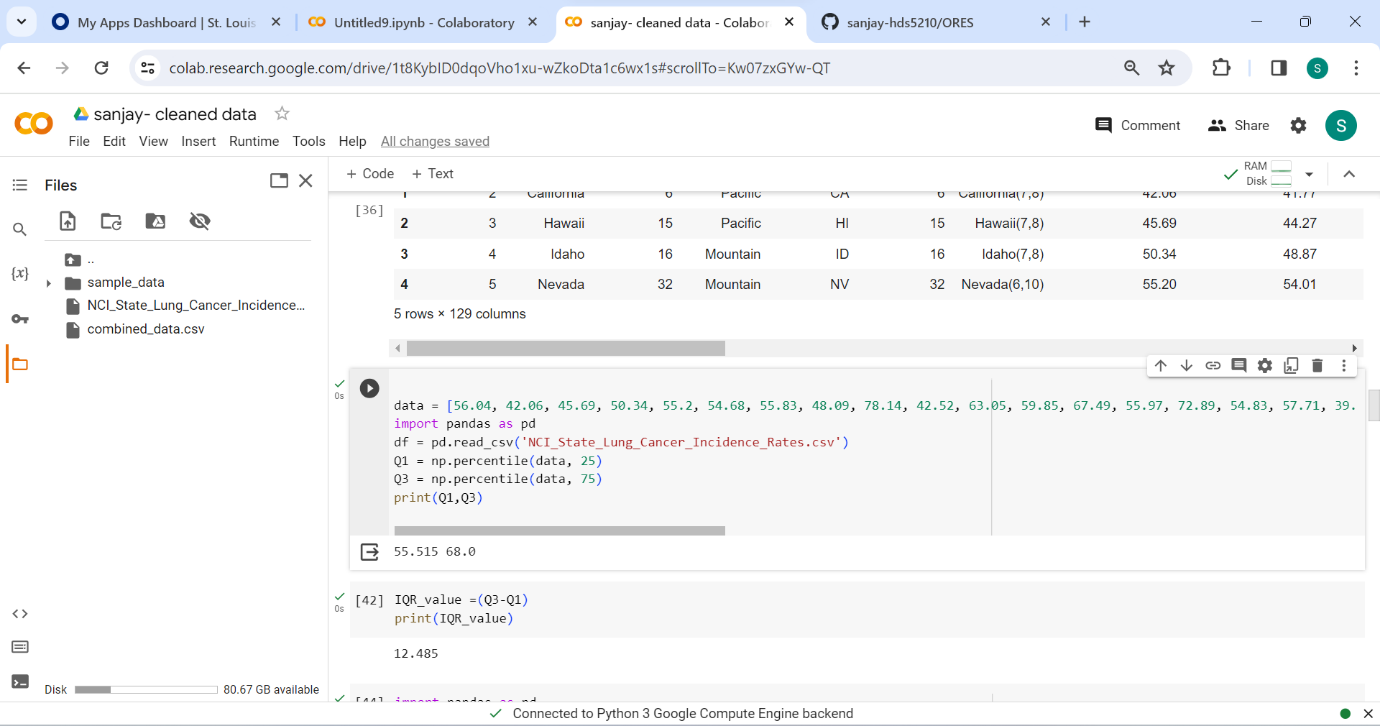
Here, I gave a function to create a TOTAL\_RATE as calculated field by adding ALLAGE\_B\_AA\_RATE and ALLAGE\_M\_AA\_RATE.



