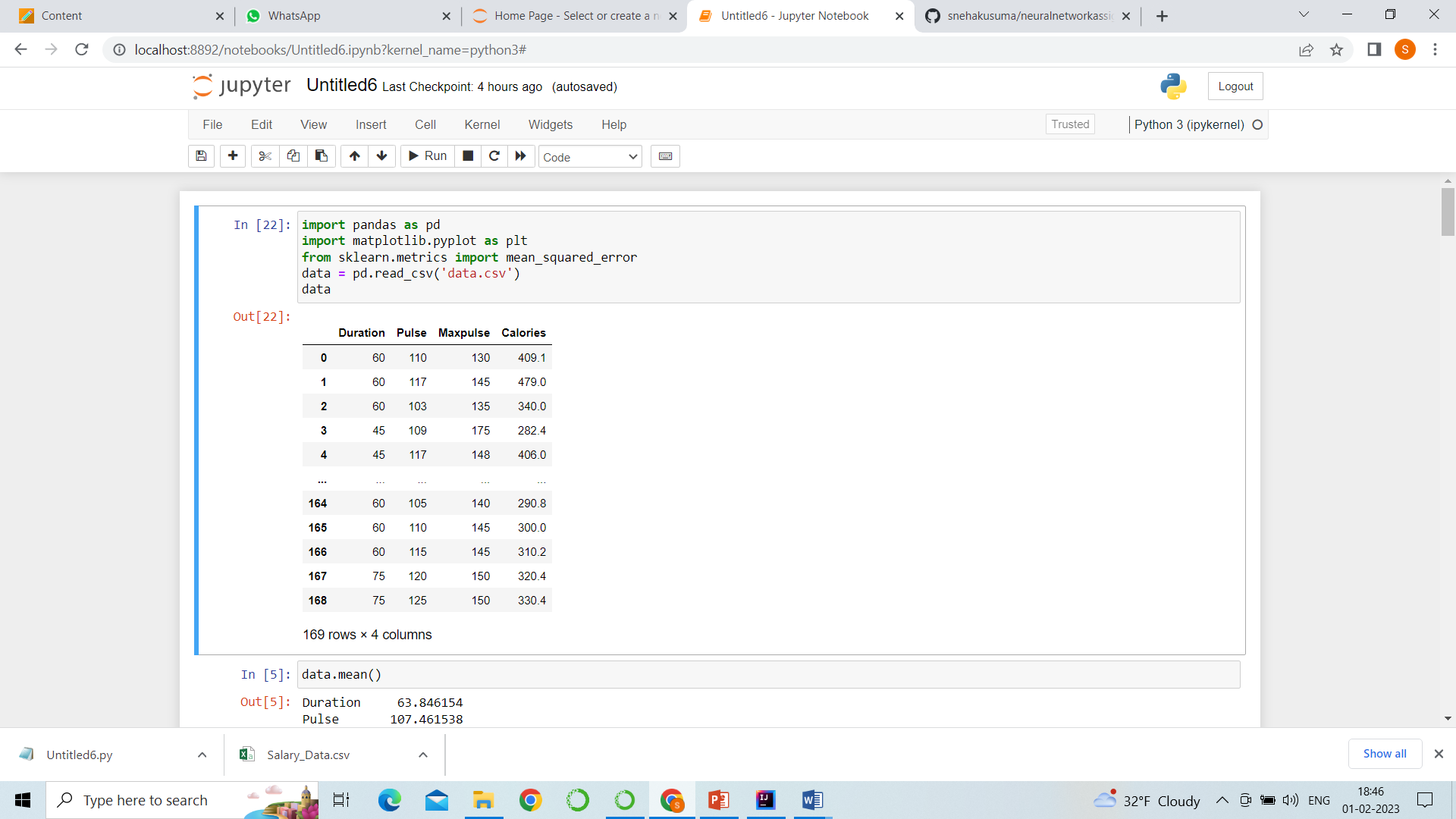
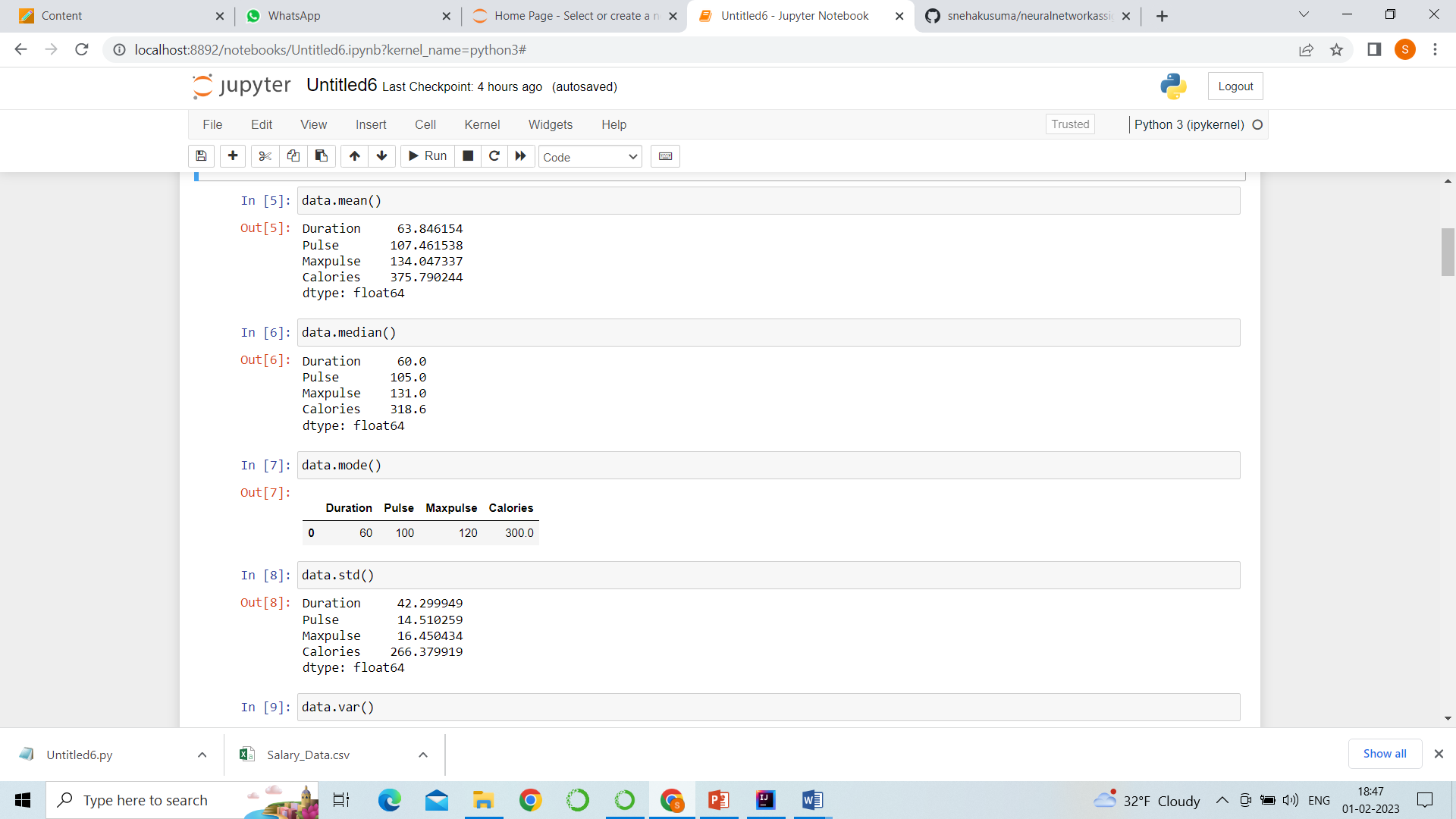
