**Importing the Libraries:**

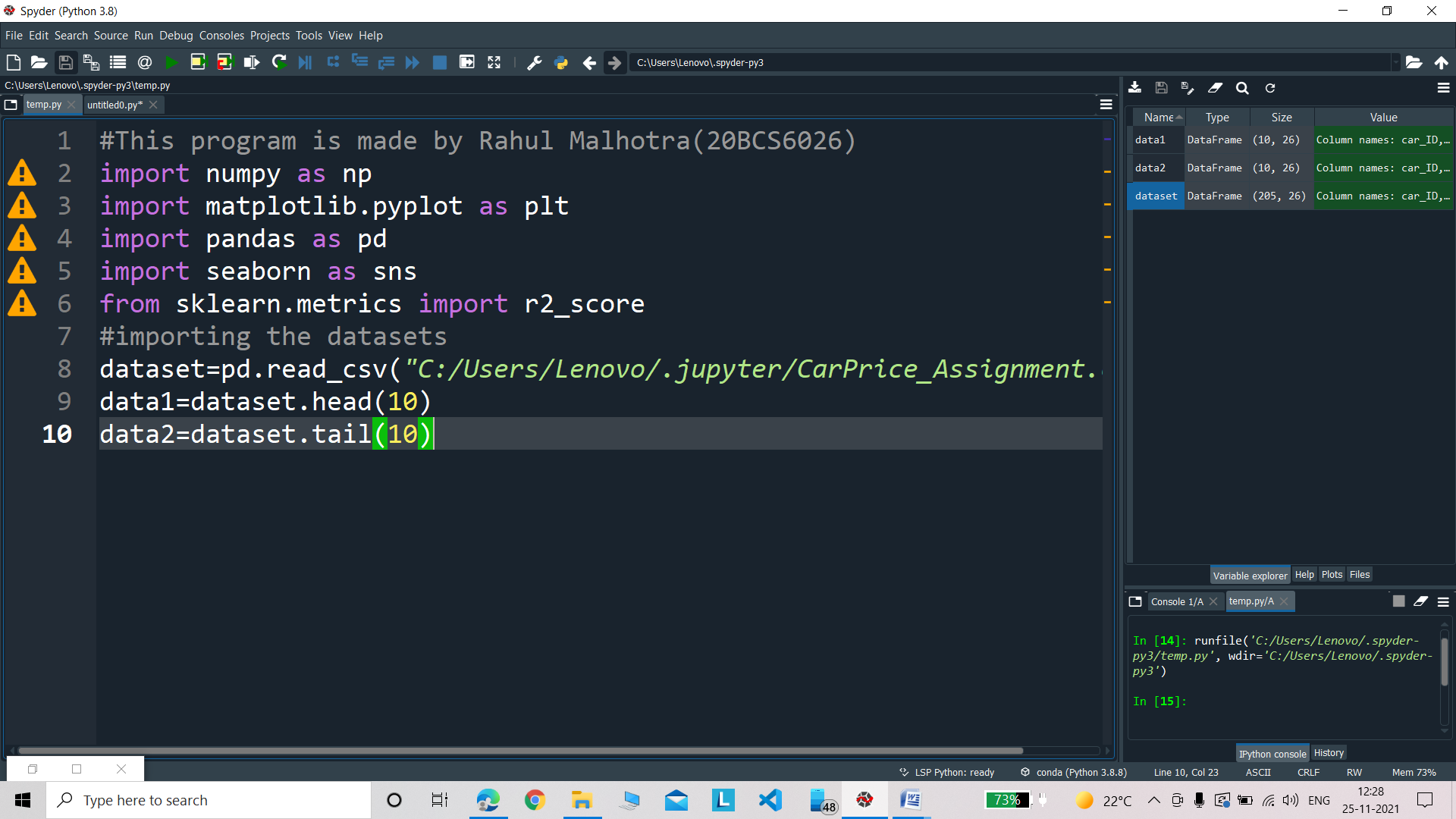
import numpy as np

import matplotlib.pyplot as plt

import pandas as pd

import seaborn as sns

from sklearn.metrics import r2\_score

****

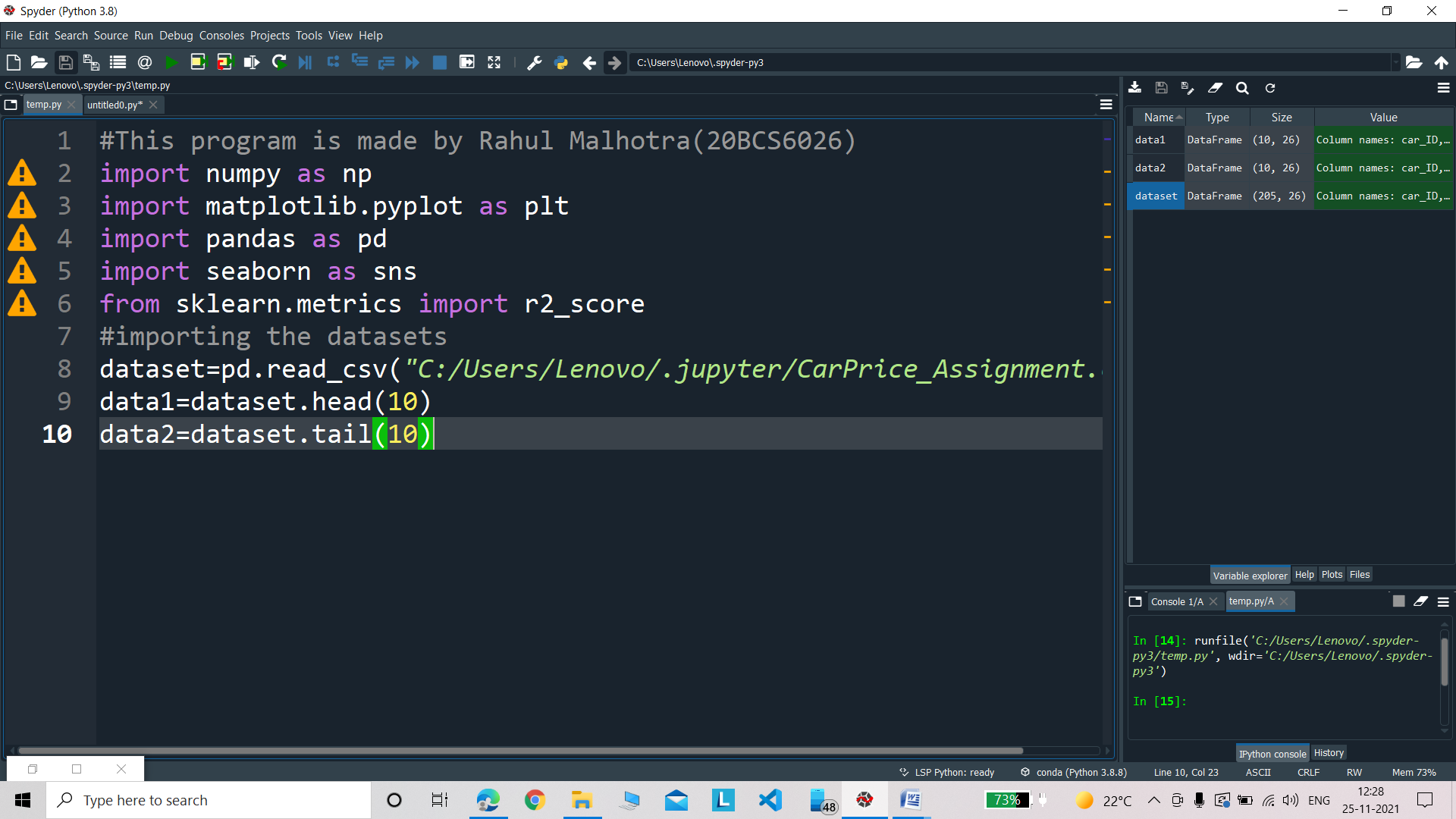
**Importing the datasets**

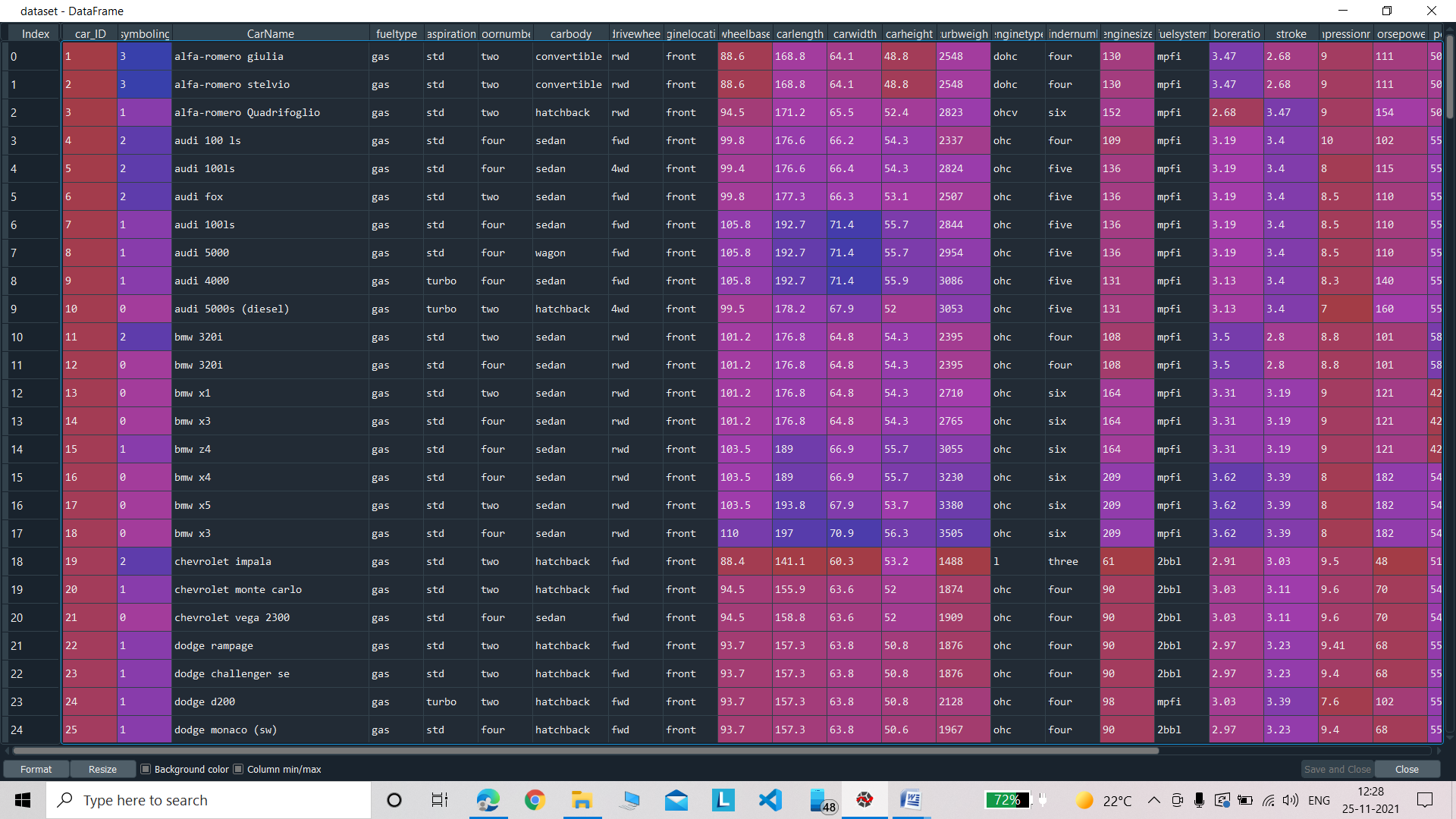
