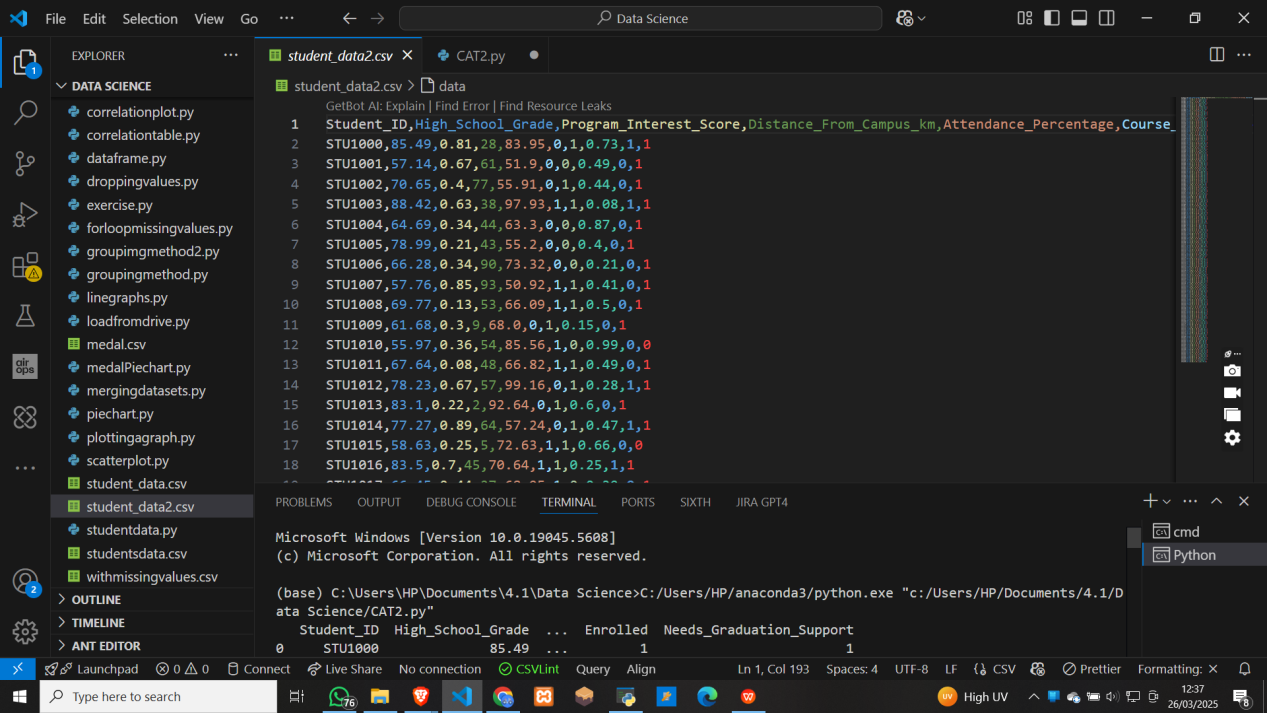
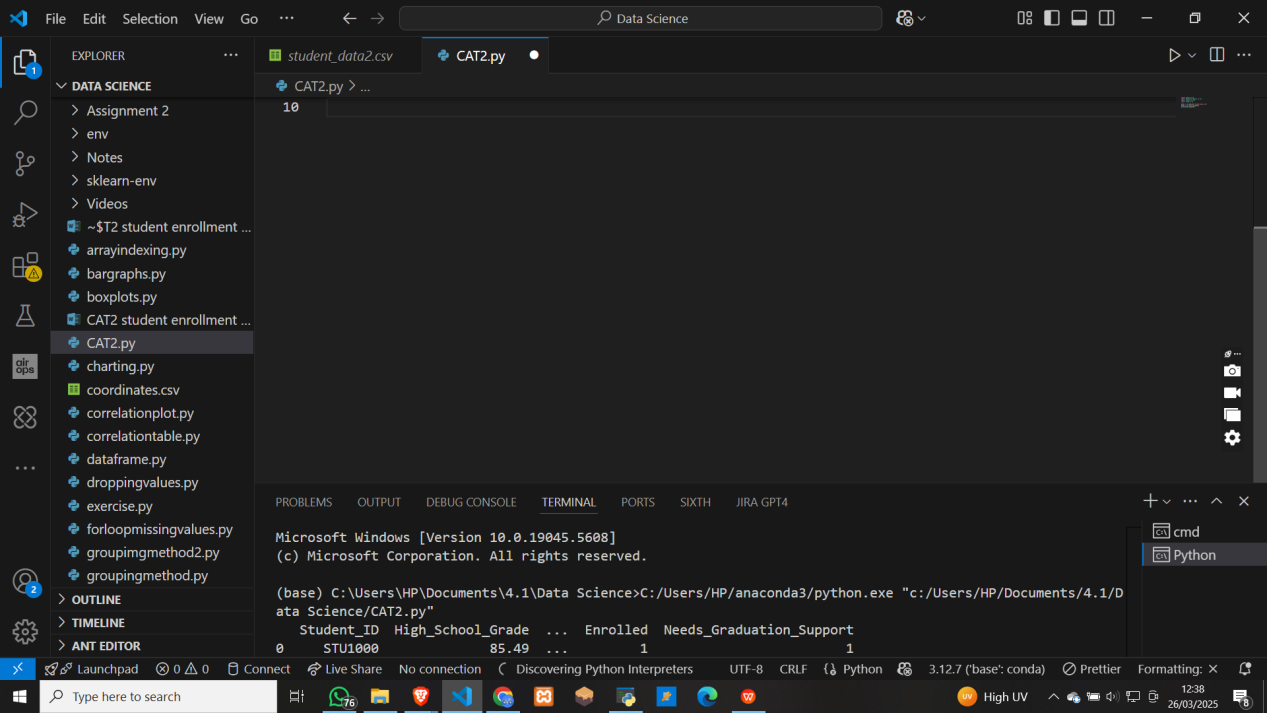
### CAT2. Document - Student Enrollment Prediction