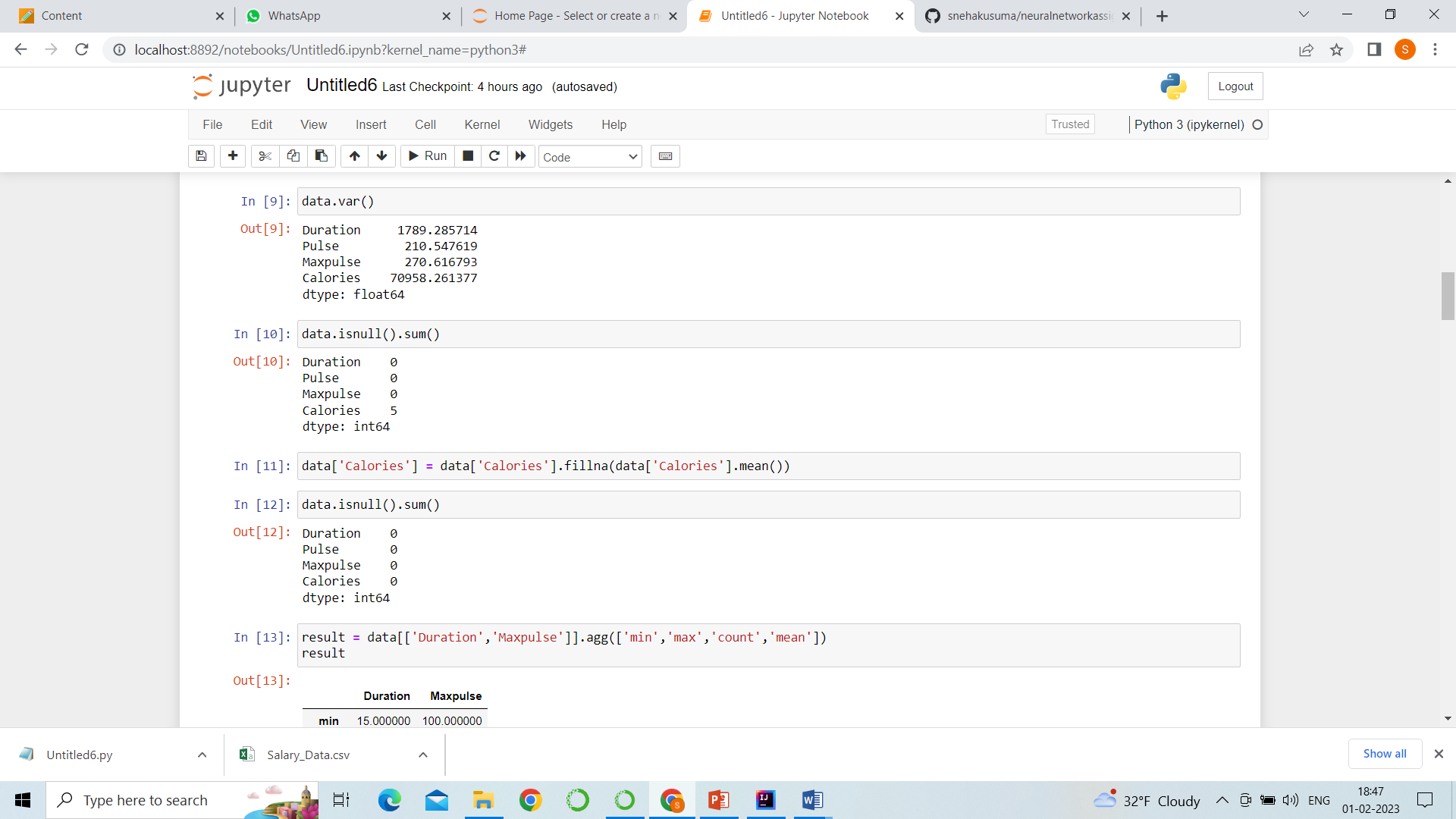
**QUESTION-1:DATA MANIPULATION**



