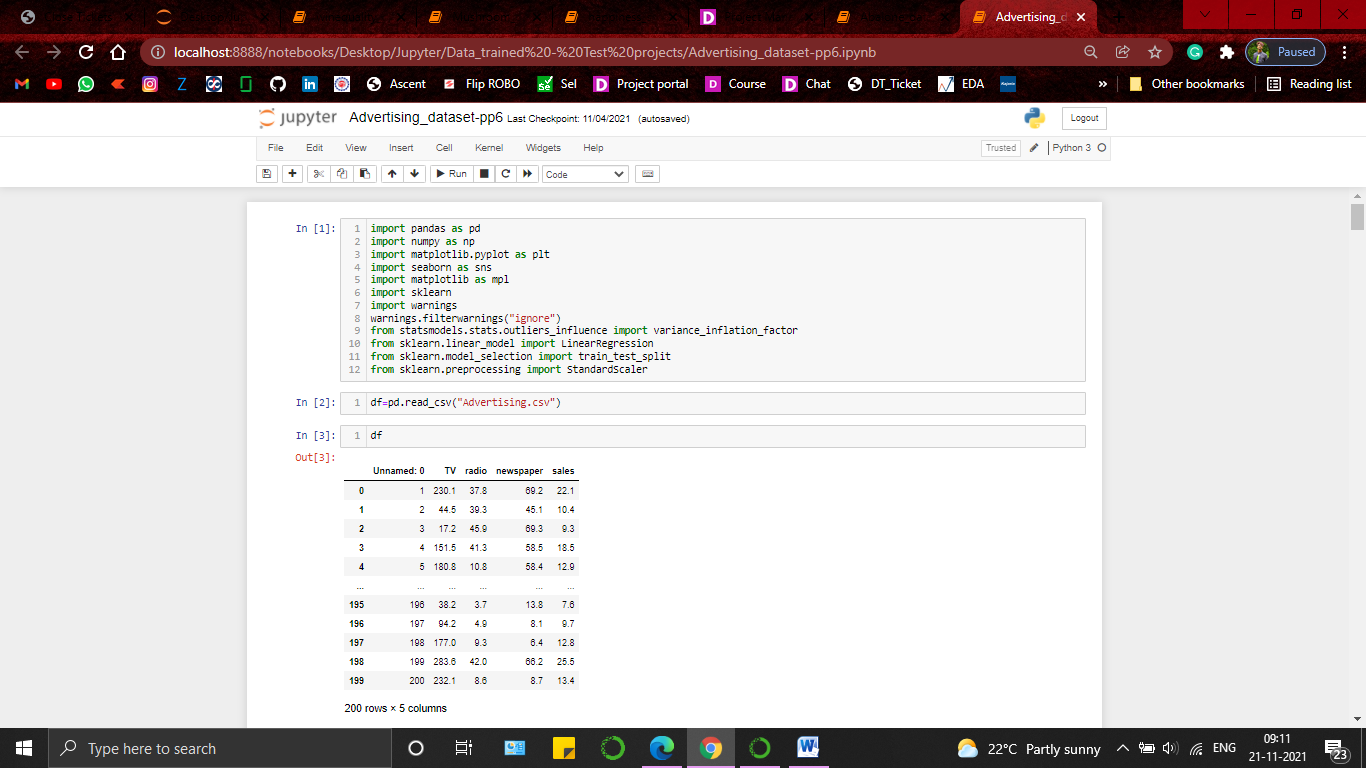
Advertisement dataset analysis

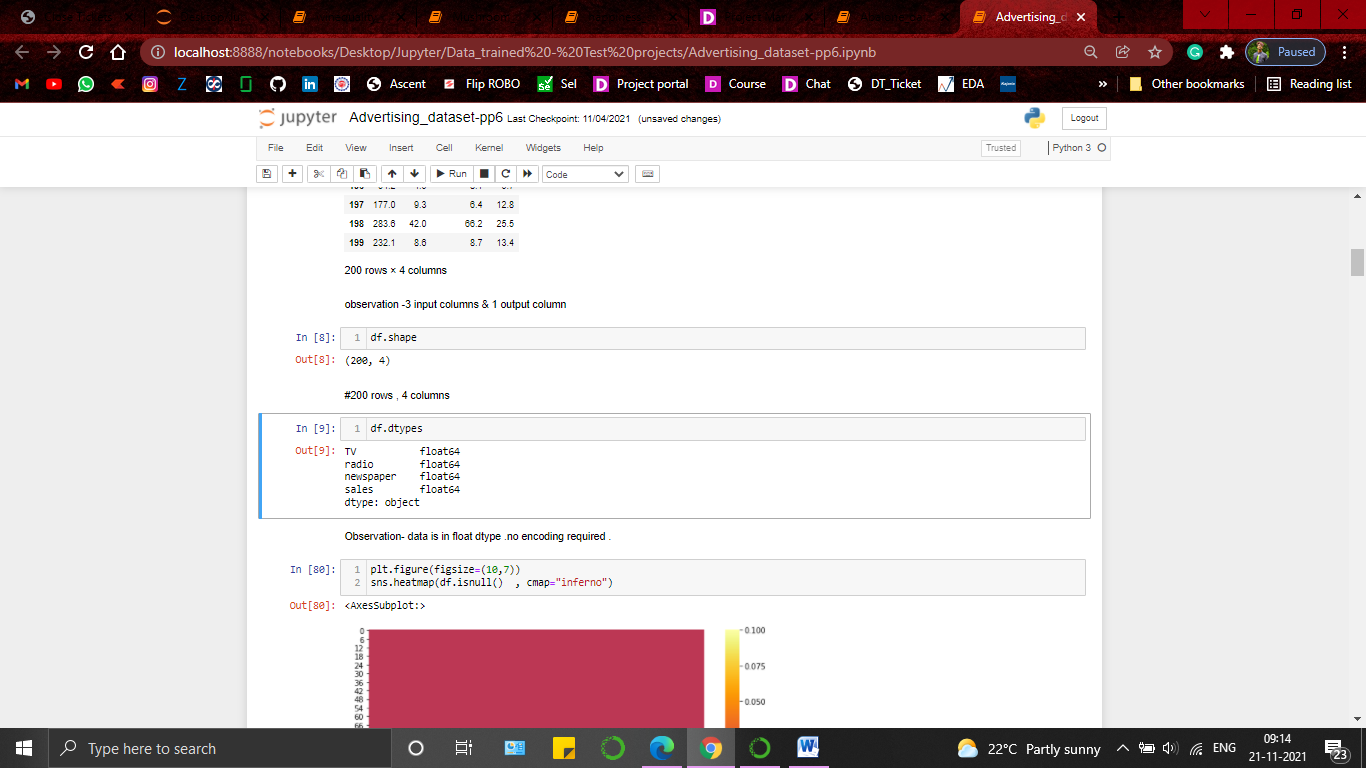
1. Problem Definition.

Every channel has a product or service to deliver and to do so they must acquire a good market to generate and leverage good sales . to Reach the customer company uses different platforms to advertise their product for a better reach but a company should have a better understanding about which platform is giving a better sales generation , basing which they can take further decision like whether to invest more on the high contributing platform or to improve the quality and reach of advertisements in other platforms . to do so we the company needs to analyse its existing sales generation data to predict the outcomes .