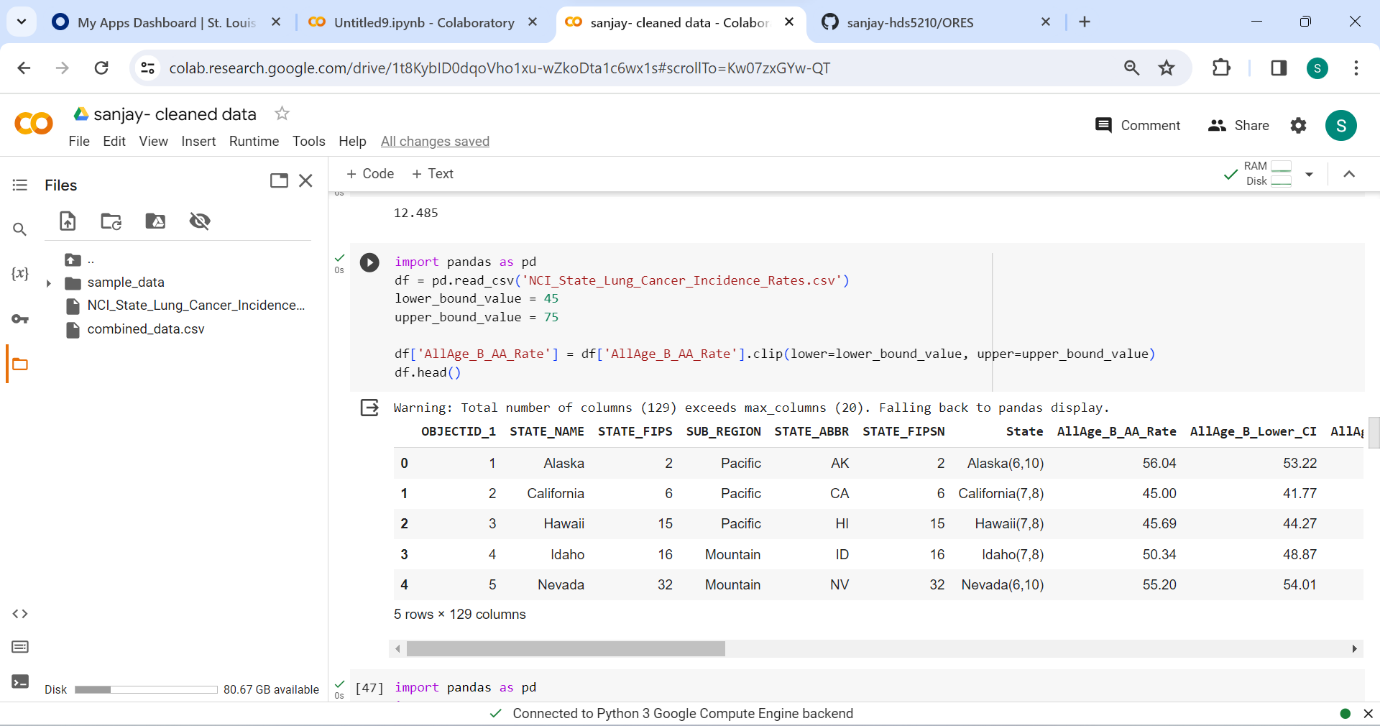
This function is to drop the duplicates. I observed that there is no change in the output as there is no duplicates in my dataset.



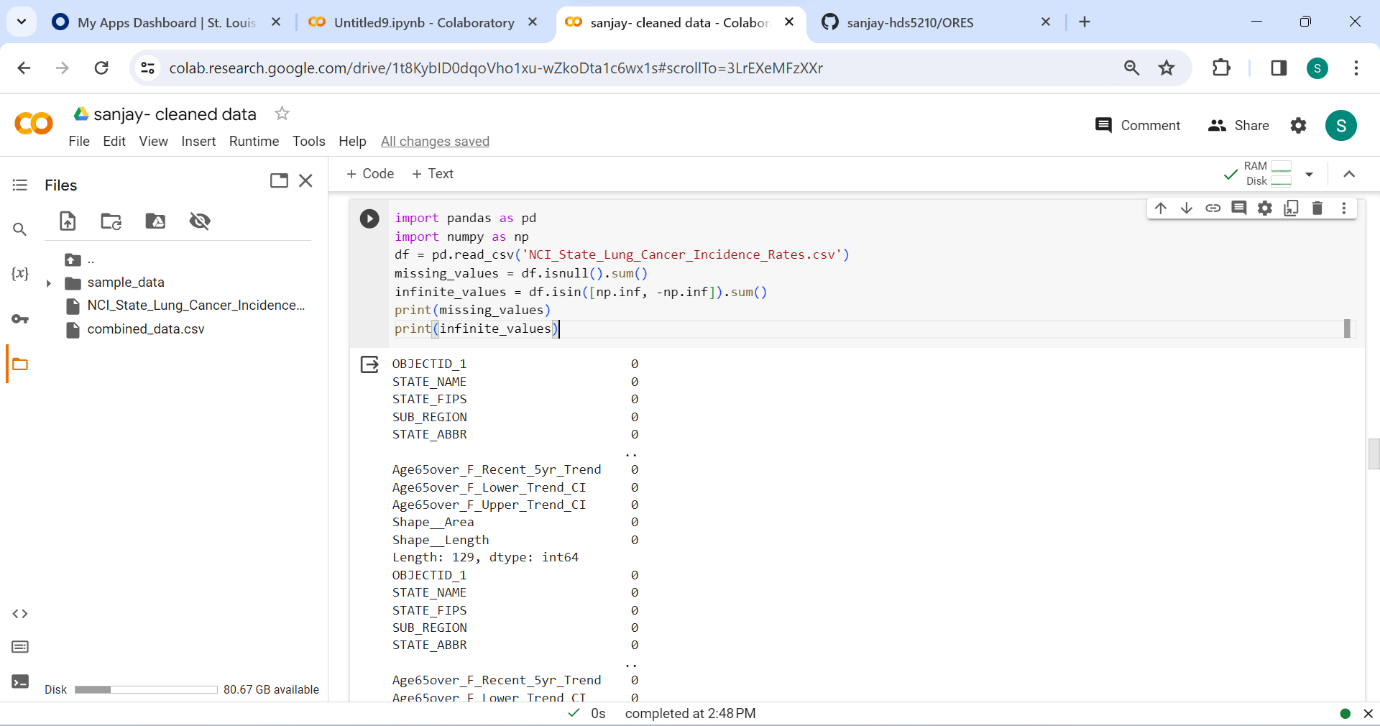
I ran this function to identify the outliers and IQR value in the dataset.