1. We are importing pandas library to read the data
2. Data is read from data.csv file and stored in data.
3. So, the output contains duration, pulse, maxpulse and calories.



1. data.mean is a function to get the mean of the data .
2. data.median , data.mode and standard deviation data as well.

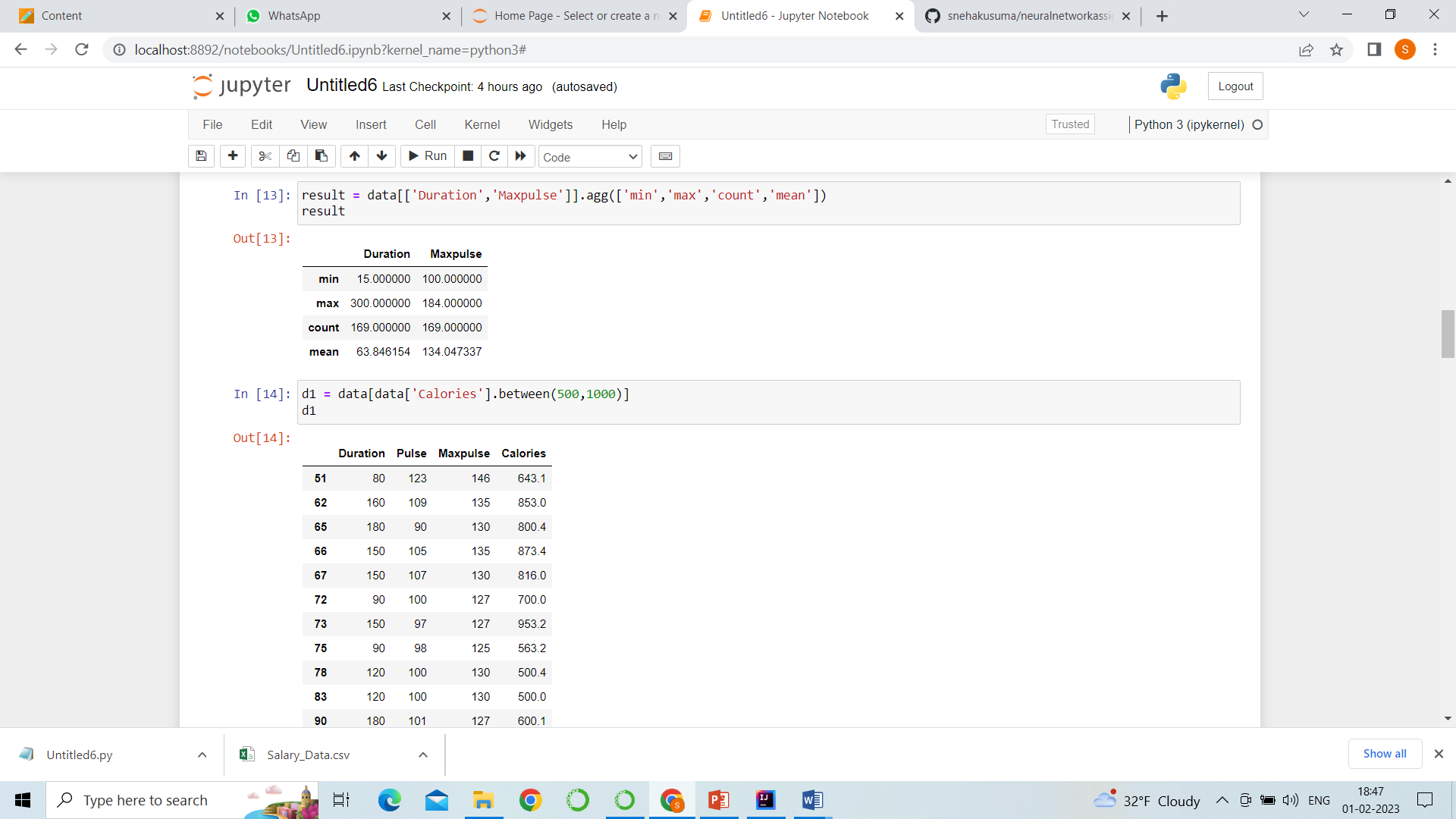


6.with data.var function, we found the variance of the data.

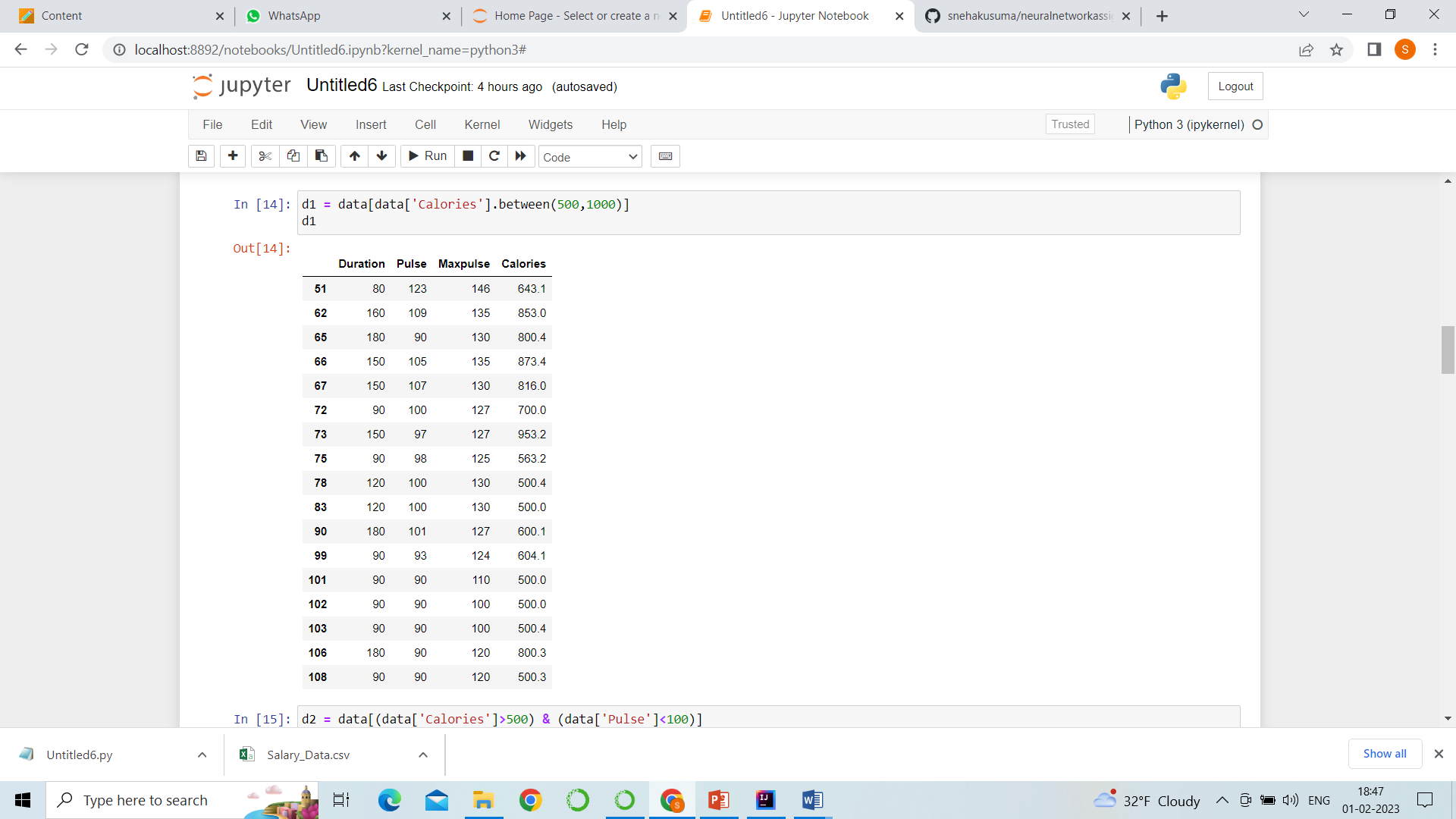
7. data.isnull().sum() which used to get the null values in the stored data and here we are summing

up the null values.