#importing the datasets

dataset=pd.read\_csv("C:/Users/Lenovo/.jupyter/CarPrice\_Assignment.csv")

data1=dataset.head(10)

data2=dataset.tail(10)

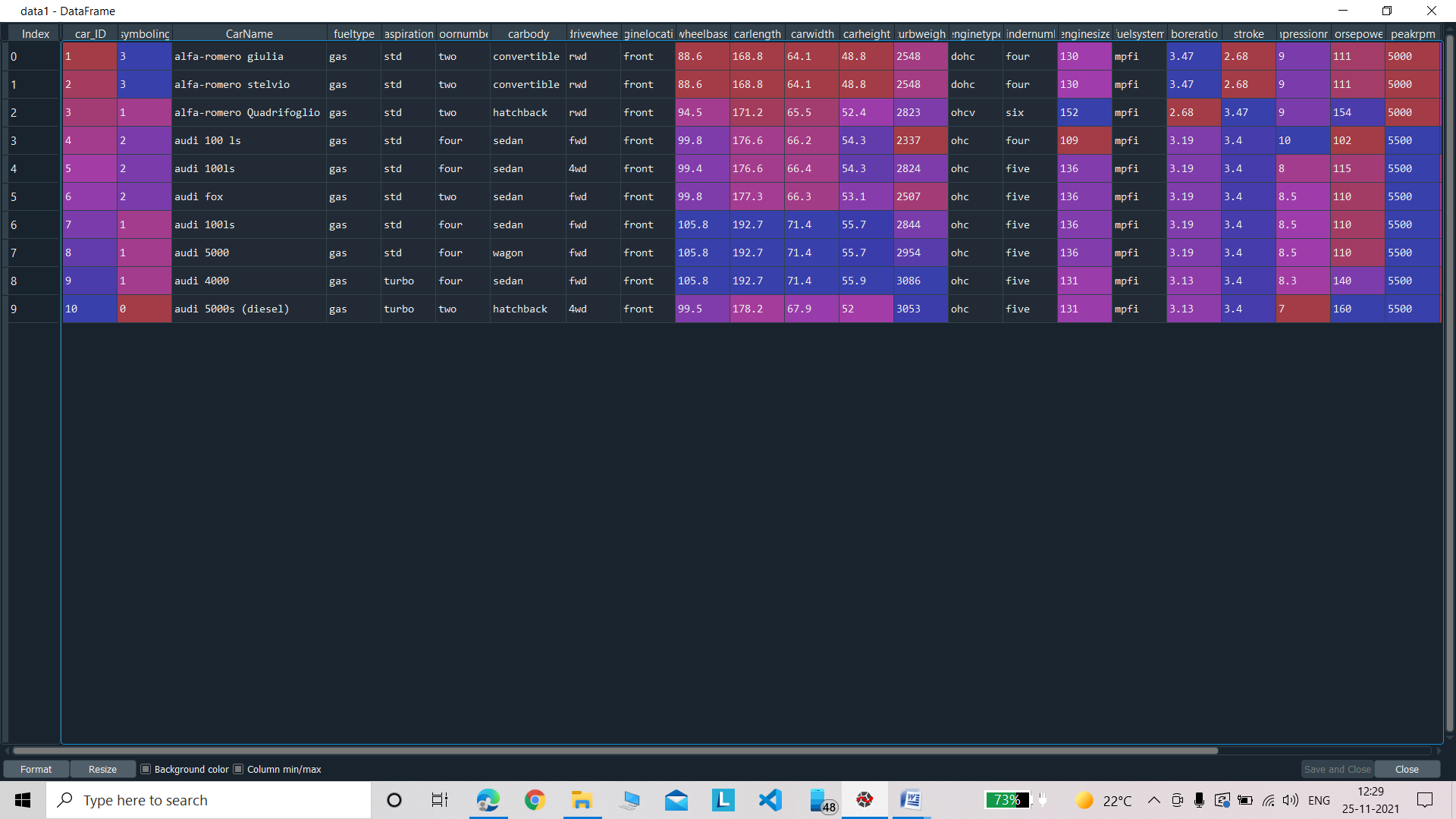


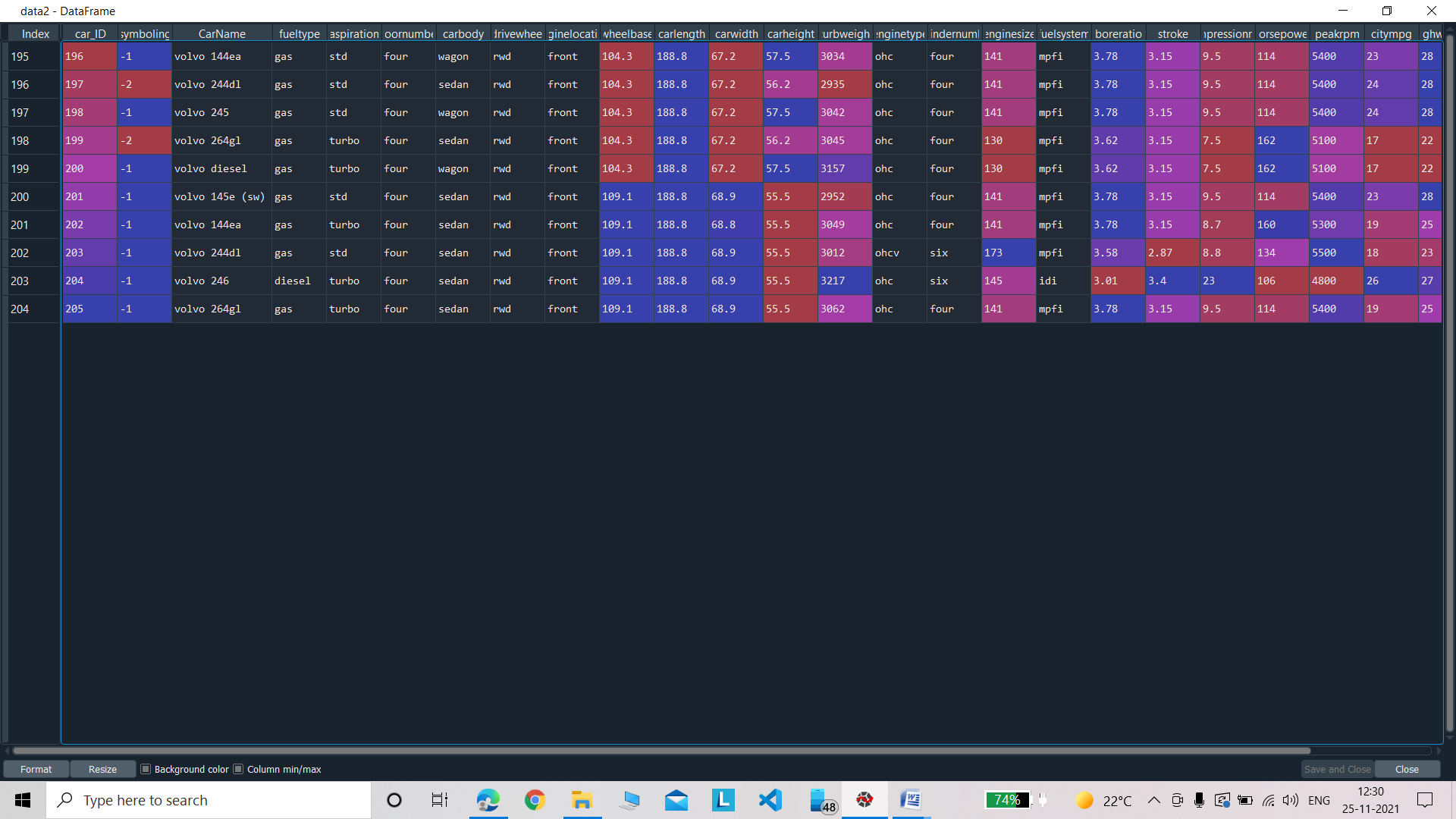


**Analysing the first and last values of the dataset:**

data1=dataset.head(10)

data2=dataset.tail(10)