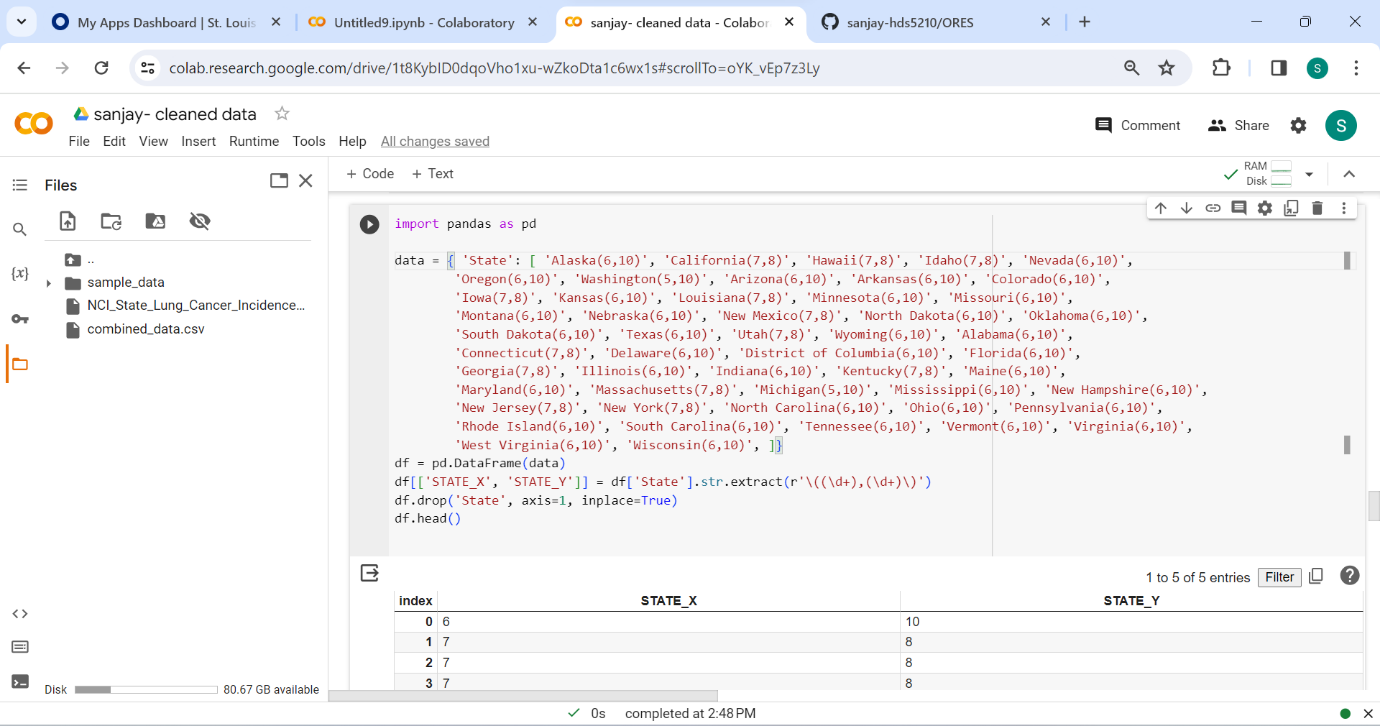
In this picture, I used this function to clip the values of the column ALLAGE\_B\_AA\_AGE from 45 to 75 by using clipping method