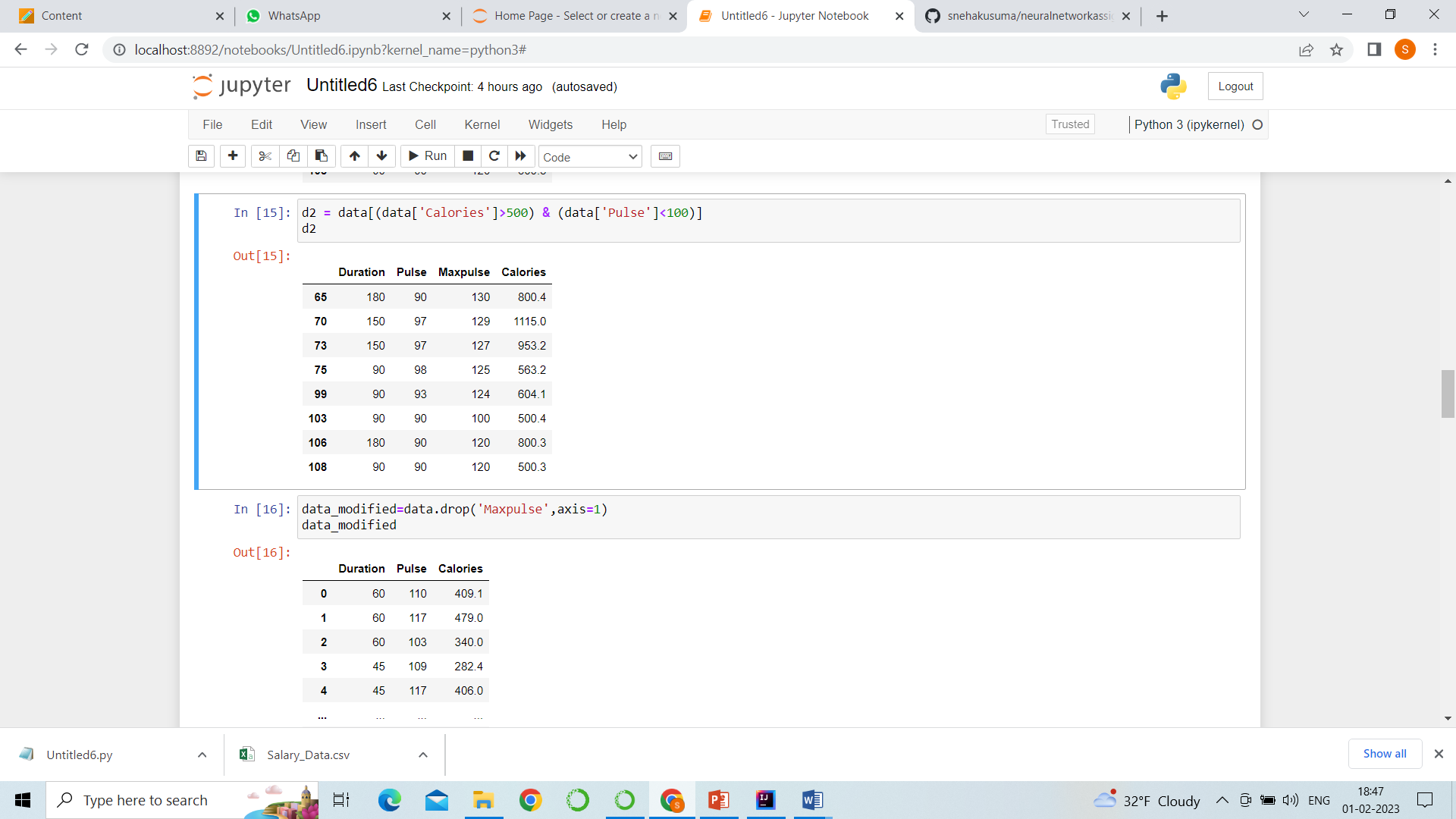
8. next we are finding the null values in a calories column and performing a mean



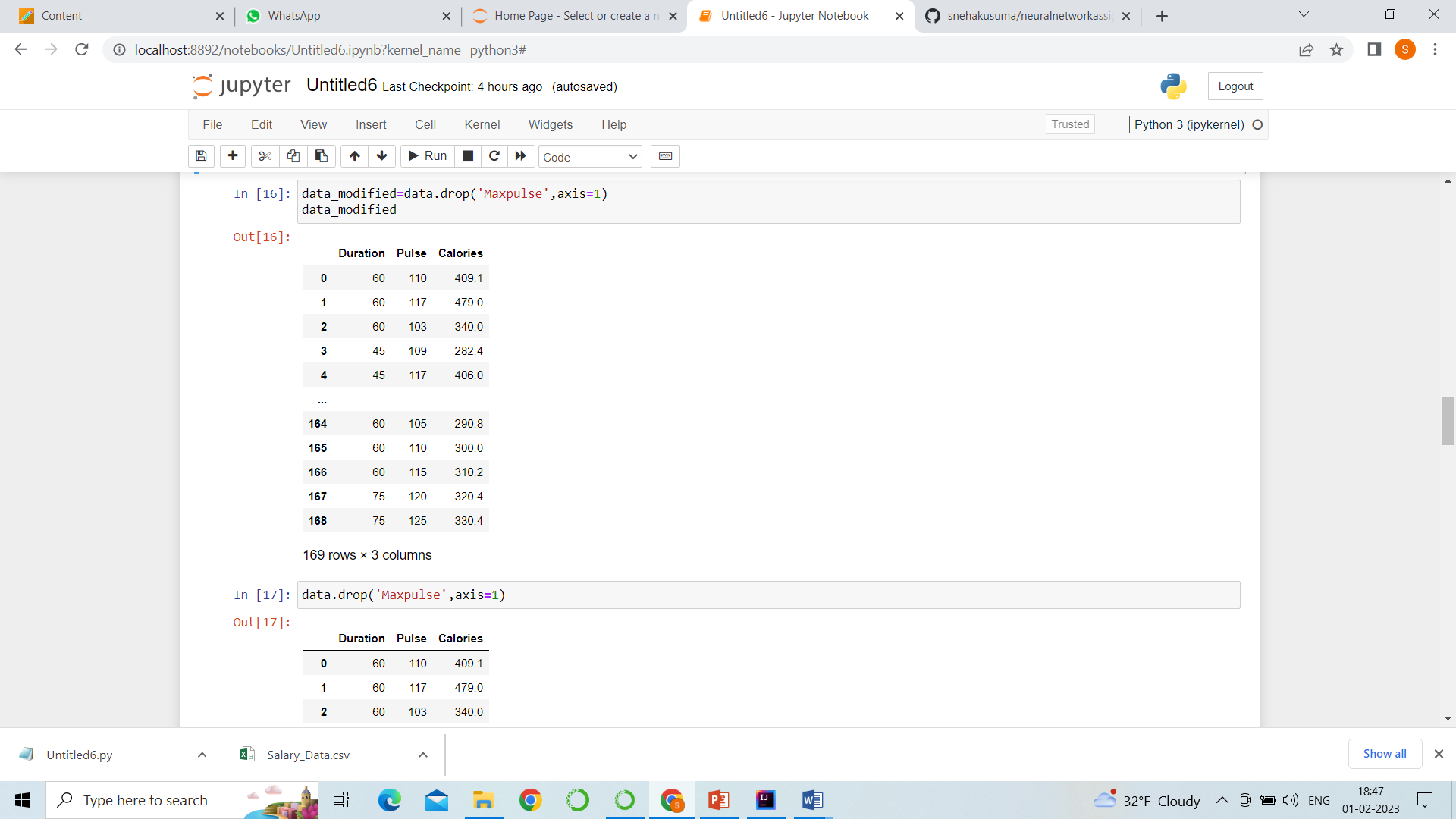
9. Here, we are storing the columns (duration, maxpulse) in result and from the columns of duration and maxpulse we got max, min values and we got mean value.