****

****

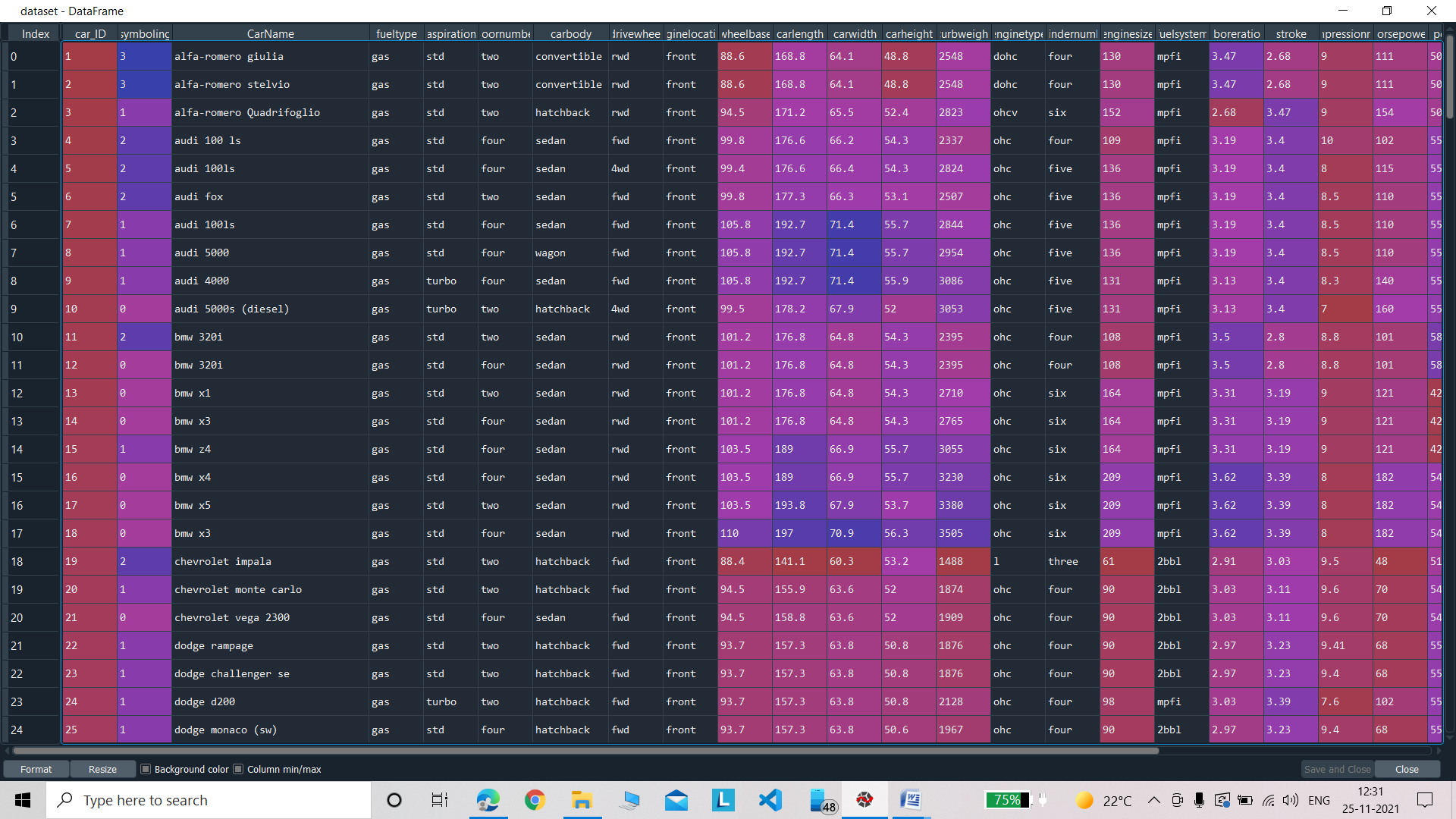
**Dropping the duplicate and null values of the dataset:**

#dropping the duplicates

dataset = dataset.drop\_duplicates()

#dropping the Null Values

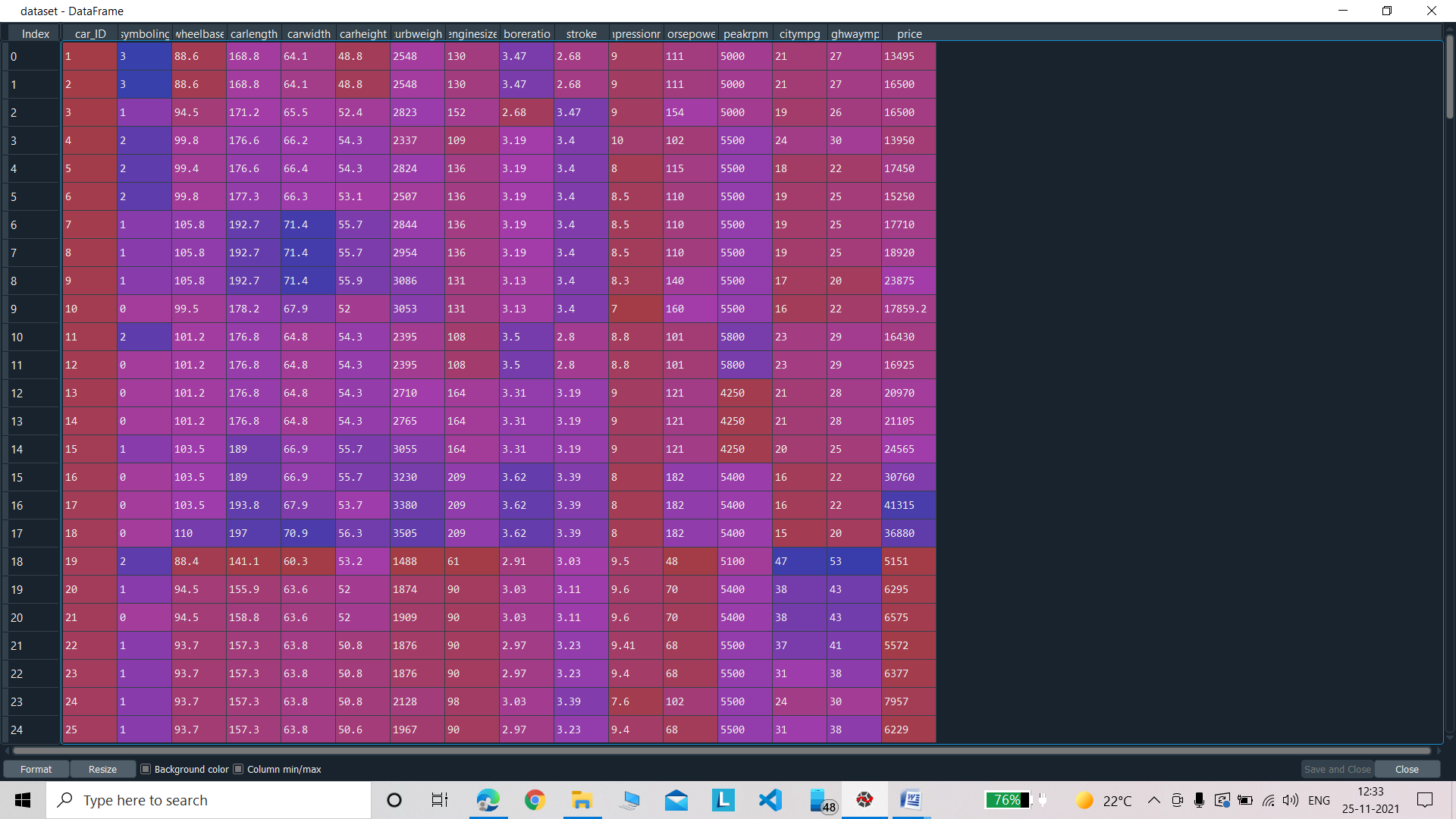
dataset = dataset.dropna()



**Dropping** **the non important independent variables:**

#dropping the non important independent variables

dataset = dataset.drop(['CarName', 'fueltype', 'aspiration', 'doornumber', 'carbody', 'drivewheel','enginelocation','enginetype','cylindernumber','fuelsystem'], axis=1