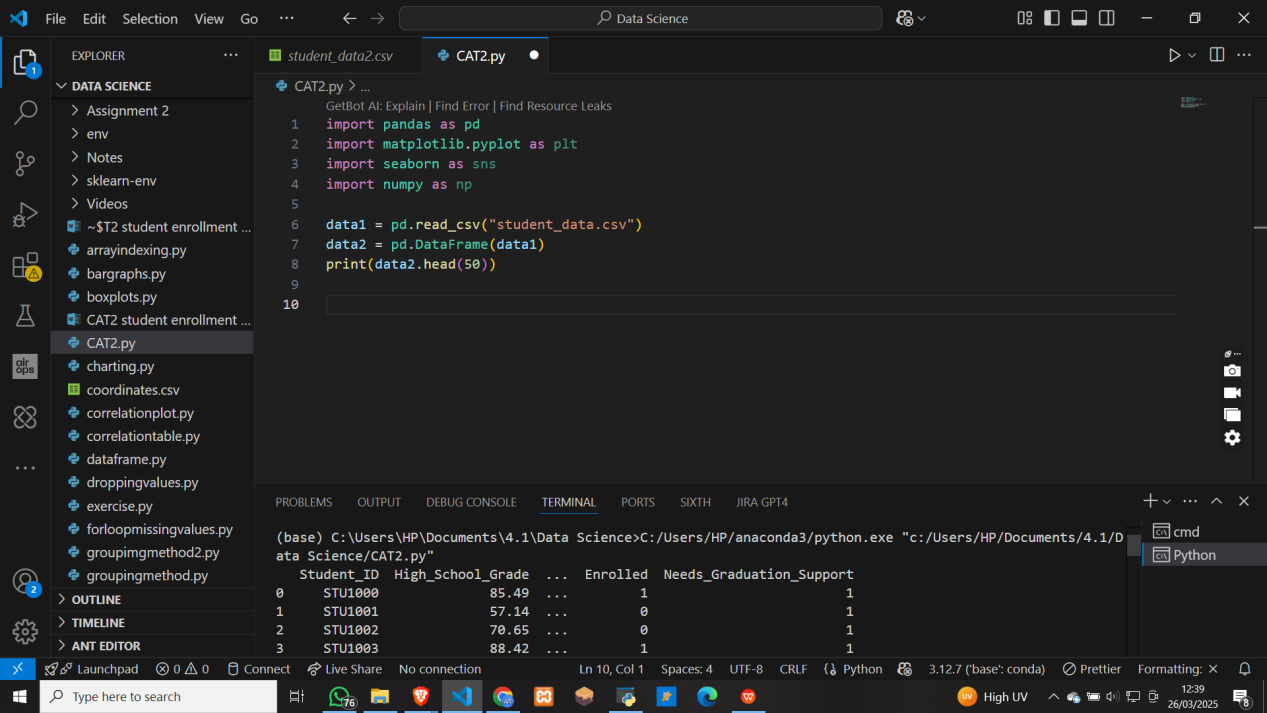
Hello sir  
Now I started looking for the dataset  
I wasn’t sure I would find the data on <https://www.kaggle.com/search?q=students+enrollment>So, I created my own, or I had help creating it



I then created a python file and named it CAT2.py



Then went ahead and loaded the csv file into my python file and ran it to make sure that my data was correct

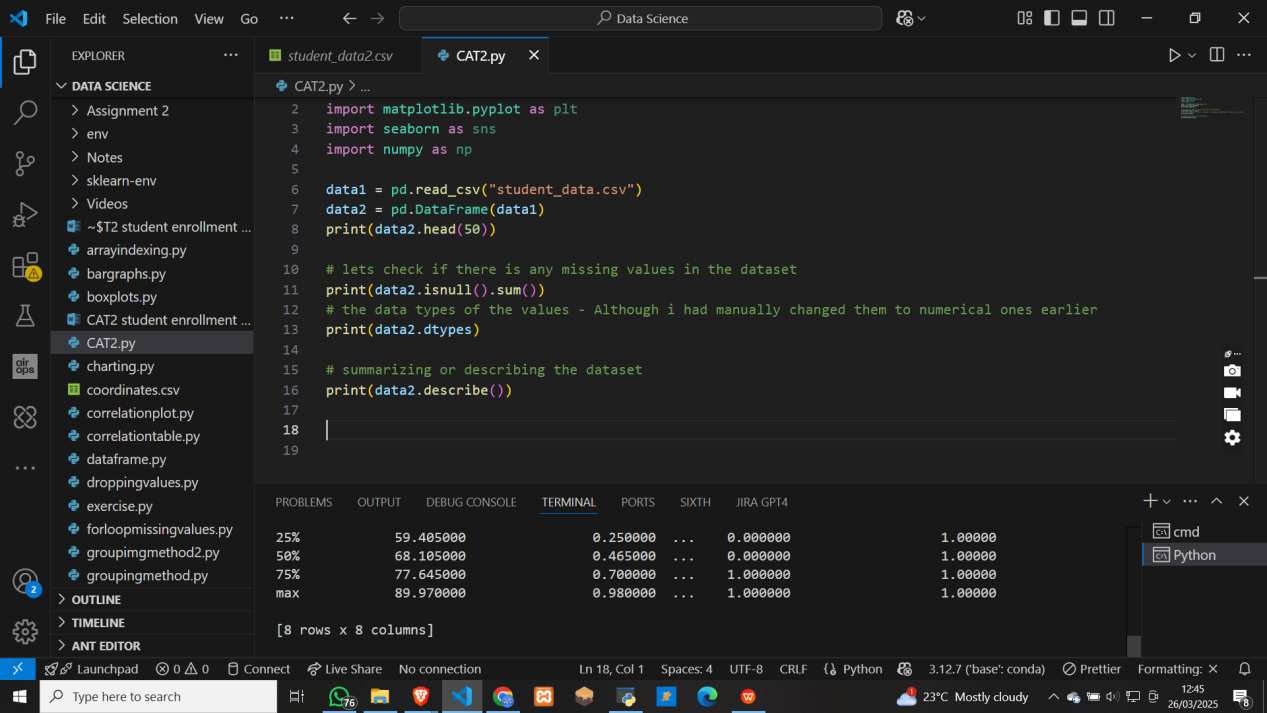


**Just a point I failed to mention, I changed all the string values in the dataset to numerical values for easier development and training of the moel**

I then move to check if there are any missing marks in the datasey - and found none



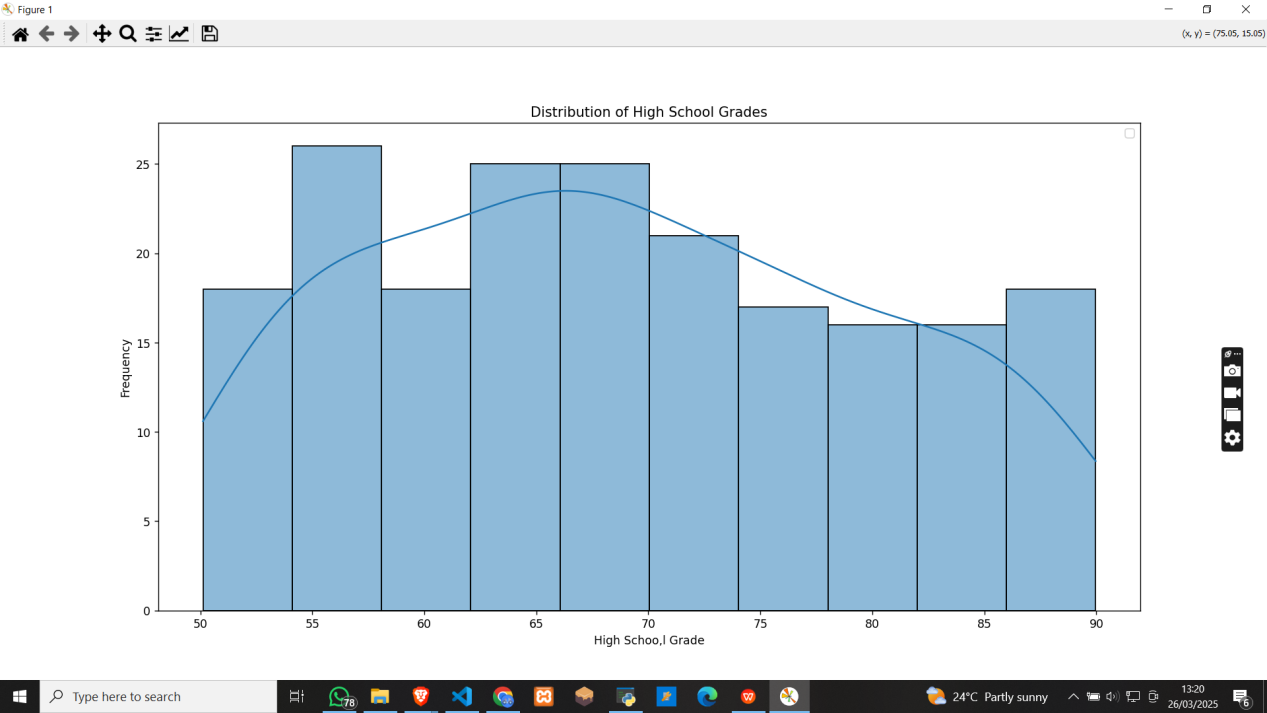
Then I went to check the few common things like describing the dataset and such