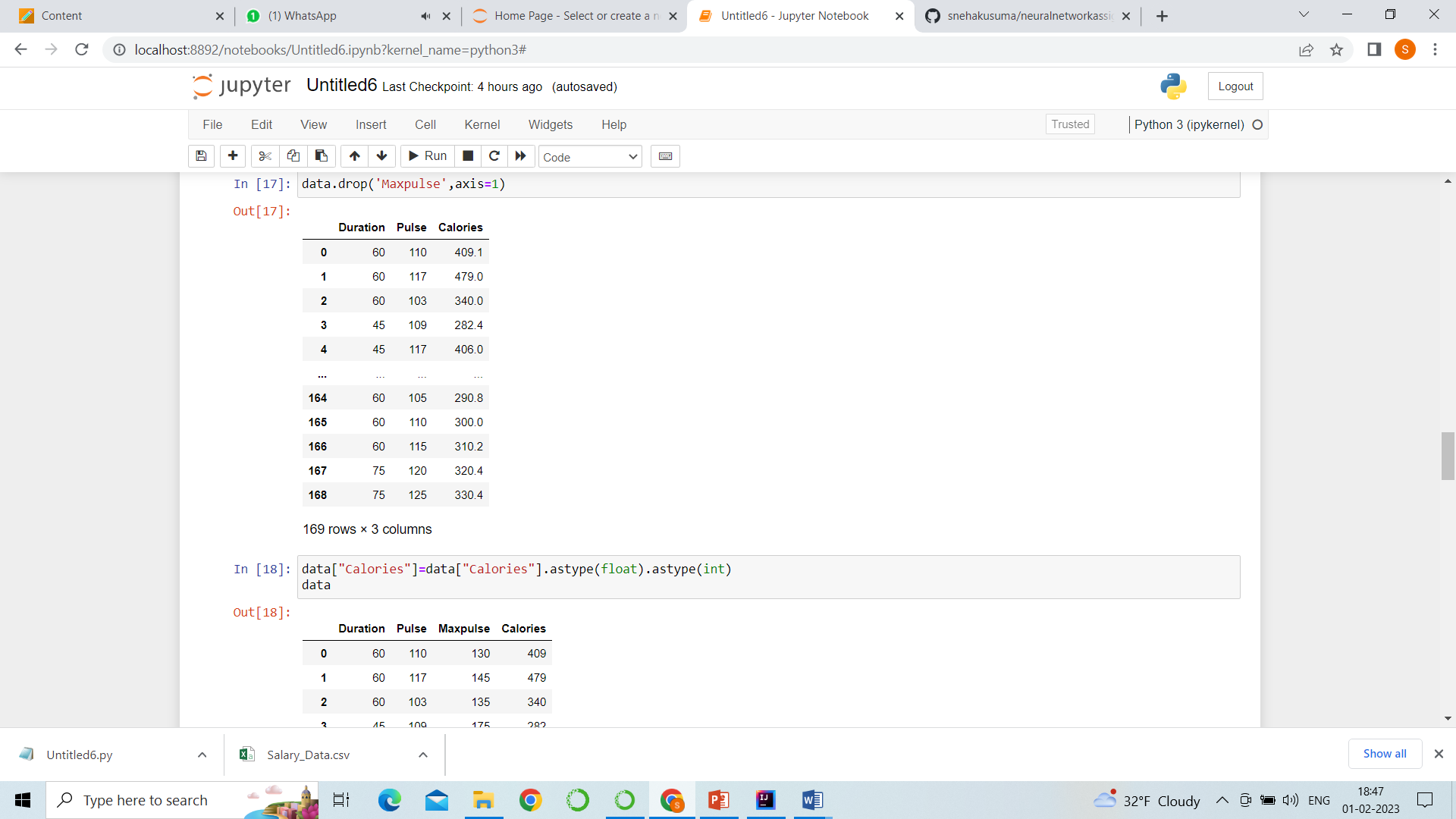
10. Here, from the calories column we have some values, so here we got values in range from 500 to 1000 and it was being stored in d1.