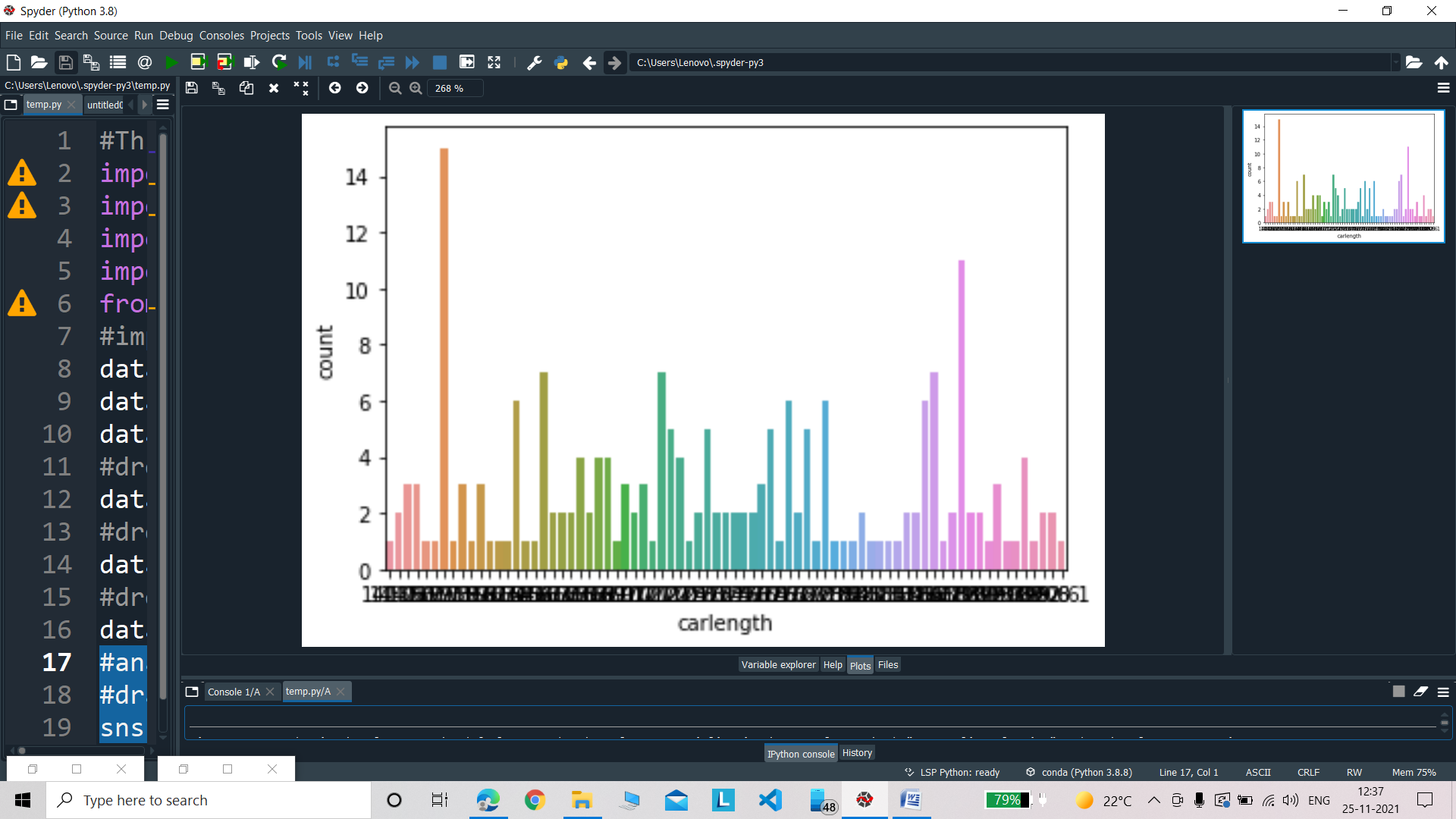
**Analyzing the dataset visually:**

**Drawing the countplot:**

#analyzing the dataset visually

#drawing the countplot

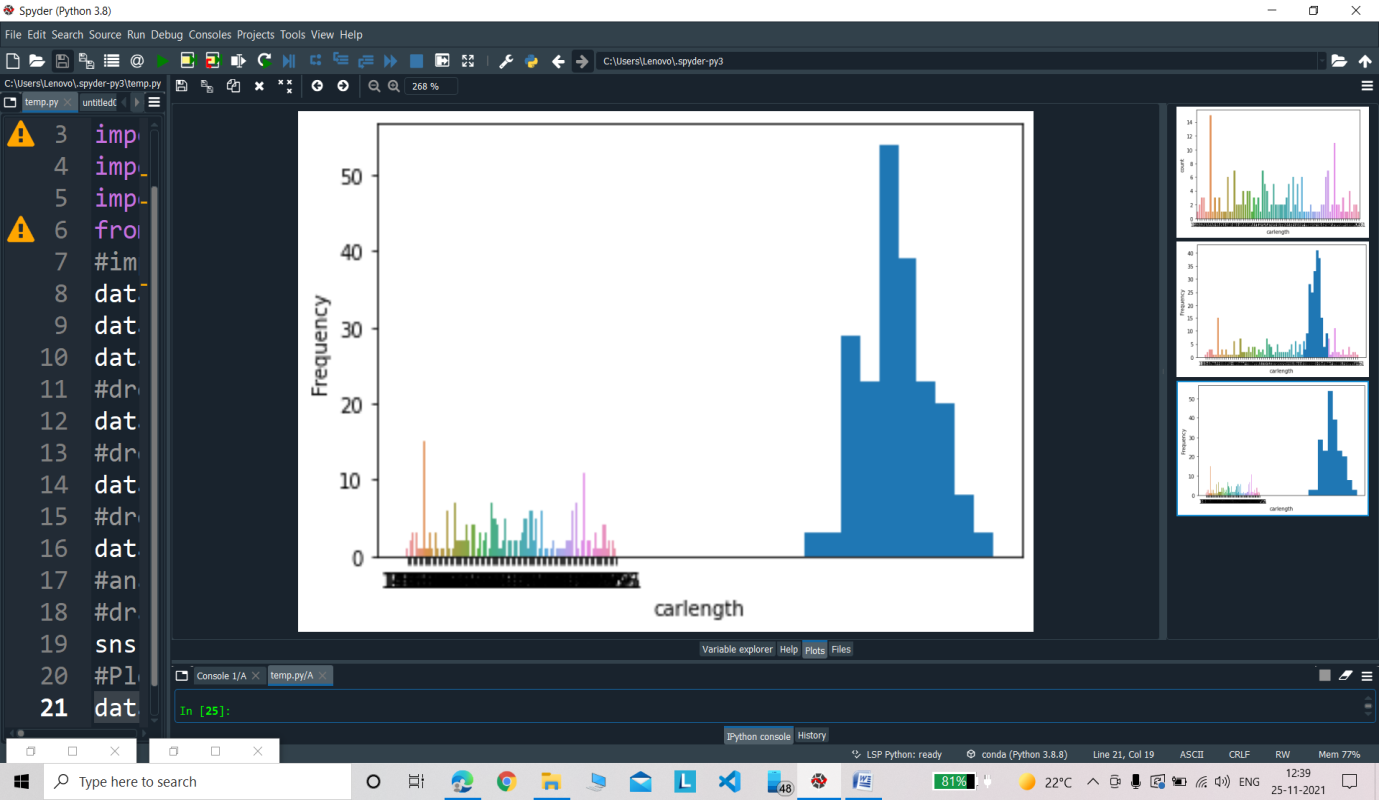
sns.countplot(x="carlength", data=dataset)

****

**Plotting the Histogram:**

#Plotting the Histogram

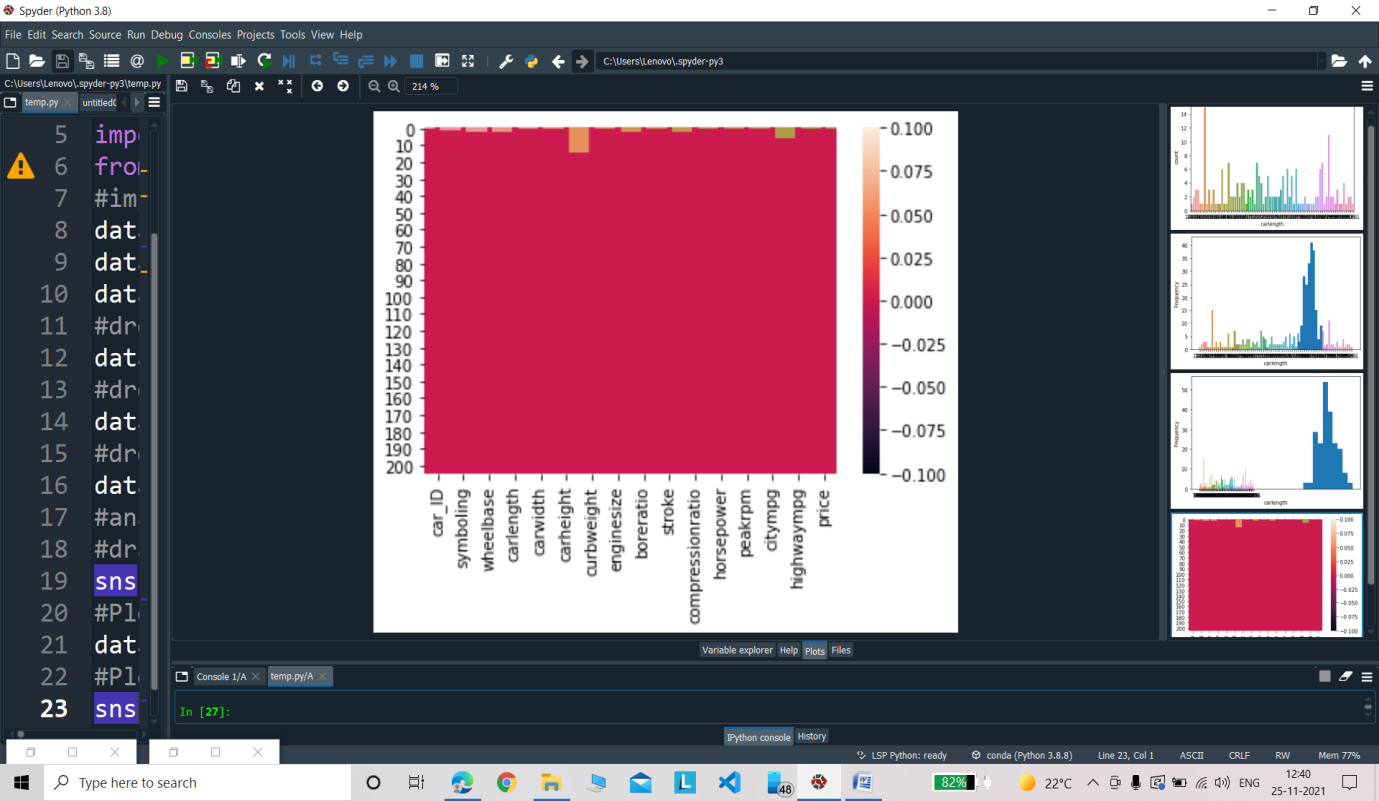
dataset["carlength"].plot.hist()