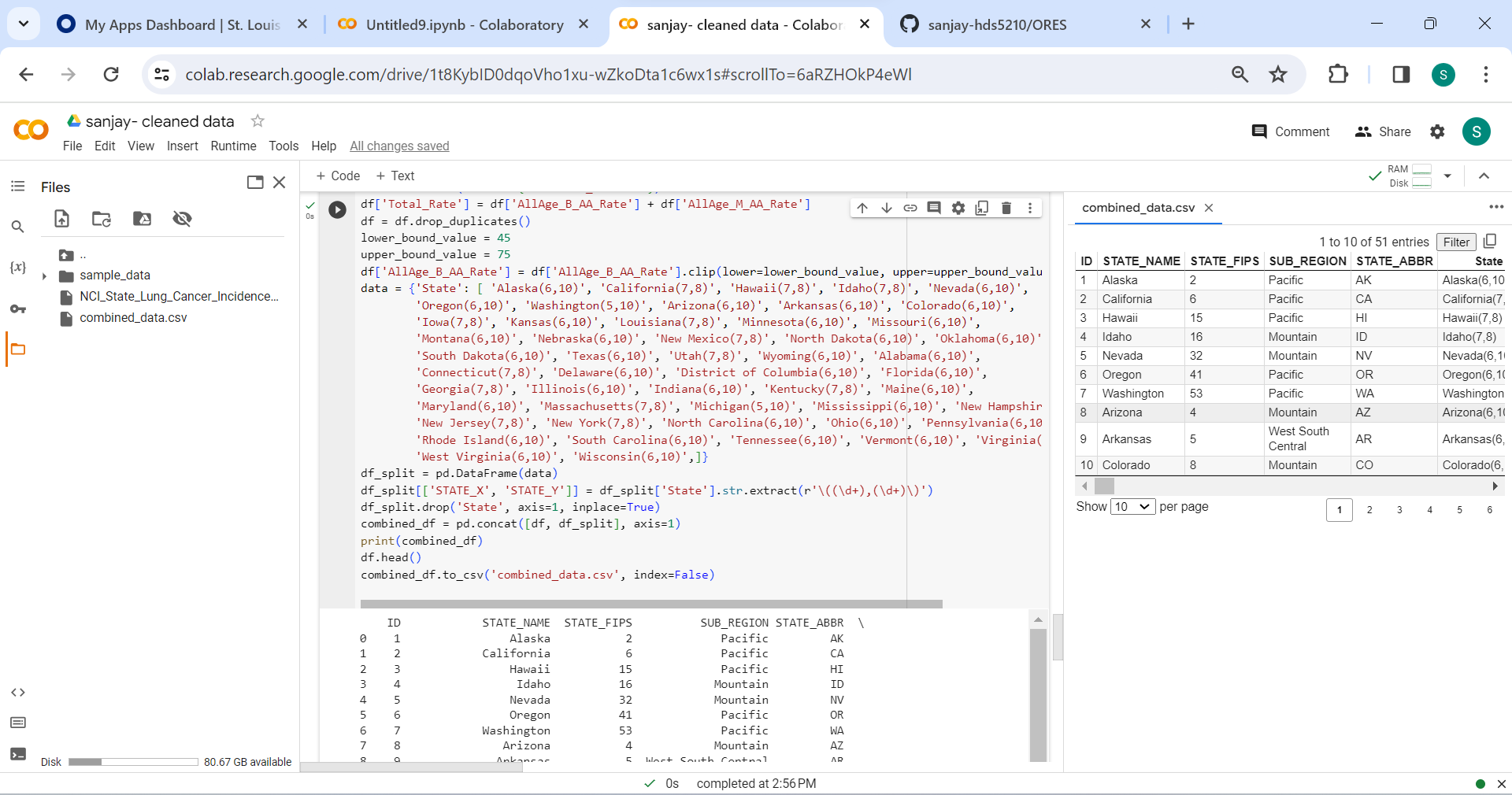
This function is to deal with null or infinite values. Output shows that there is no null or infinite values in my dataset.



To perform other transformation in my dataset, I chose the splitting approach to split the column ‘state’ into two columns as ‘STATE\_X’ and ‘STATE\_Y’.



I used this function to combine and save all the codes to produce the cleaned dataset.



APPENDIX AND CITATIONS:

Drop columns and rows in pandas data frame-<https://youtu.be/oeGk7hrXMlY>

How to rename columns in pandas data frame-<https://youtu.be/6eJlRVygEMk>

How to change the data type of the column-<https://youtu.be/evKYySLSzyk>

Remove duplicate data from pandas-<https://youtu.be/JU1EhK89ThE>

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

CHAT GPT:

Prompt: Give me code to combine all the outputs of the code and download it