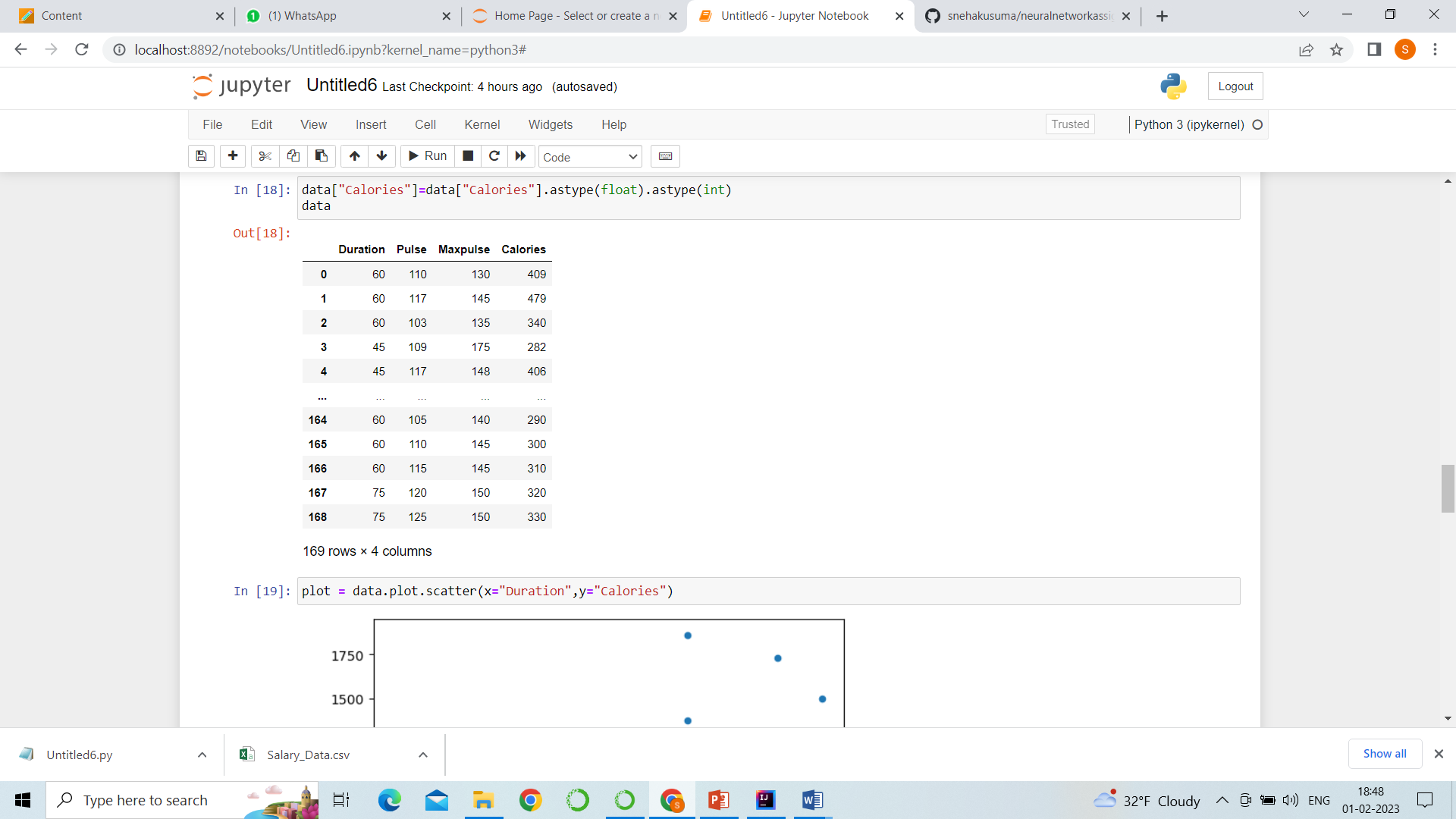


11. from calories column the values which are greater than 500 and data pulse which are less than 100 are been printed and stored in d2.

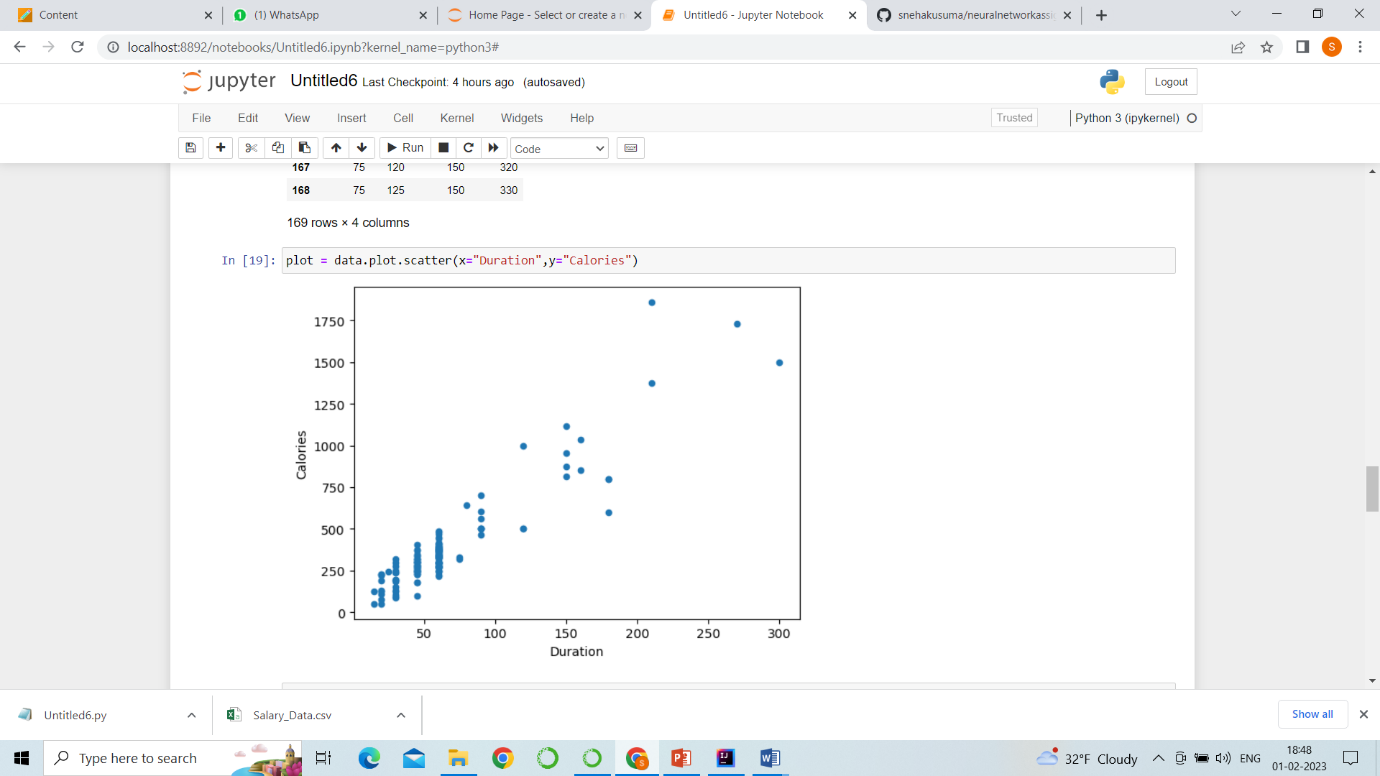


12. here, the maxpulse column was dropped and remaining data is displayed and it is being stored in data modified.