****

**Plotting the Heatmap:**

#Plotting the Heatmap

#sns.heatmap(dataset.isnull())

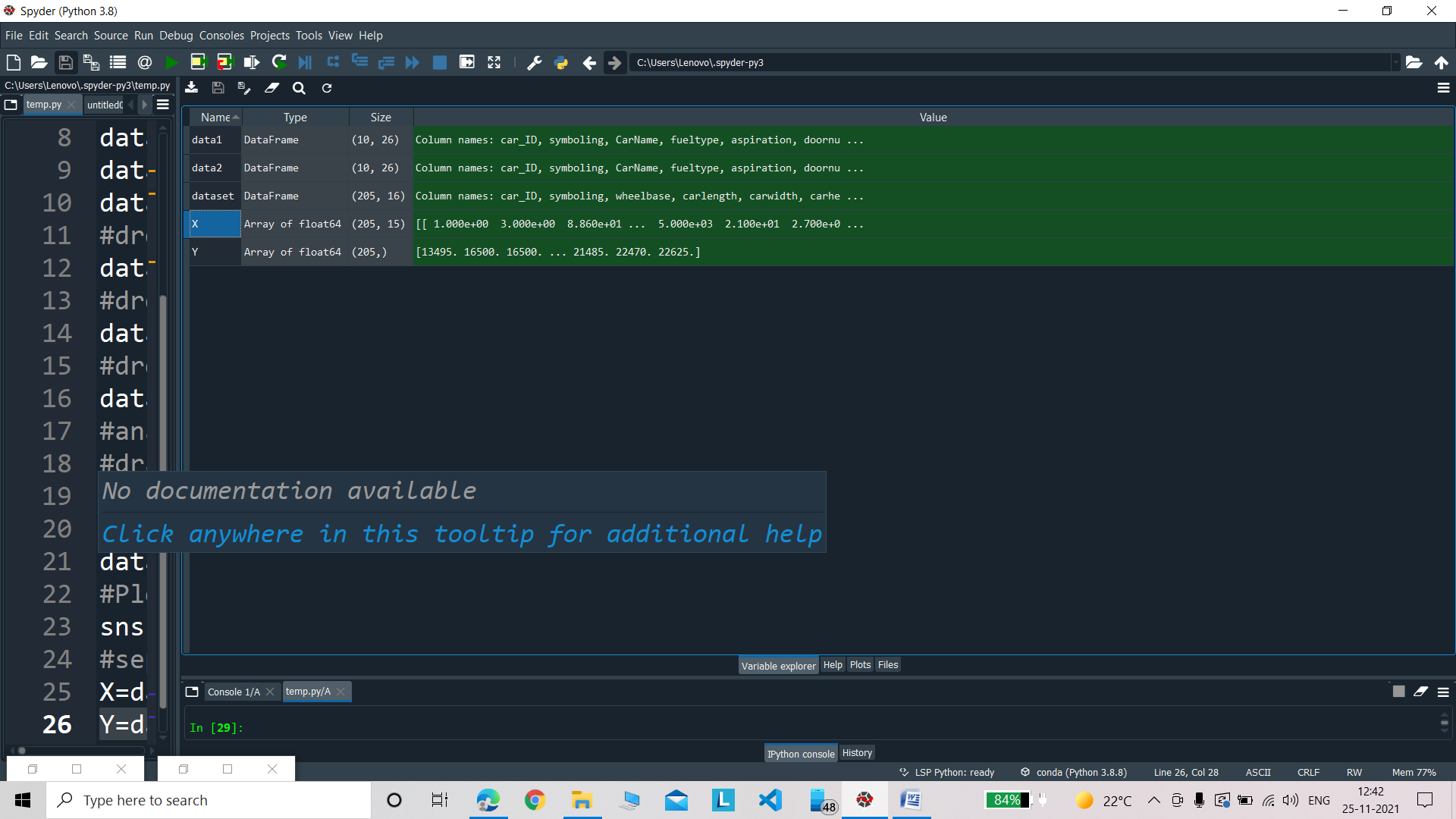
****

**Seperate the dependet and the independent variables**

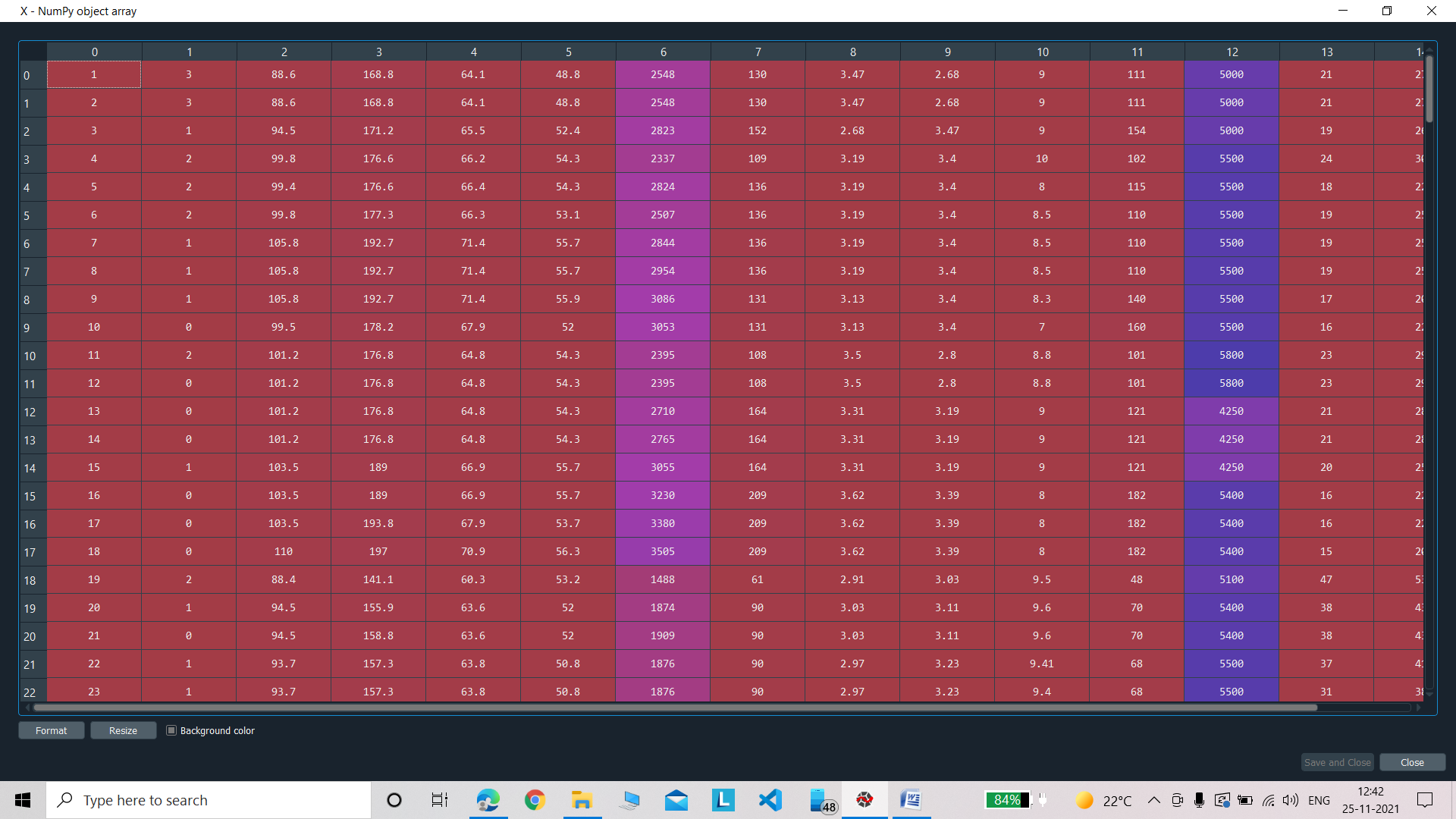
#seperate the dependet and the independent variables