So, I moved to visualize how the students performed in high school or their grades distribution using a histogram

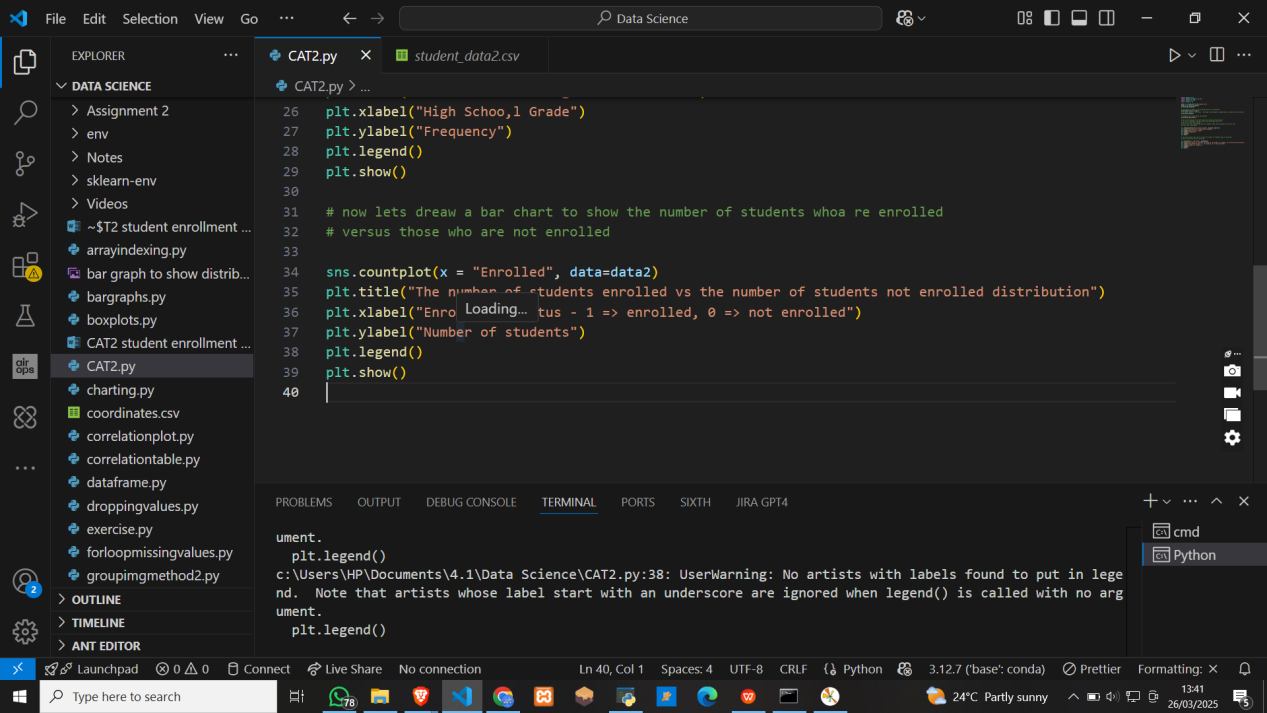


ANd here are the results of the histogram



Got a bit stuck at bar graph. I later switched from plt.bar() to sns.countplot(). This eased the issue of going back to the dataset and counting manually