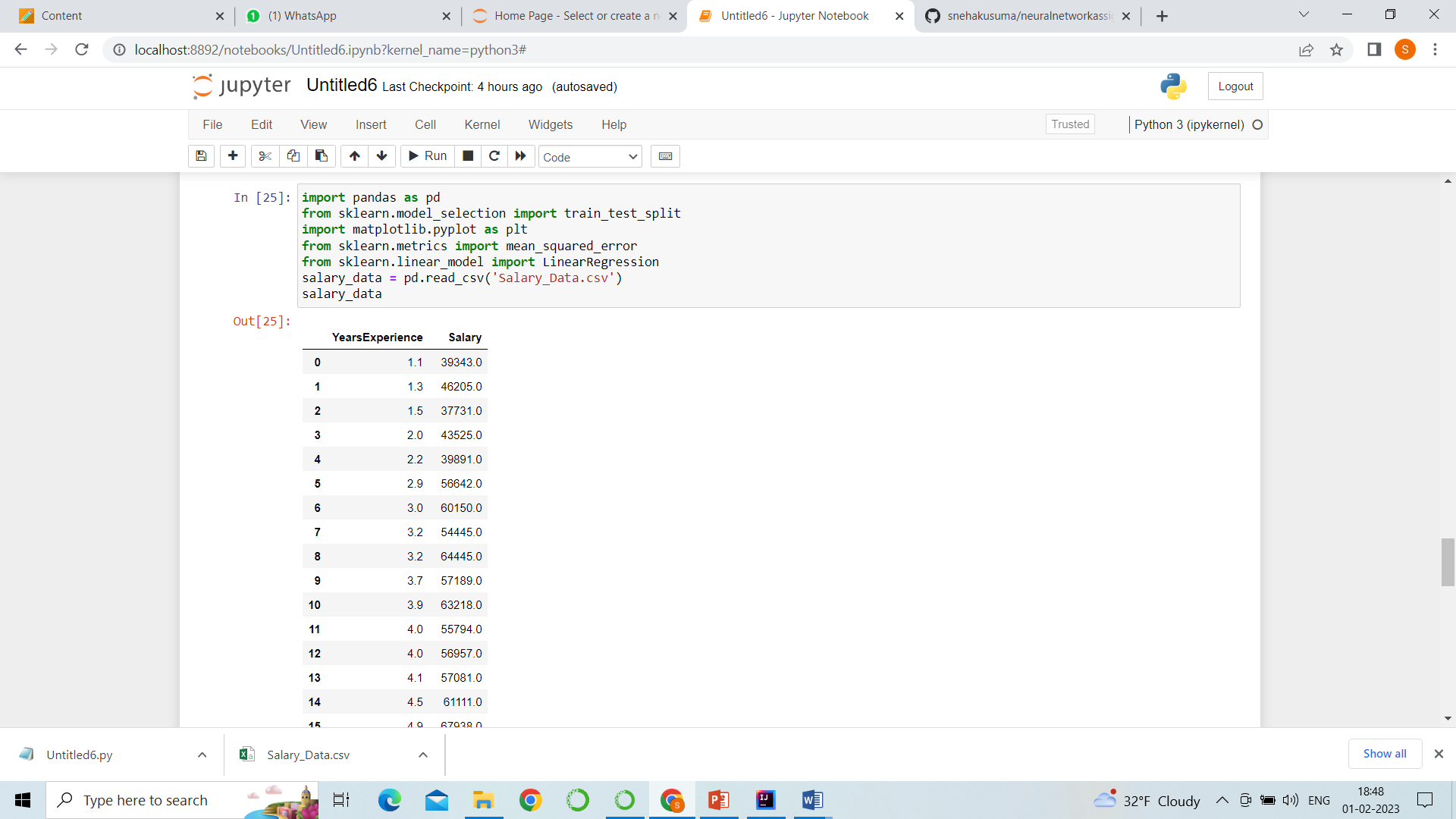


13. float values which are in calories column is converted into integer type with the help this function .

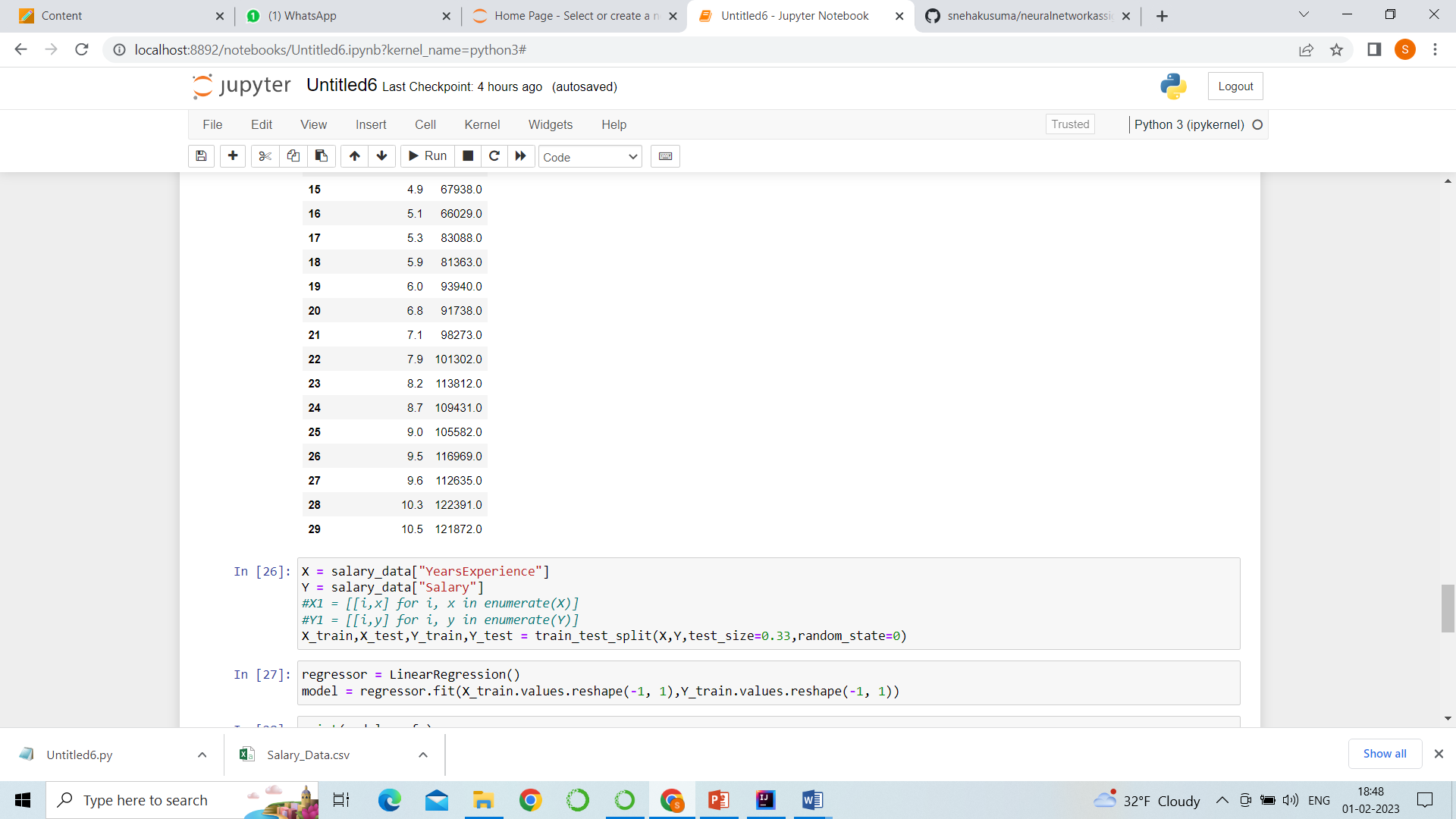


14.so, finally this is the output, the function we used in this program is used to plot the values.