X=dataset.iloc[:,:-1].values

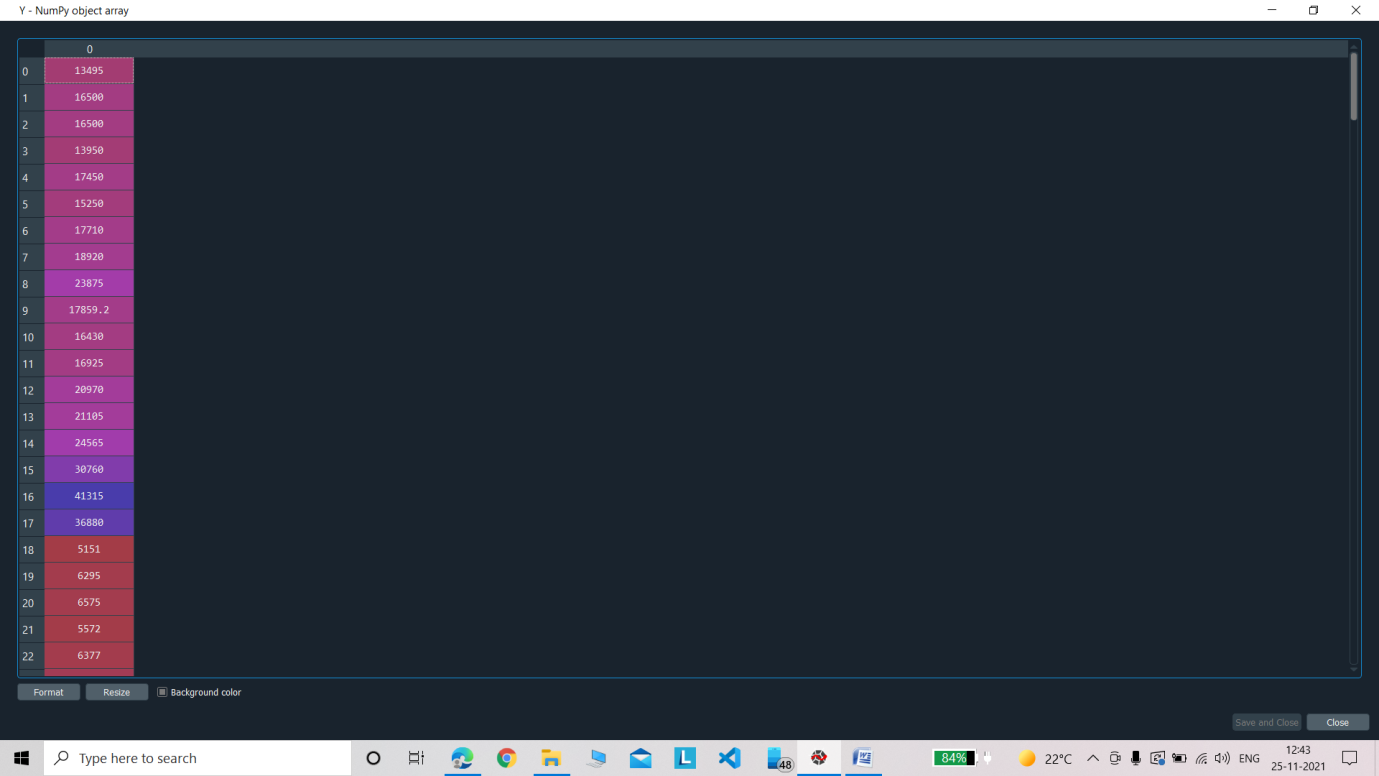
Y=dataset.iloc[:,-1].values



**Independent Variables:**



**Dependent Variables:**

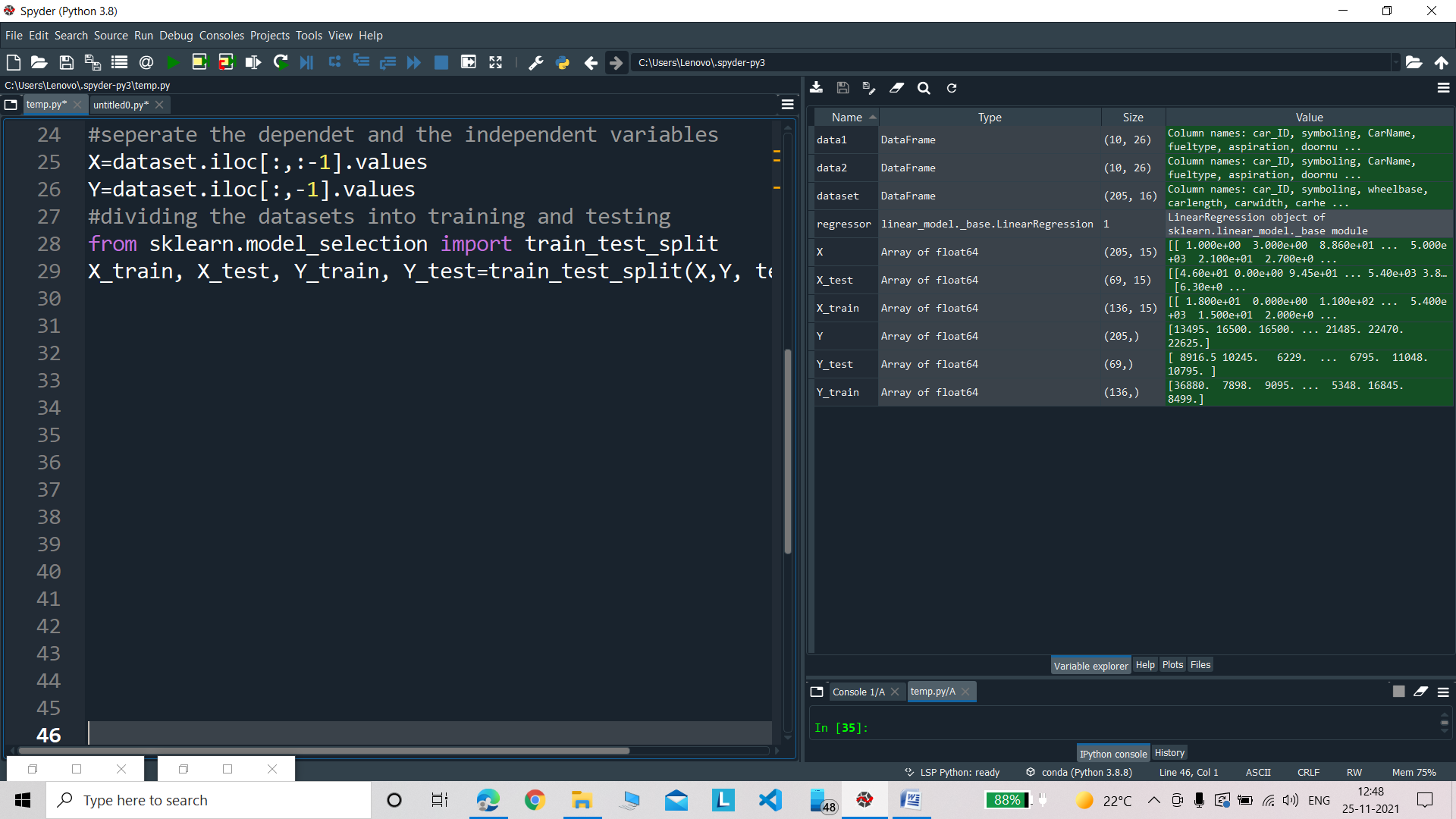


**Dividing the datasets into training and testing**

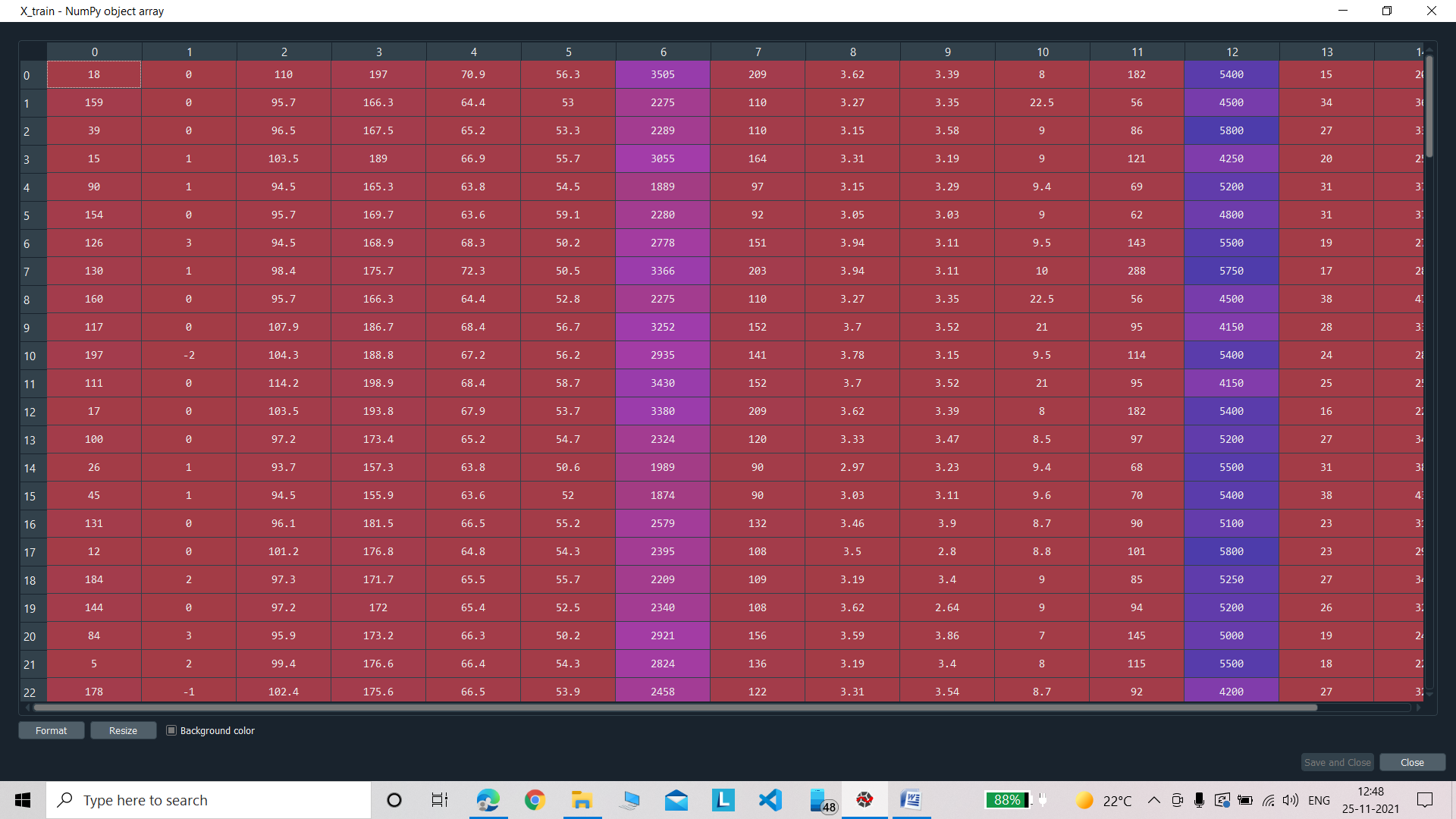
#dividing the datasets into training and testing