1. Data Analysis.

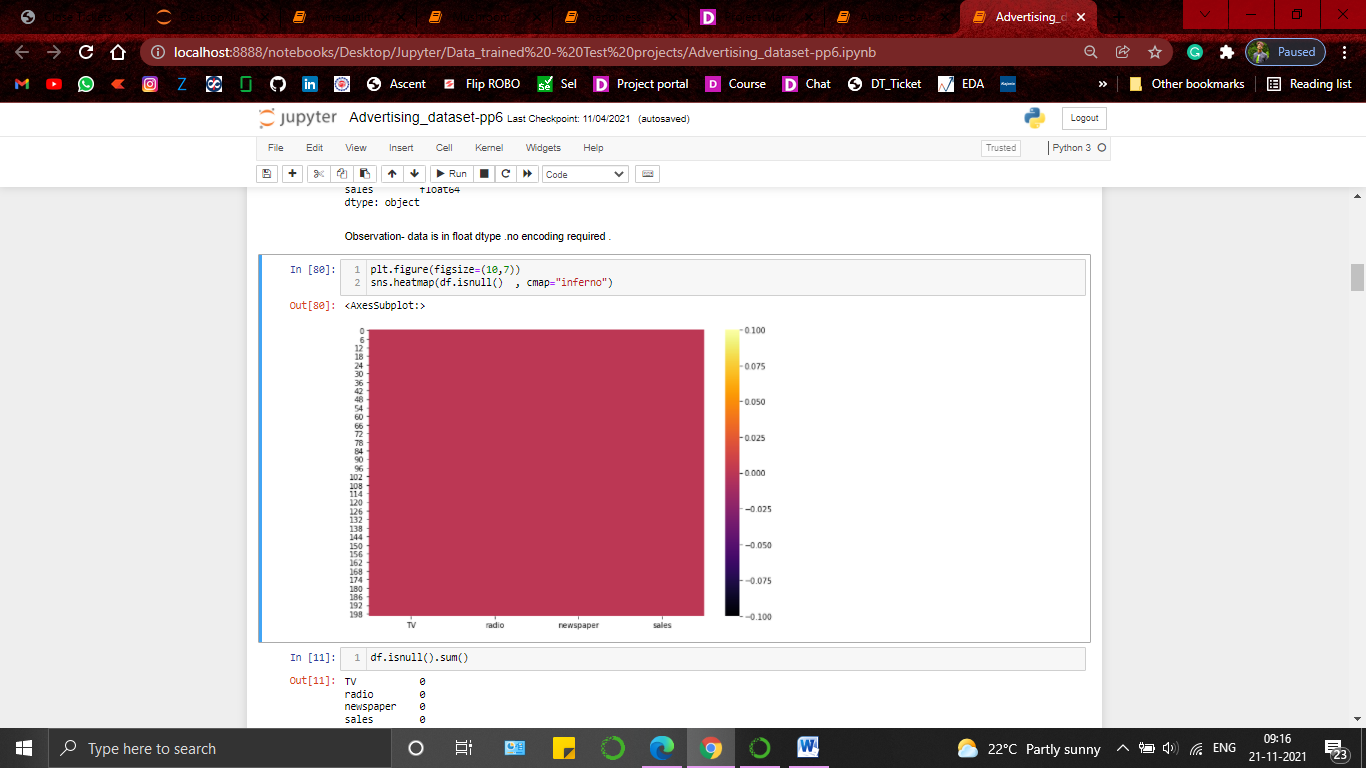
We have a data set containing sales generation data of a company . the distributions is among 3 different platforms . Thhose are tv , radio , newspaper . 

We have 200 rows and 4 columns which includes 200 instances of log entry and all 3 mediums with the sales generation data .