But I was able to do it after almost 2 hours



Here are the results for the bar graph

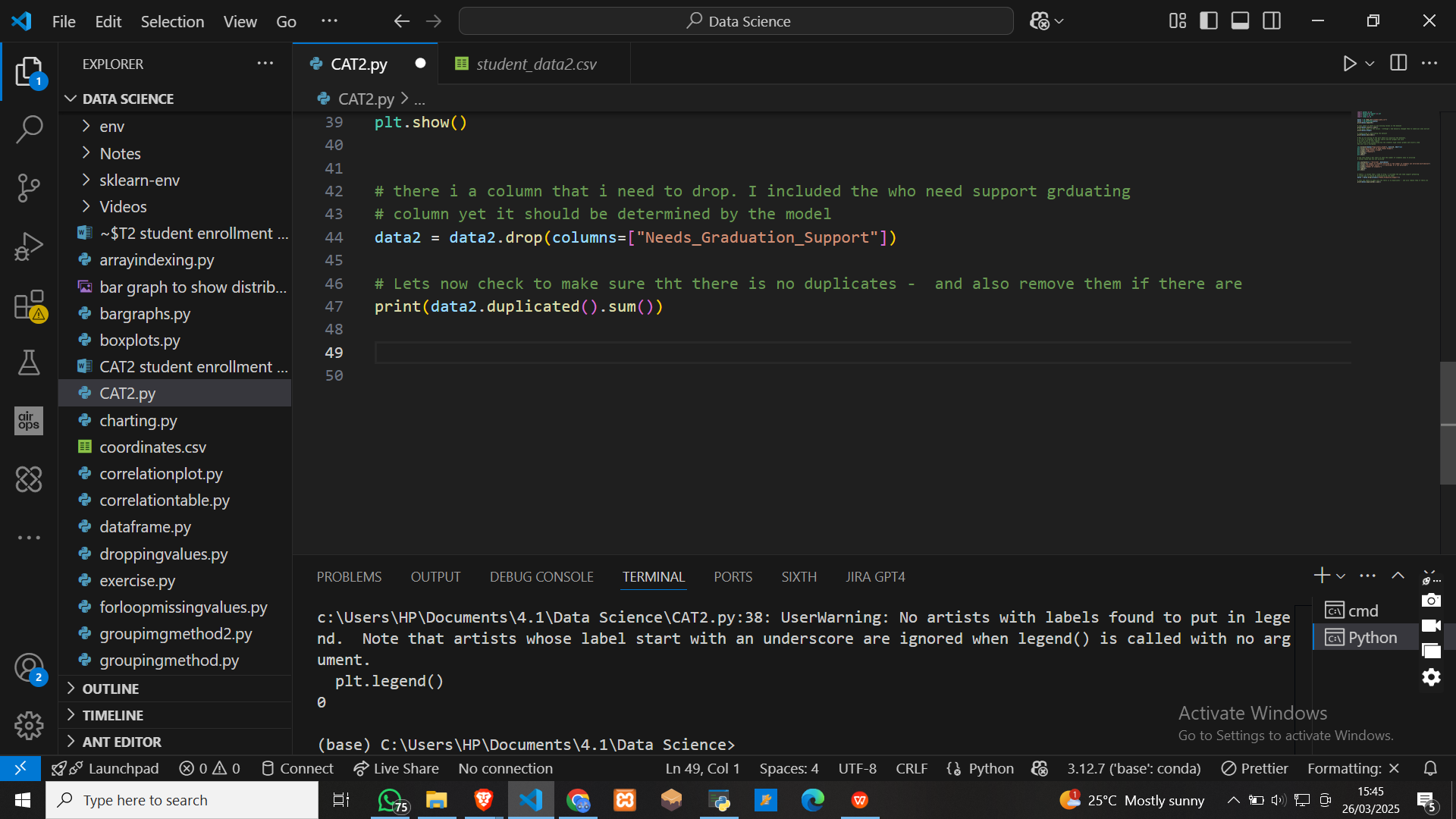


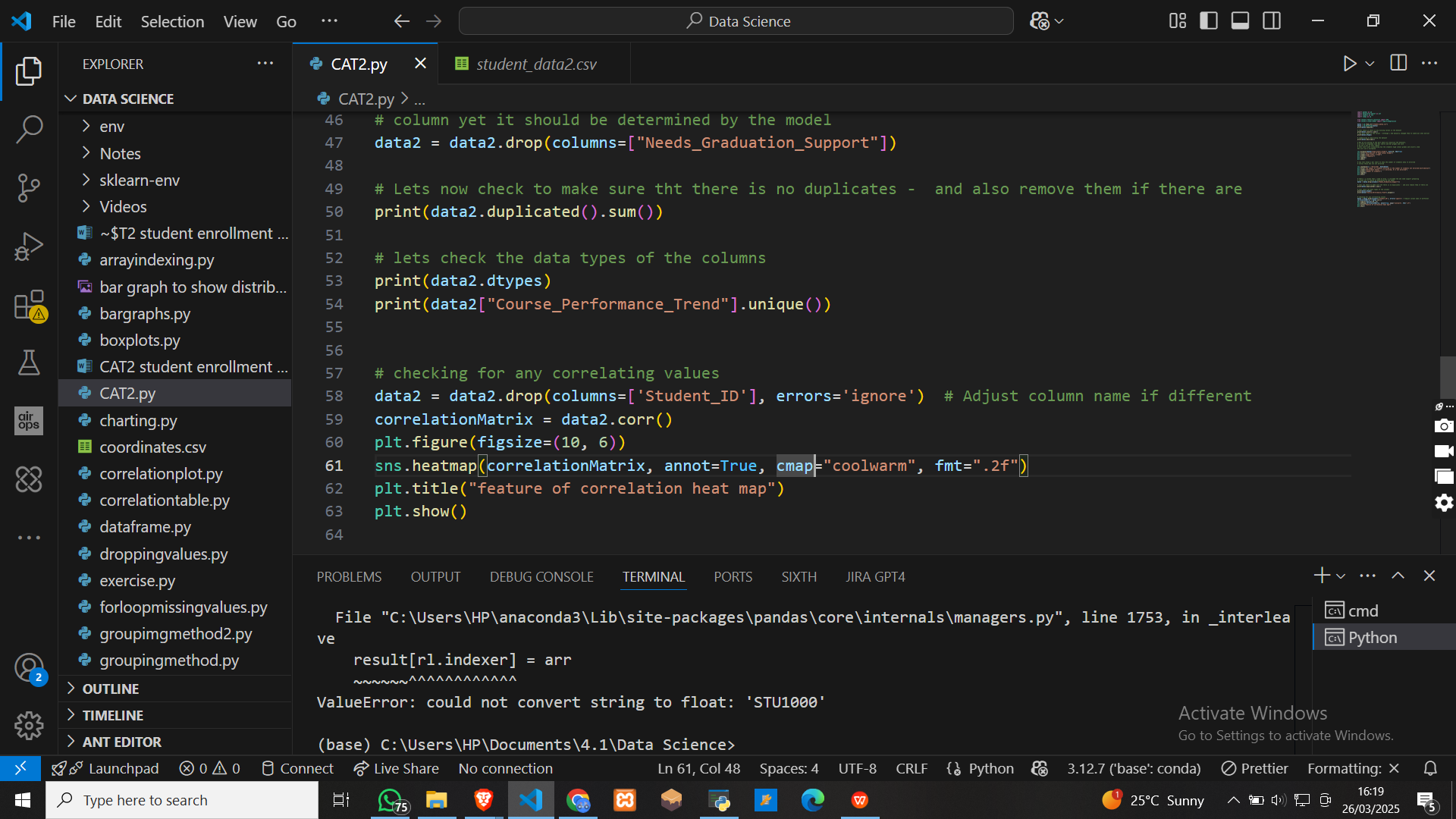
At forst, I was loooking at this project as if it was the enrollment after high school. But later thought that even in campus, one may choose to take another extra course

So, I though that maybe financial capability could also be used to determine those who can be targeted for the course.

ANd now am rambling -

Lets now check to make sure tht there is no duplicates - and also remove them if there are



Had to drop student ID since it was a string  




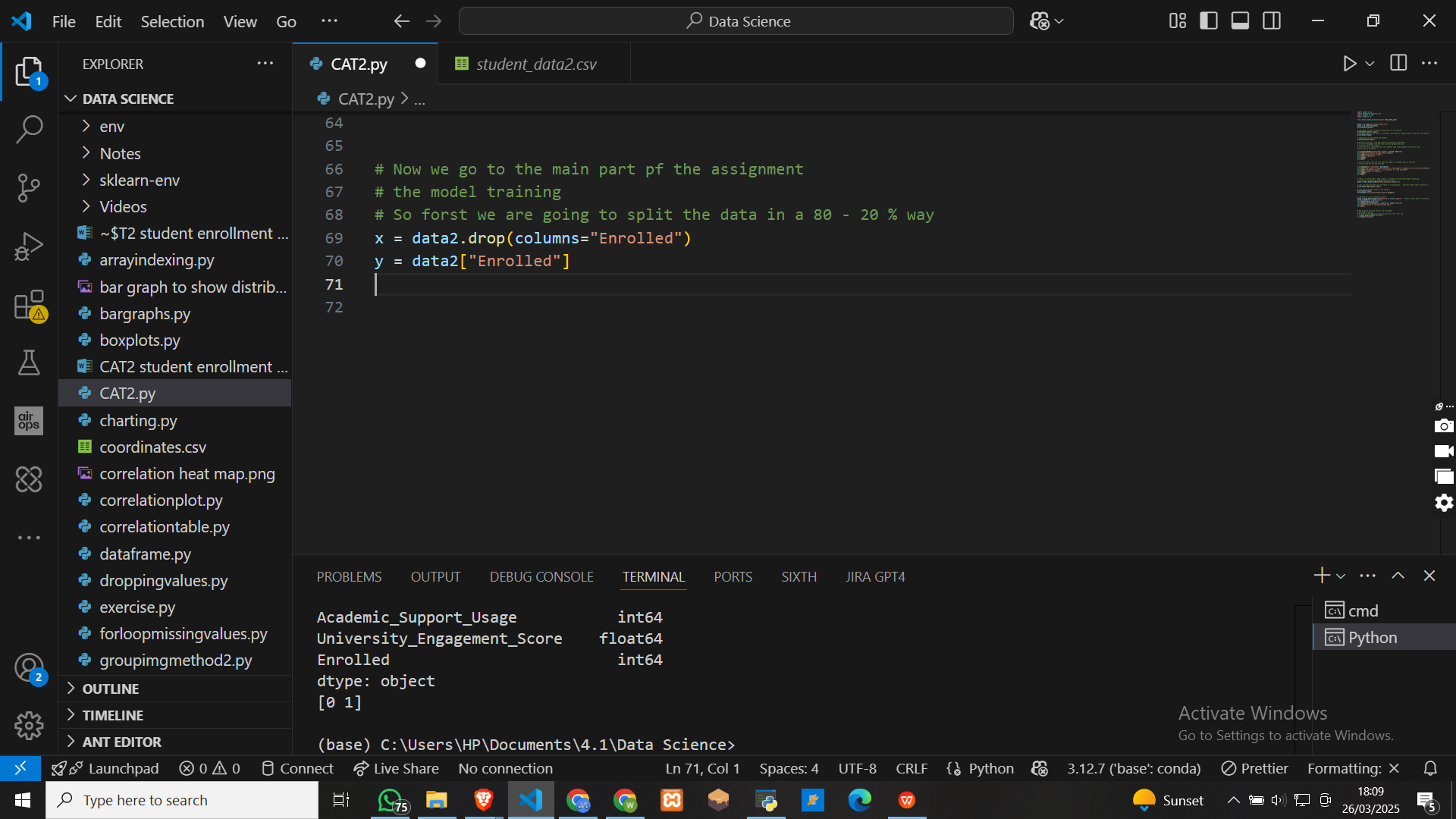
Now we go to the main part pf the assignment