**QUESTION-2:LINEAR REGRESSION**



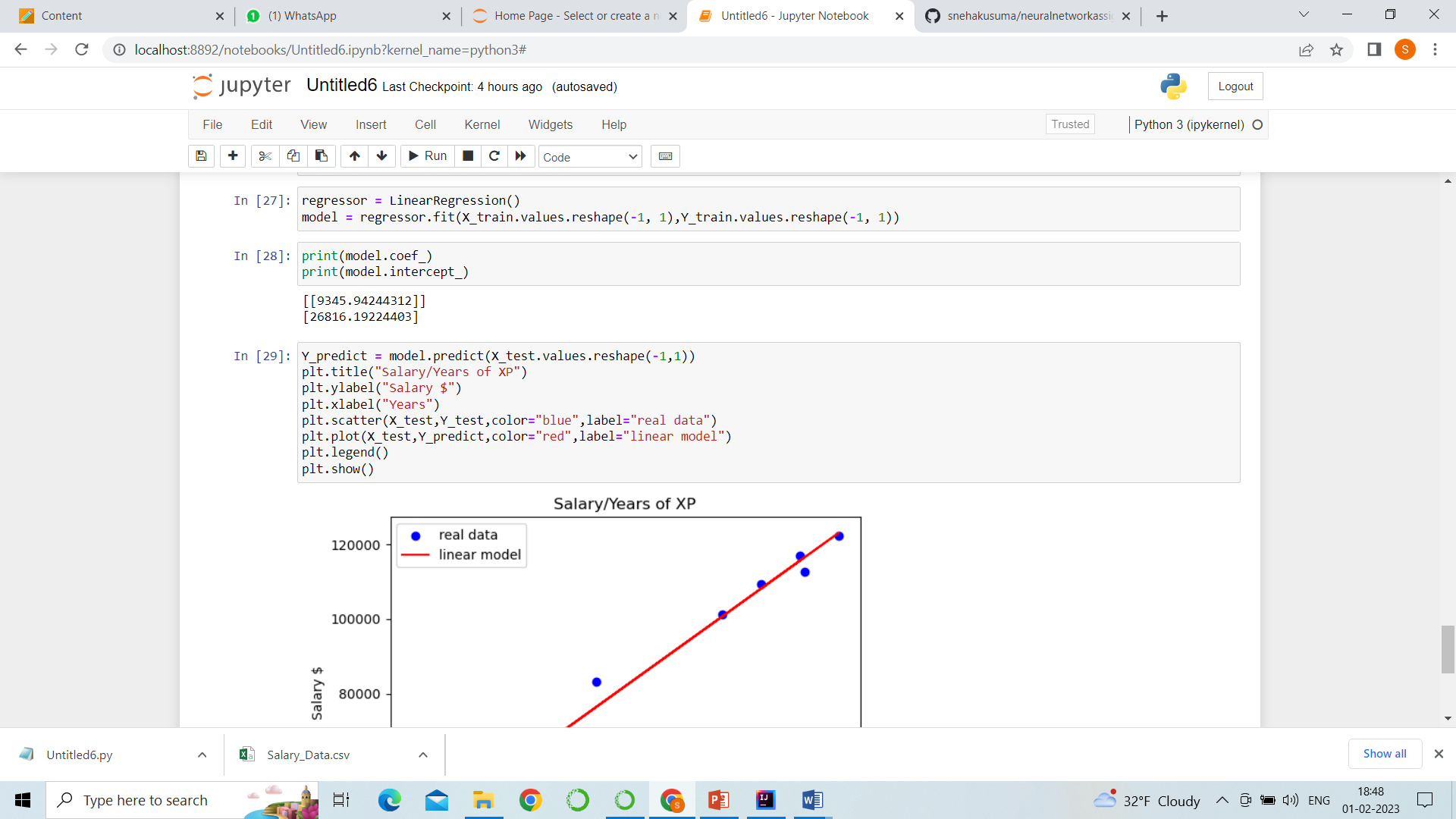
1. import panda the library is used to read the data.
2. Import train\_test\_spilt: to split and test
3. Import matplotlib.pyplot : to plot the values.
4. Data is read from data.csv file and stored in data



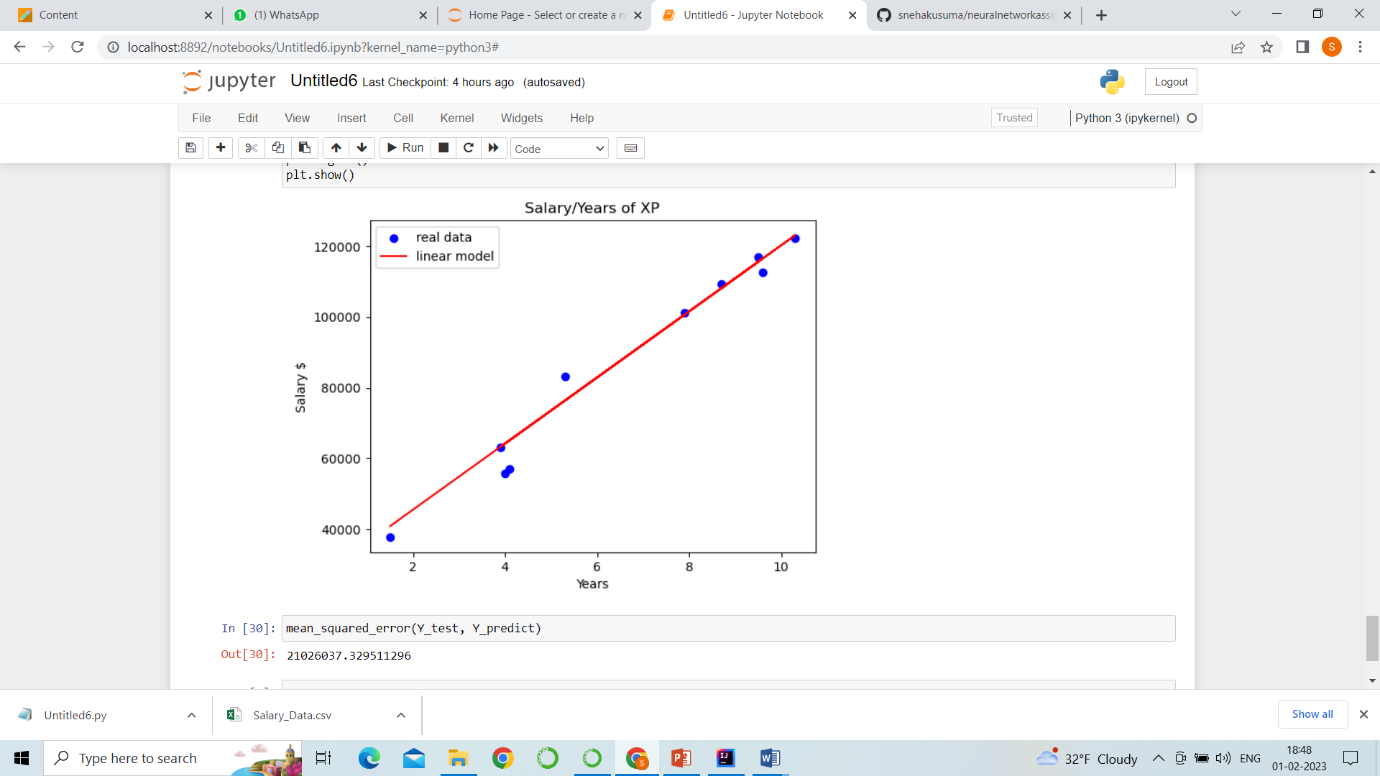
4. so, years of experience column is stored in x.

5. salary column is stored in y.

6. we imported linear regression model to train values and to reshape.



1. The predicted and reshaped model is stored in y\_predict.
2. The x axis is years and y axis is salary and the title is salary /years of xp.
3. We used scatter method to plot the original data and the displayed color is blue, label is real data.
4. We plot the original data and the color is red, label is linear model.



1. We got the mean squared error value of y test and y predict which is 21026037.329511296