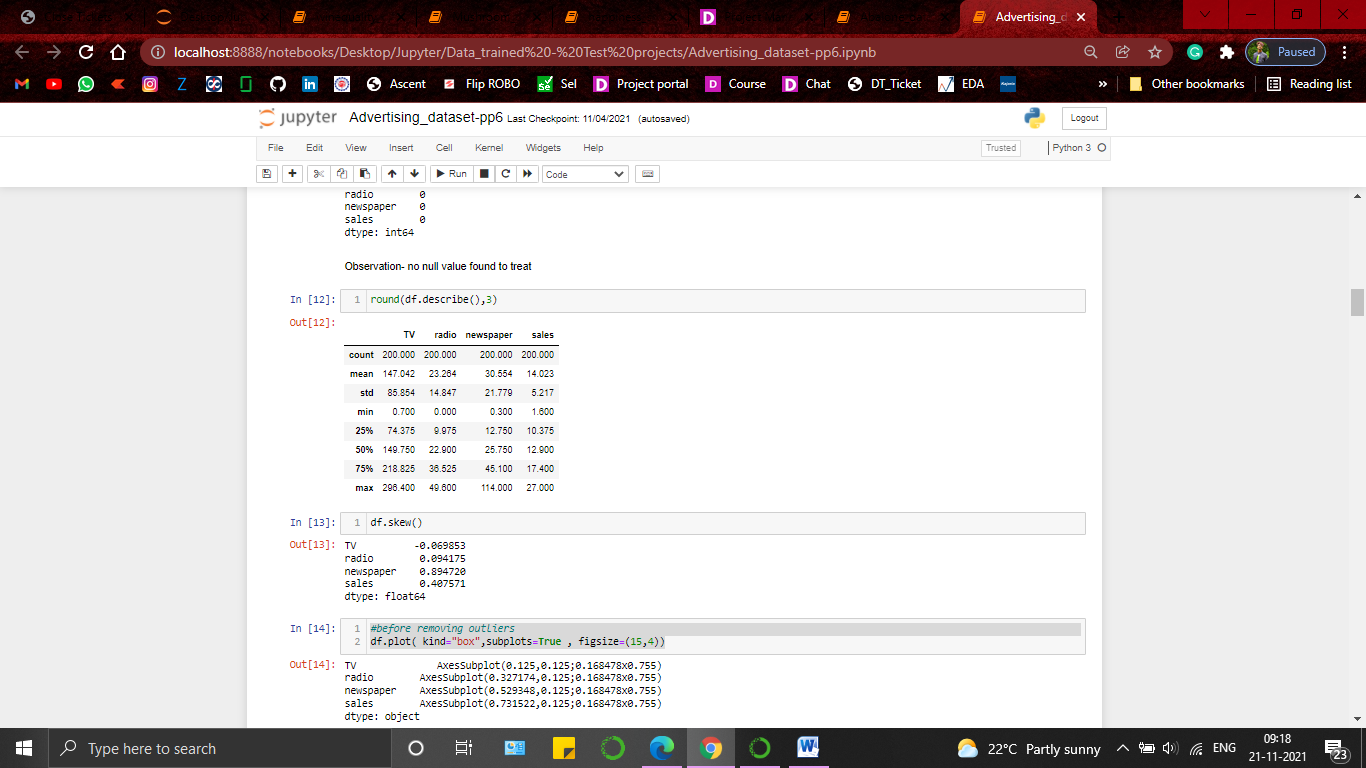


1. EDA and preprocessing .

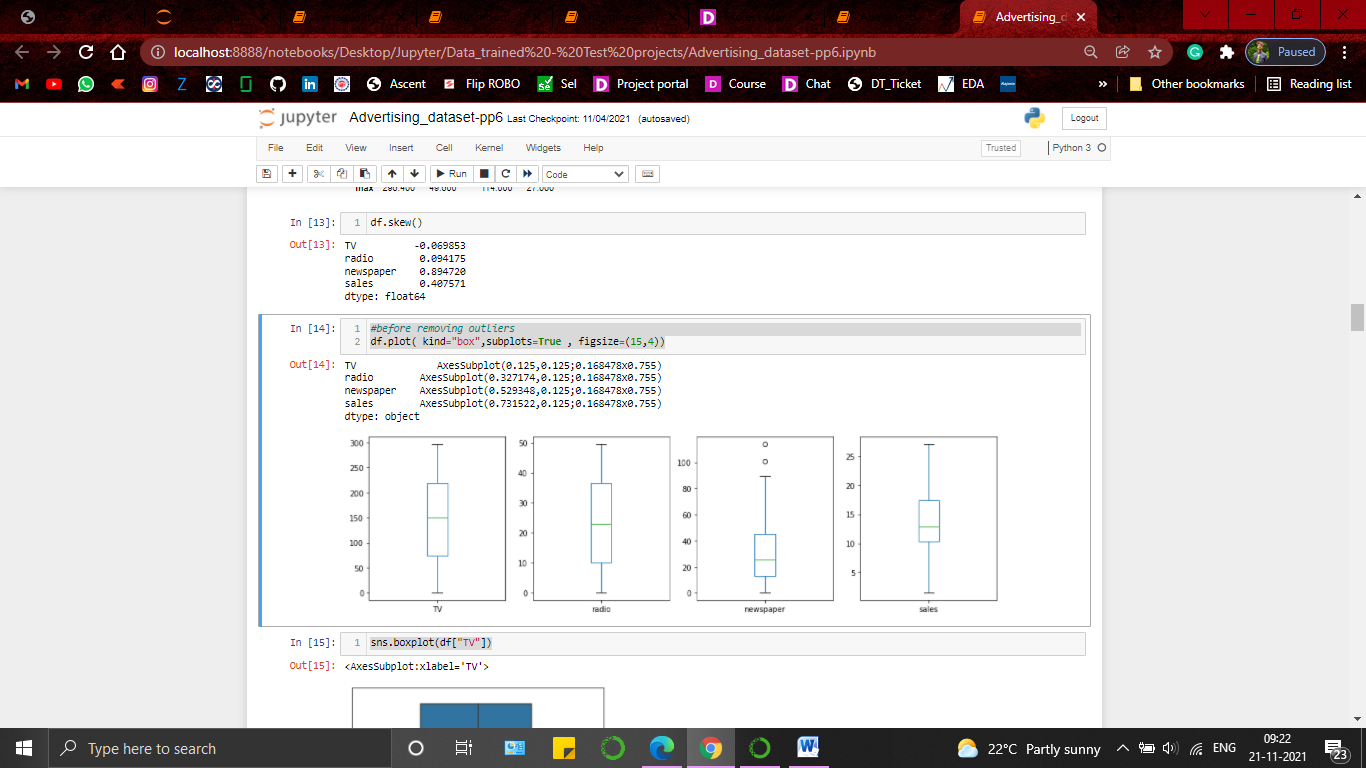
Checked for null values using boxplot , found no null value in the dataset .