# the model training

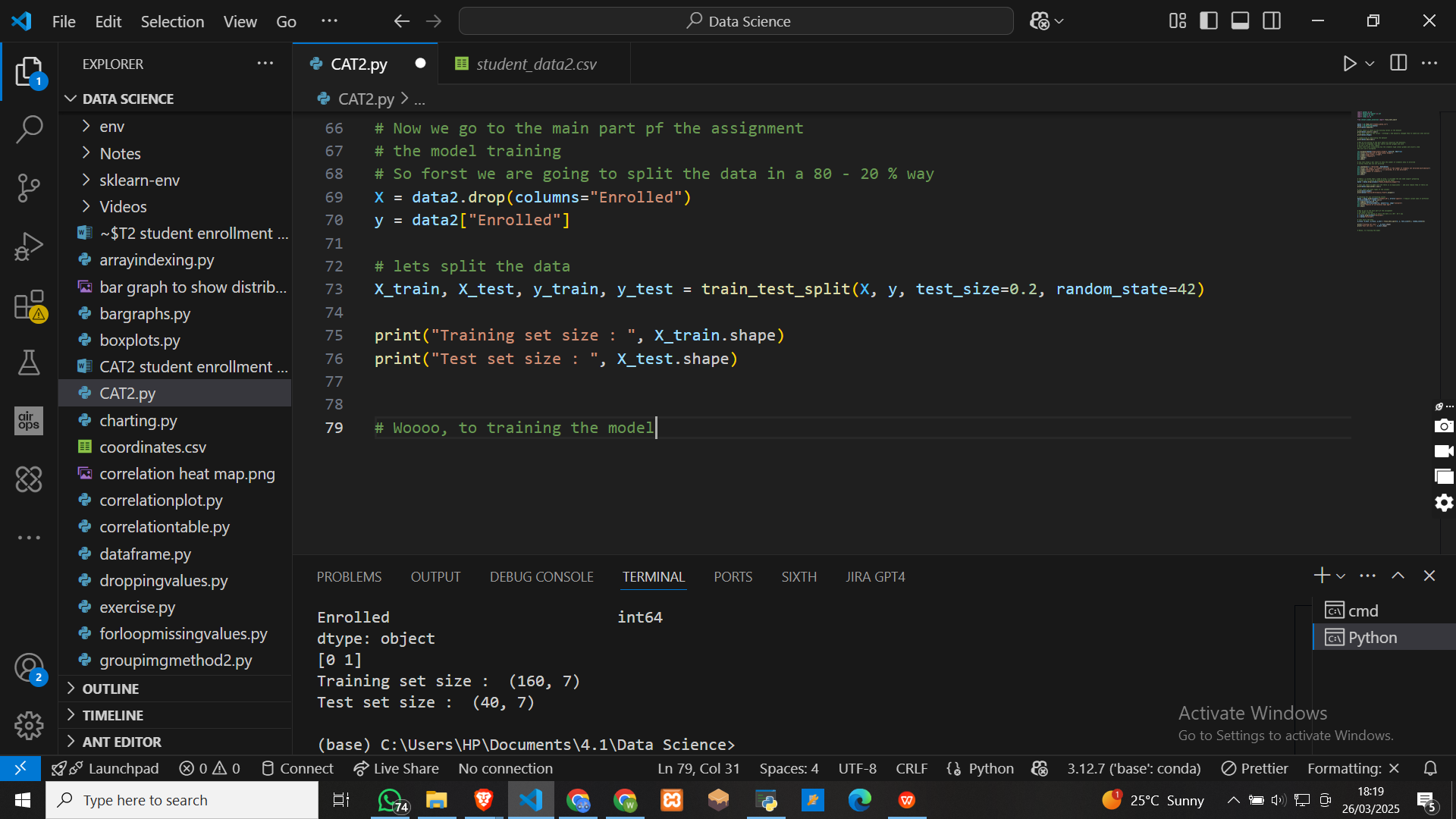
# So forst we are going to split the data in a 80 - 20 % way

I chose to use logical regression because:  
1. its simple to use and interpret

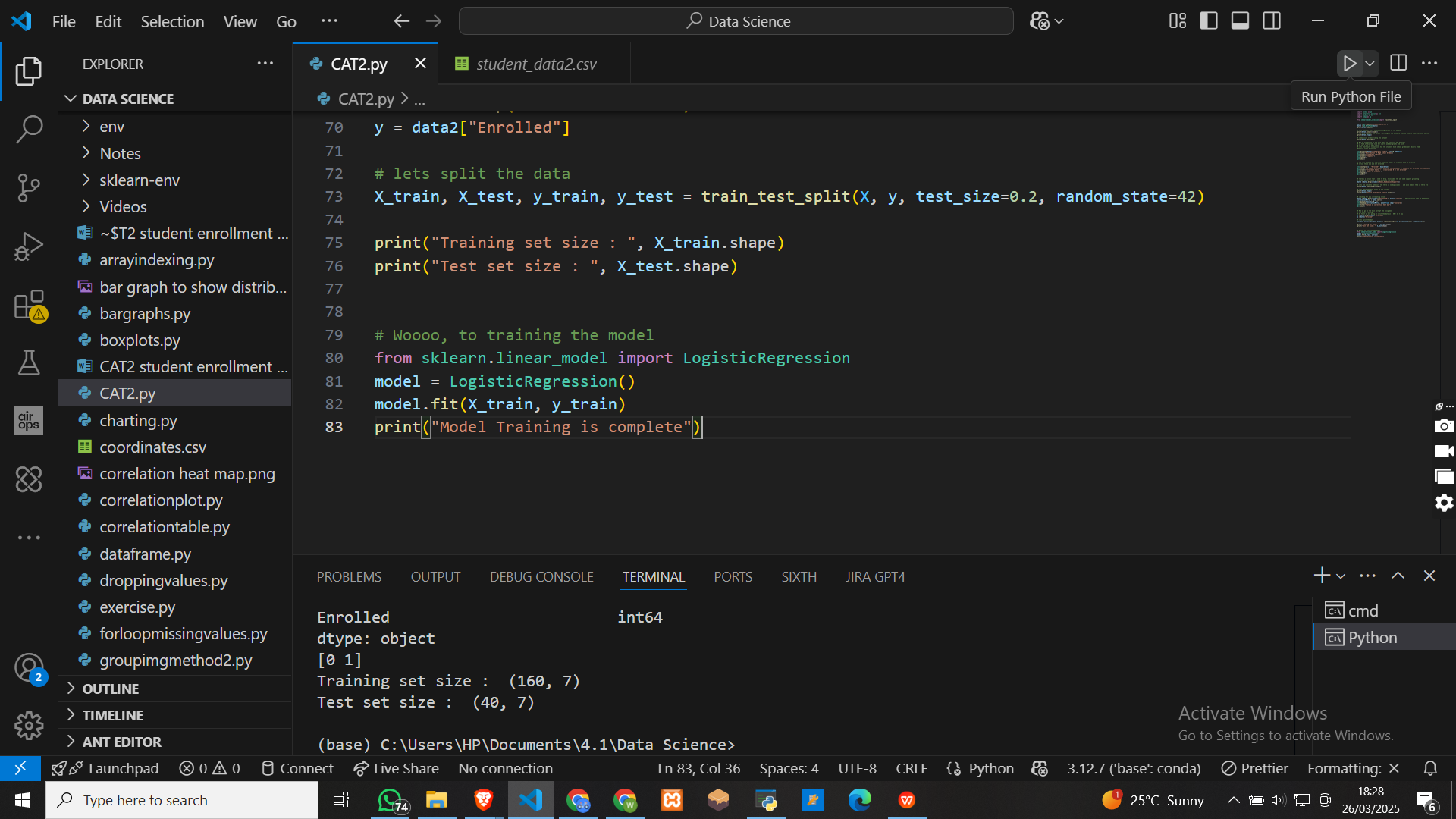
1. It works very we;l with structured data and since my data is numerical, it is best fpor this
2. It outputs the probabilities which is good for us to know those who are probable to take a certain course