from sklearn.model\_selection import train\_test\_split

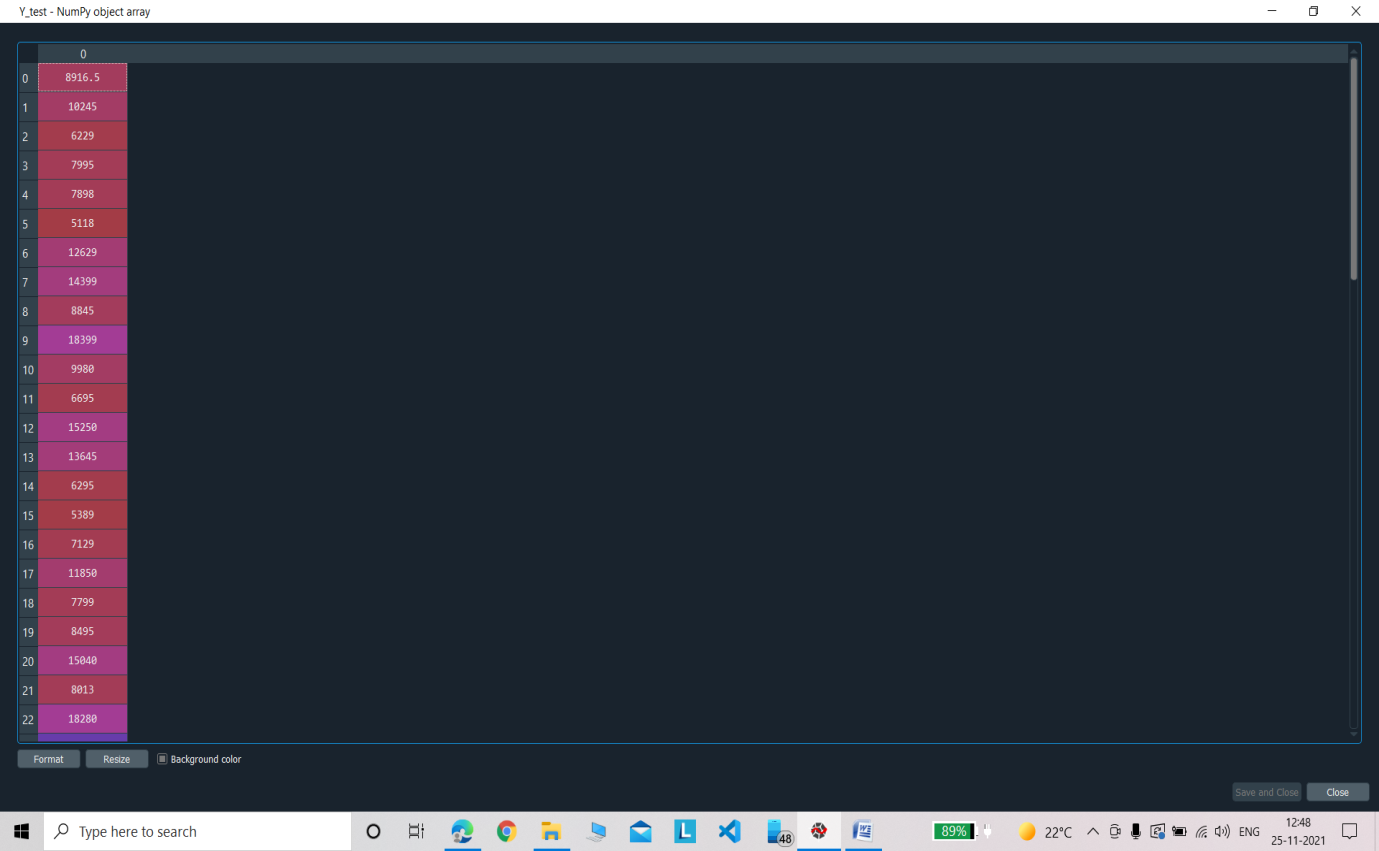
X\_train, X\_test, Y\_train, Y\_test=train\_test\_split(X,Y, test\_size=1/3)

****

**Training Dataset:**

****

**Testing Dataset:**

****

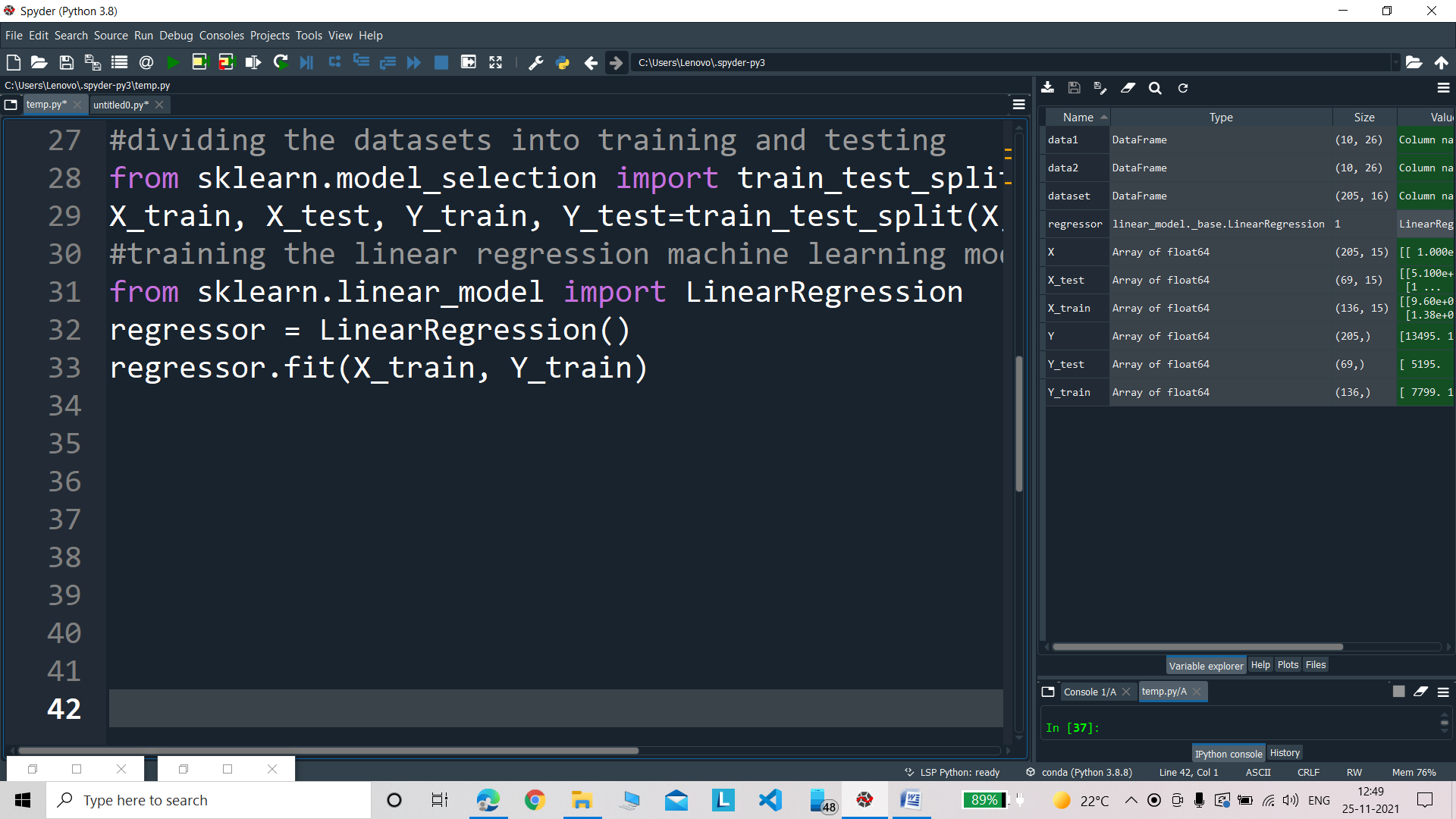
**Training the linear regression machine learning model using the training dataset**

#training the linear regression machine learning model using the training dataset

from sklearn.linear\_model import LinearRegression

regressor = LinearRegression()

regressor.fit(X\_train, Y\_train)