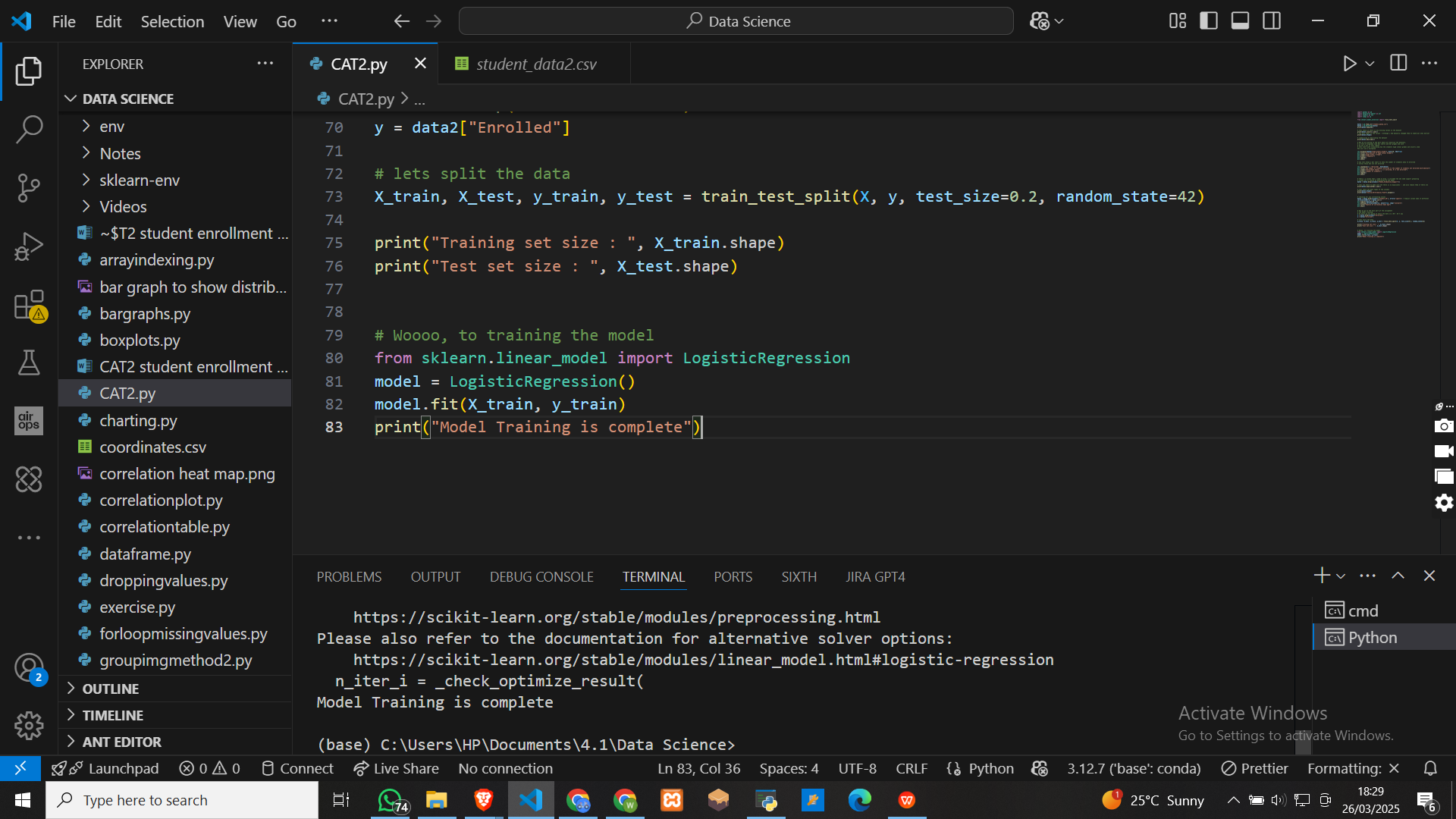


To splitting the data

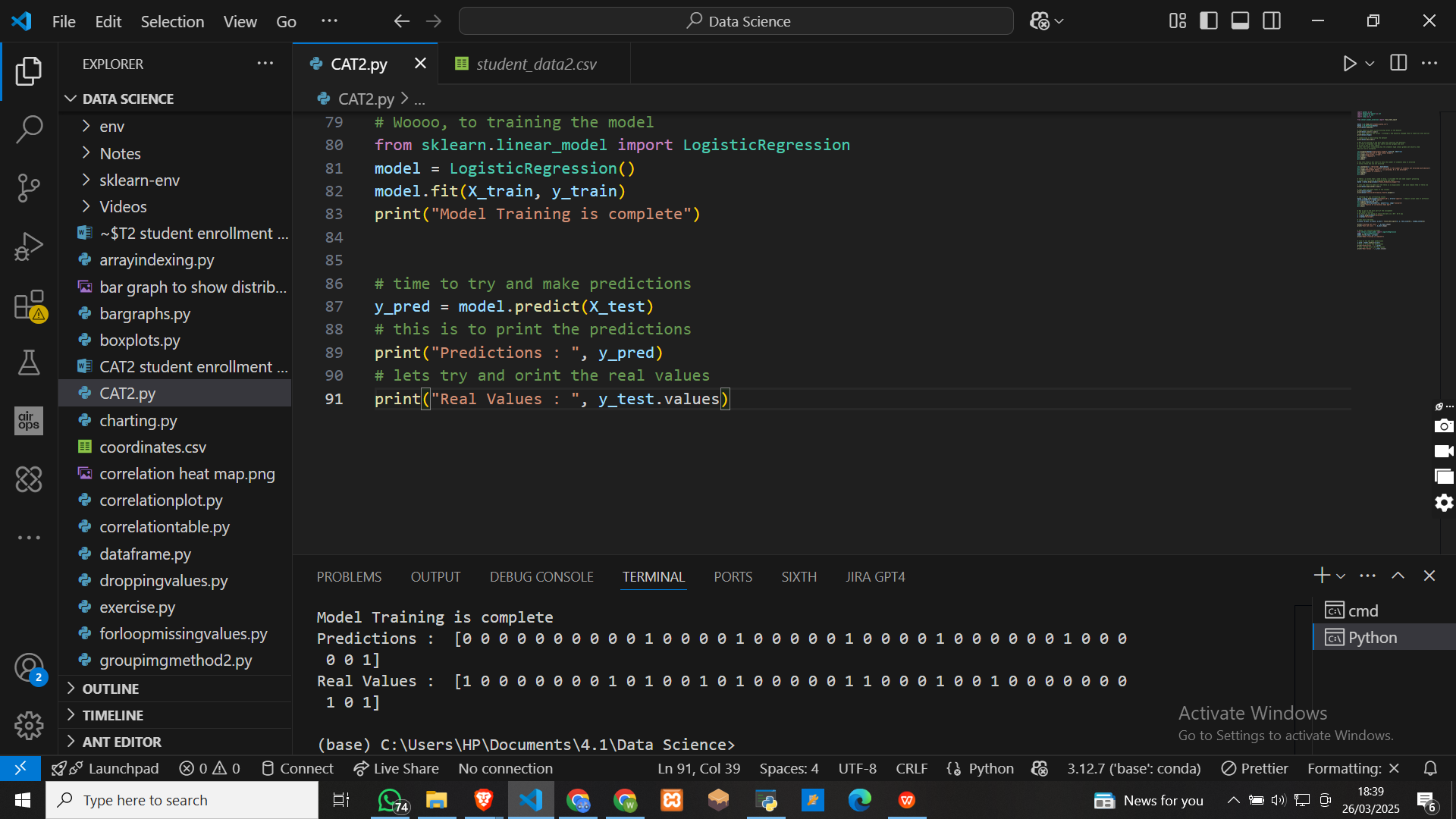


**Trainingthe model**

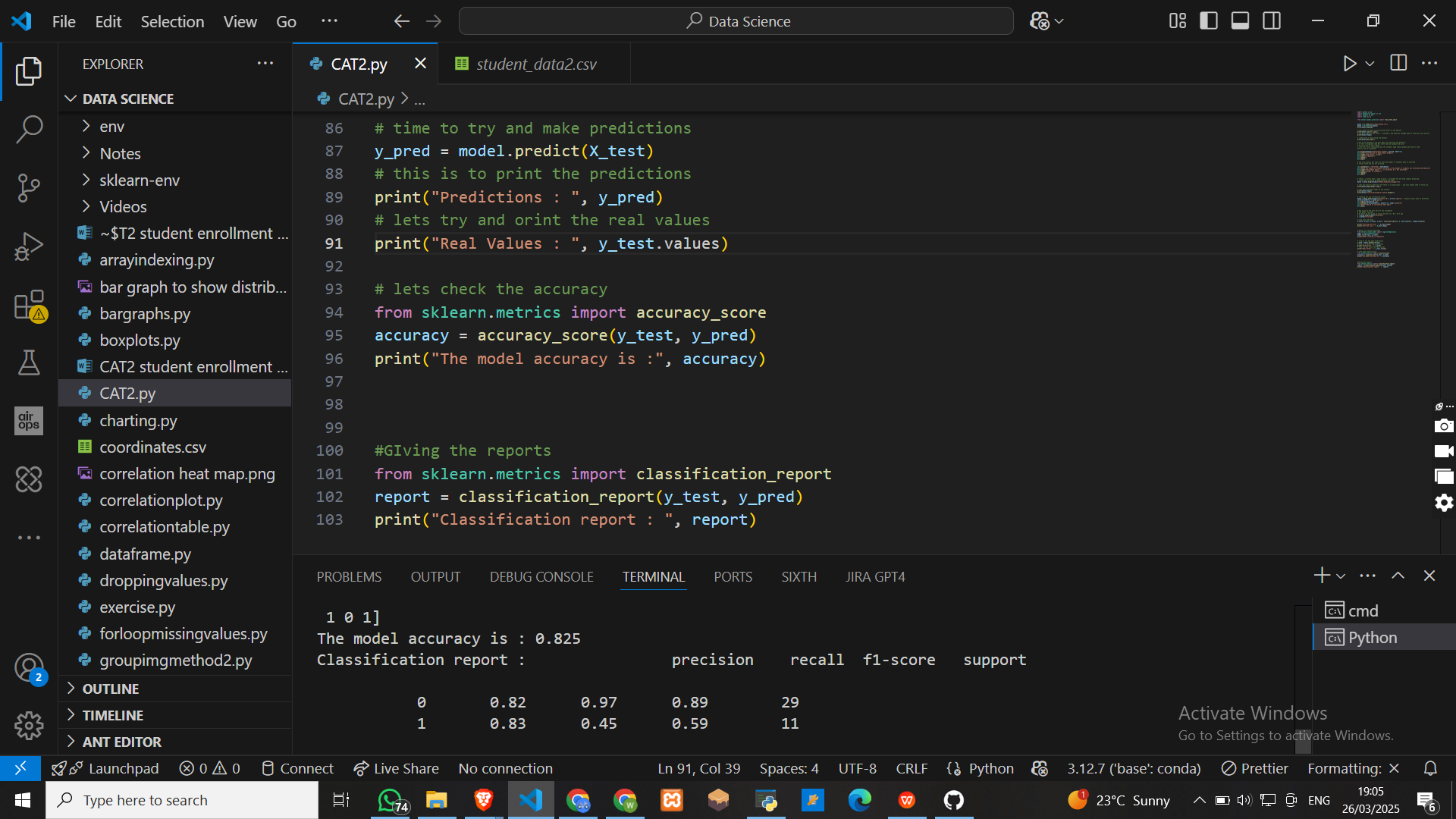




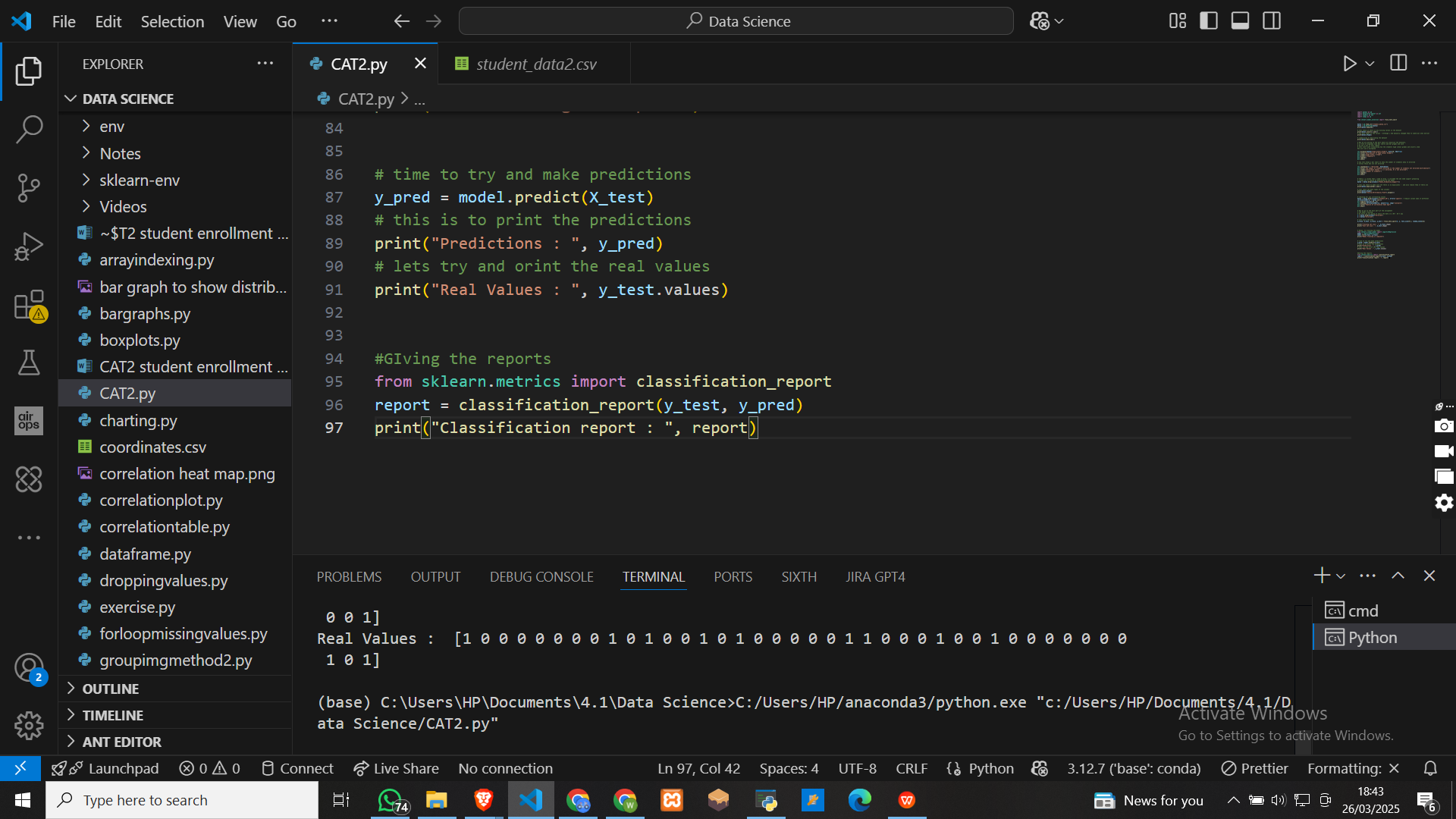
Making predictions



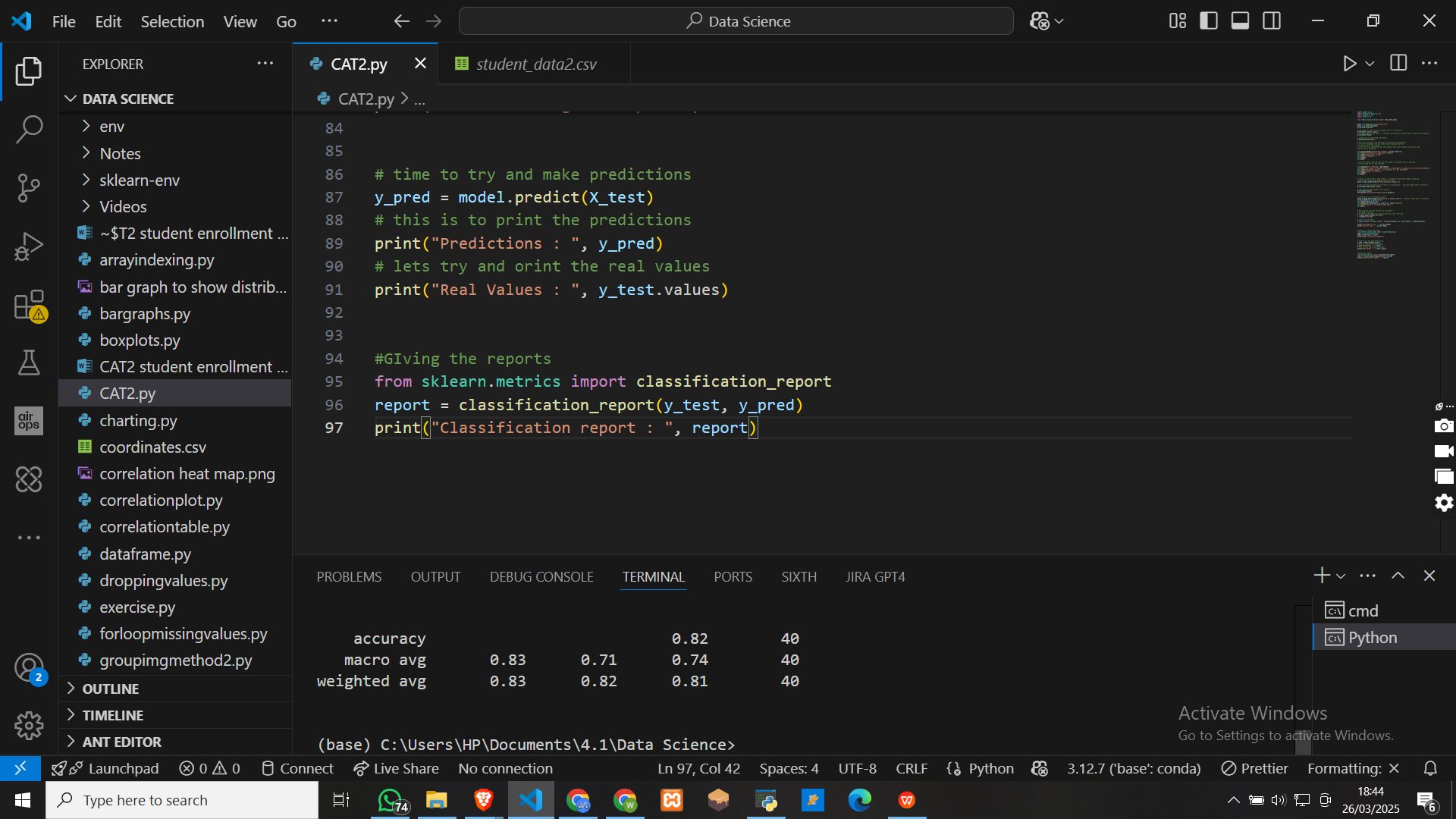
Accuracy report



Classification report



Amd the results are



**QUESTION 3**

How Can you protect the privacy of thr student data while still using it to develop predicitive models?

1. Anonymizing data by removing the personal Identifiable data or information like the full names, the emails, registration numbers, among others
2. Replacing the important details with pseudo names that don’t relate pr refer to the original information
3. Restrivting the student data to only verfied and authorized people
4. Including multi factor authentication to regulates those who have access to that information
5. Making sure that your models are in compliance with data protection laws and regulations put in place
6. Encrypting that data to regulate those whi have access to it.

**QUESTION 4**

How can you communicate the results of your model to educational institutions in a way that is actionable and informative

1. One way is to make clear and simple visualizations like bar charts, histograms, and heat maps to show the results.
2. Create a good report or maybe a dashboard that the institutions can see see and understand easily.
3. Use natural summaries like in sentences for example, 75 percent of students who score 80 percent and above in computer studies in high school have a high probability of enrolling in the course.
4. Availing yourself for face to face presentation and questions and answers session.