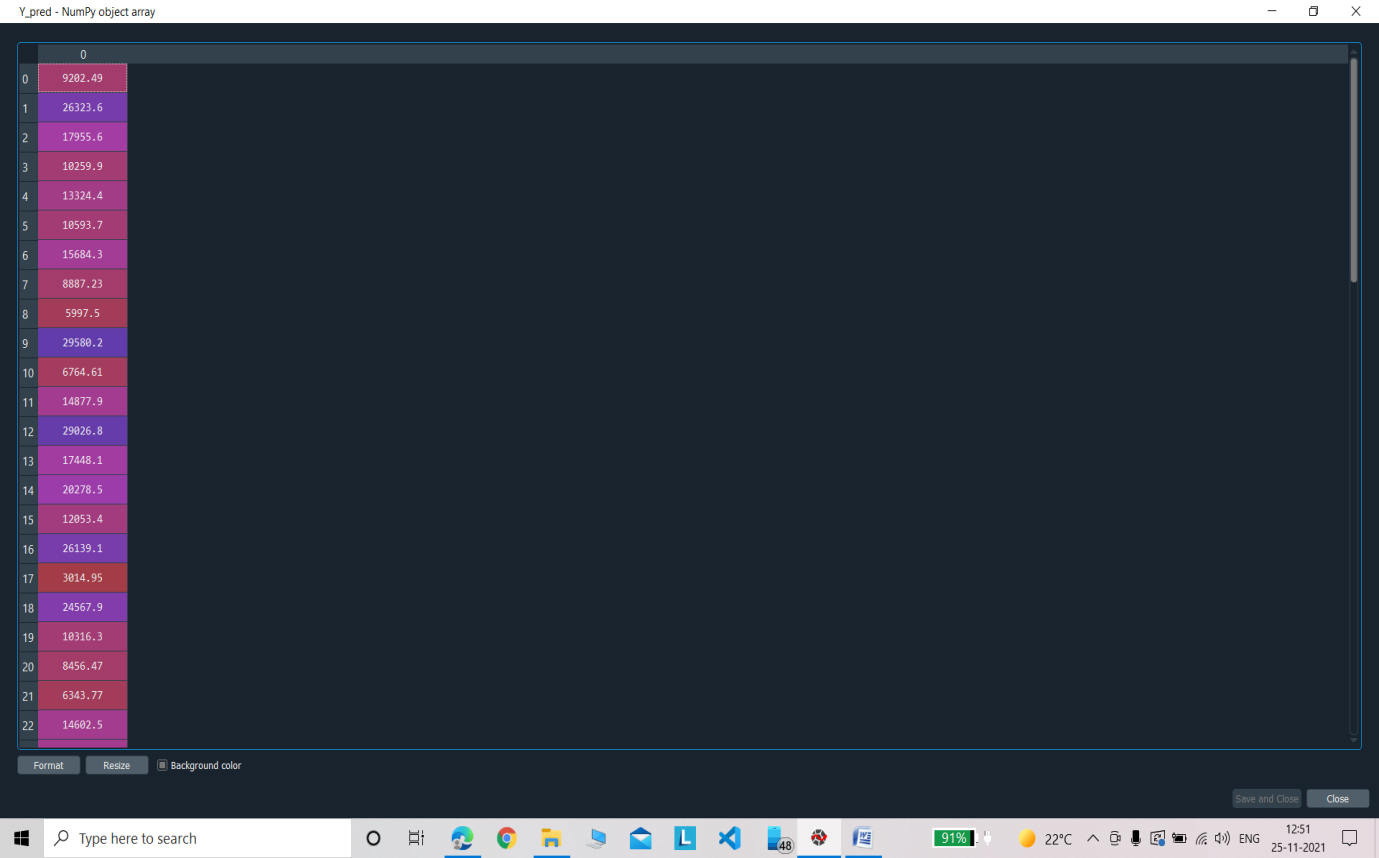


**Predicting the results for the testing datasets:**

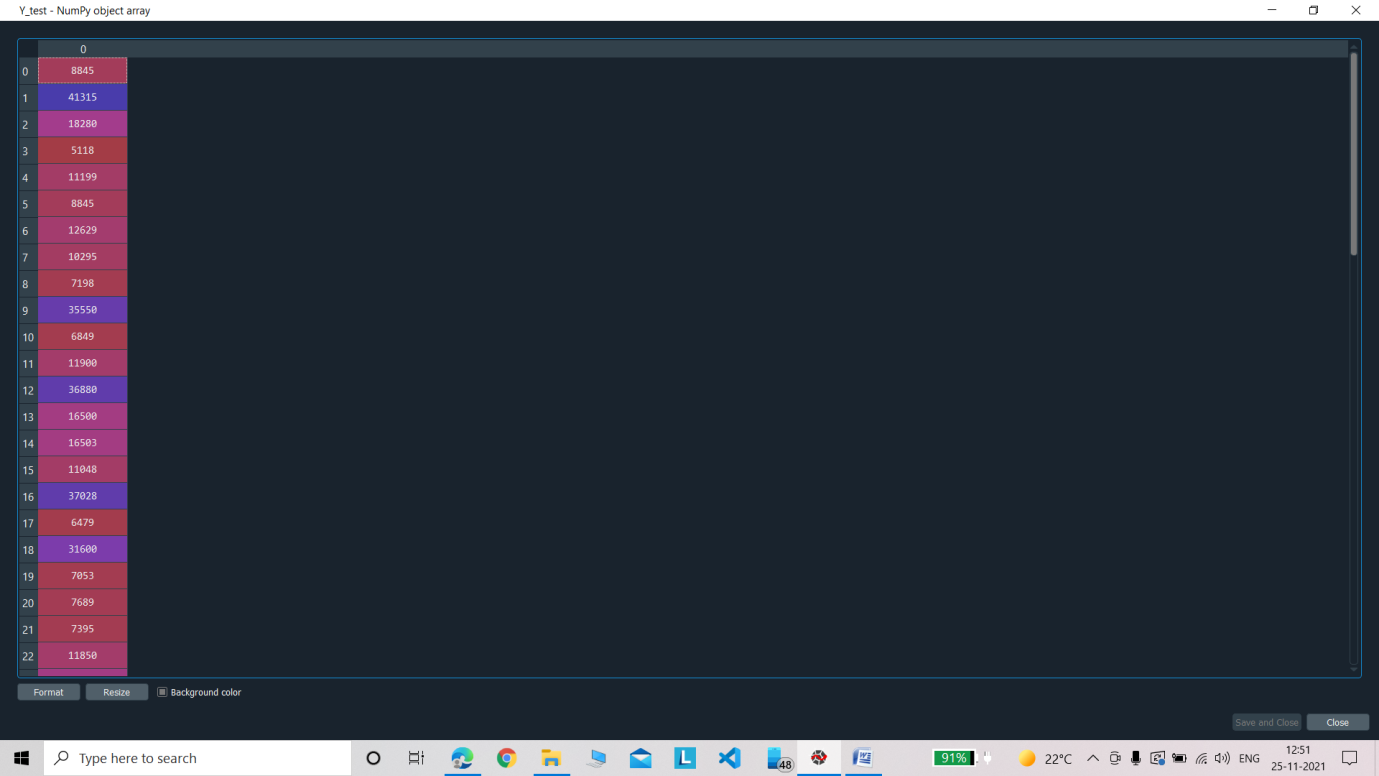
#now predicting the results for the testing datasets

Y\_pred=regressor.predict(X\_test)

**Predicted Values:**



**Testing Values:**

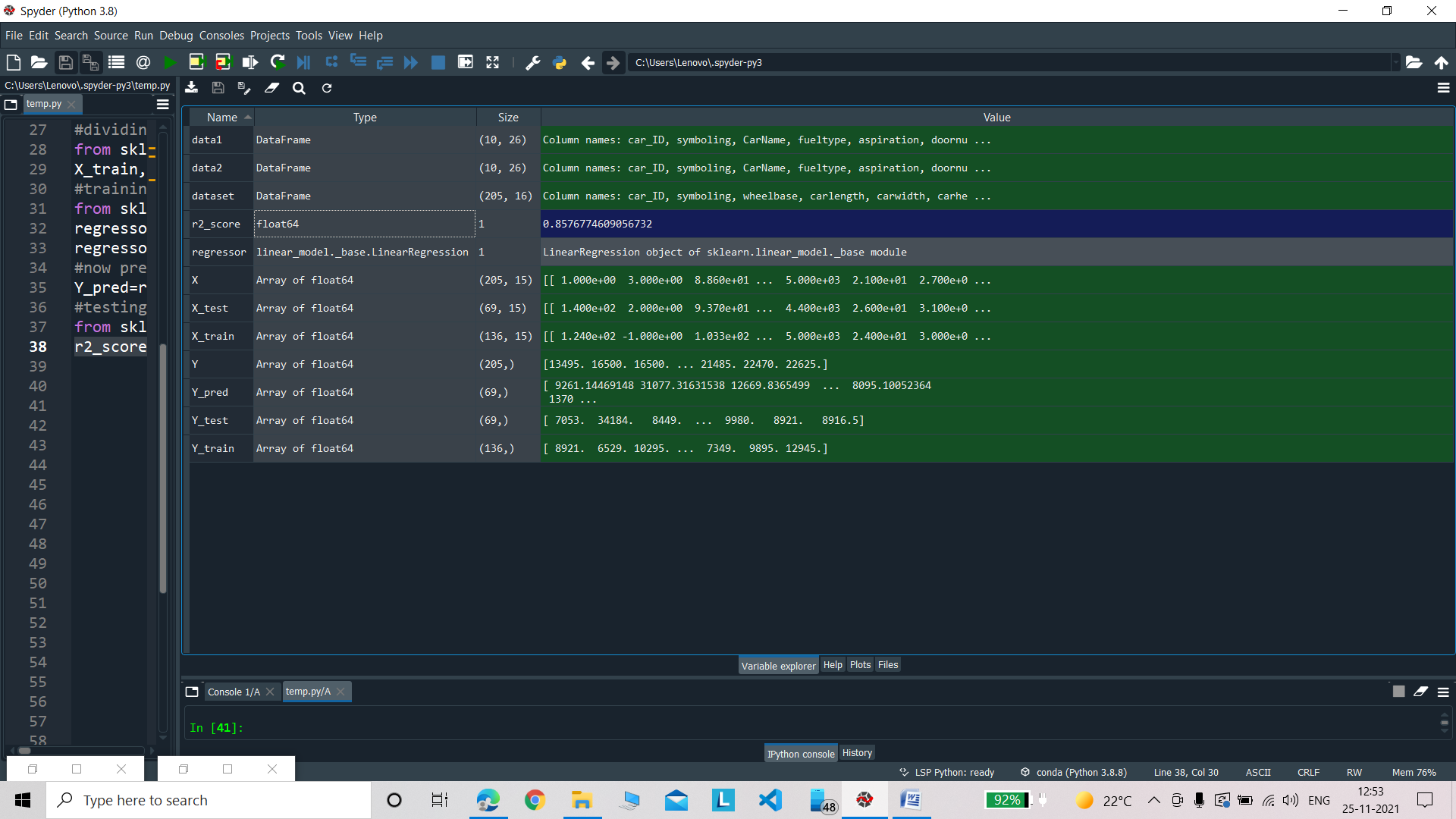


**Testing** **the accuray of the model using accuracy**

#testing the accuray of the model using accuracy

from sklearn.metrics import mean\_squared\_error

r2\_score=regressor.score(X,Y)



**R2\_Score:**

The R2\_Score of the model is 0.8576774609056732 hence our model is fairly accurate.