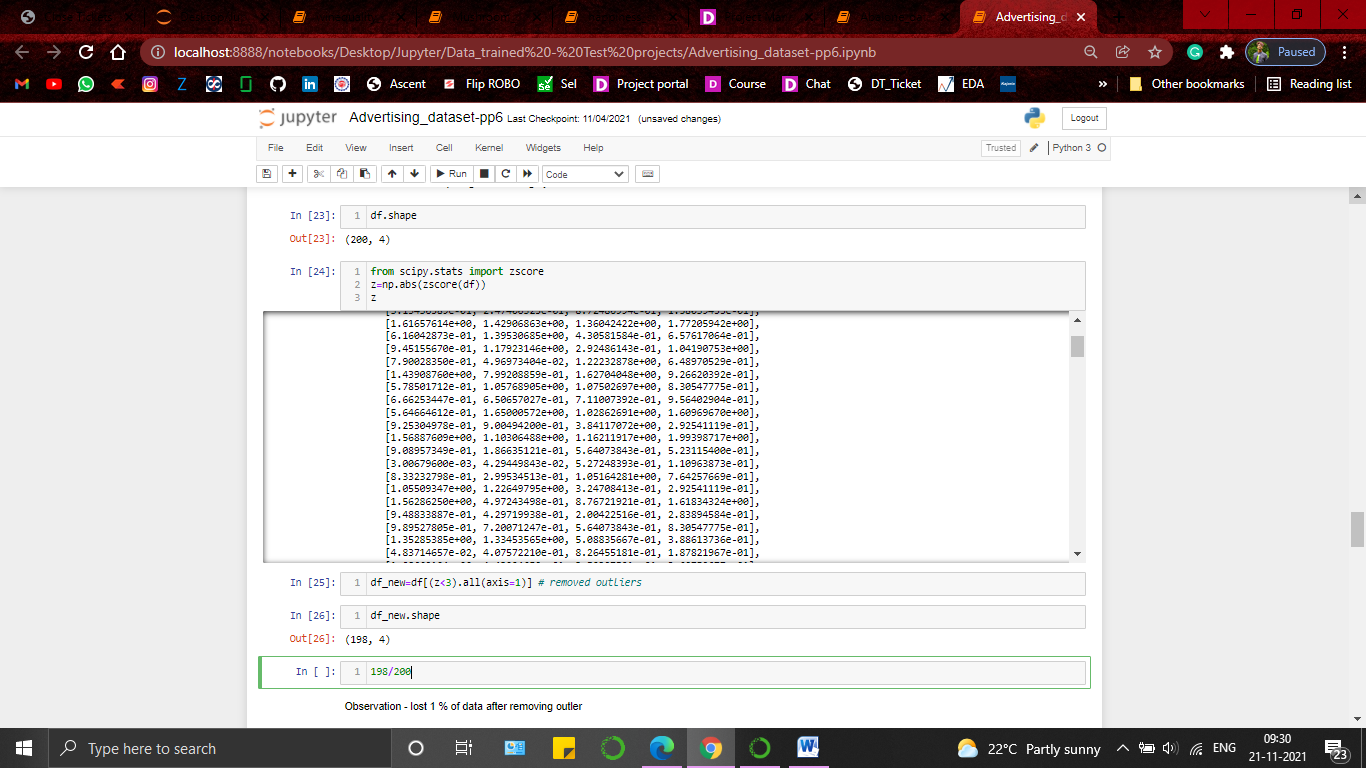
Checked the stats of the dataset . obseved that mean of newpaper and its max value has a huge difference That indicates presence of outliers in the dataset



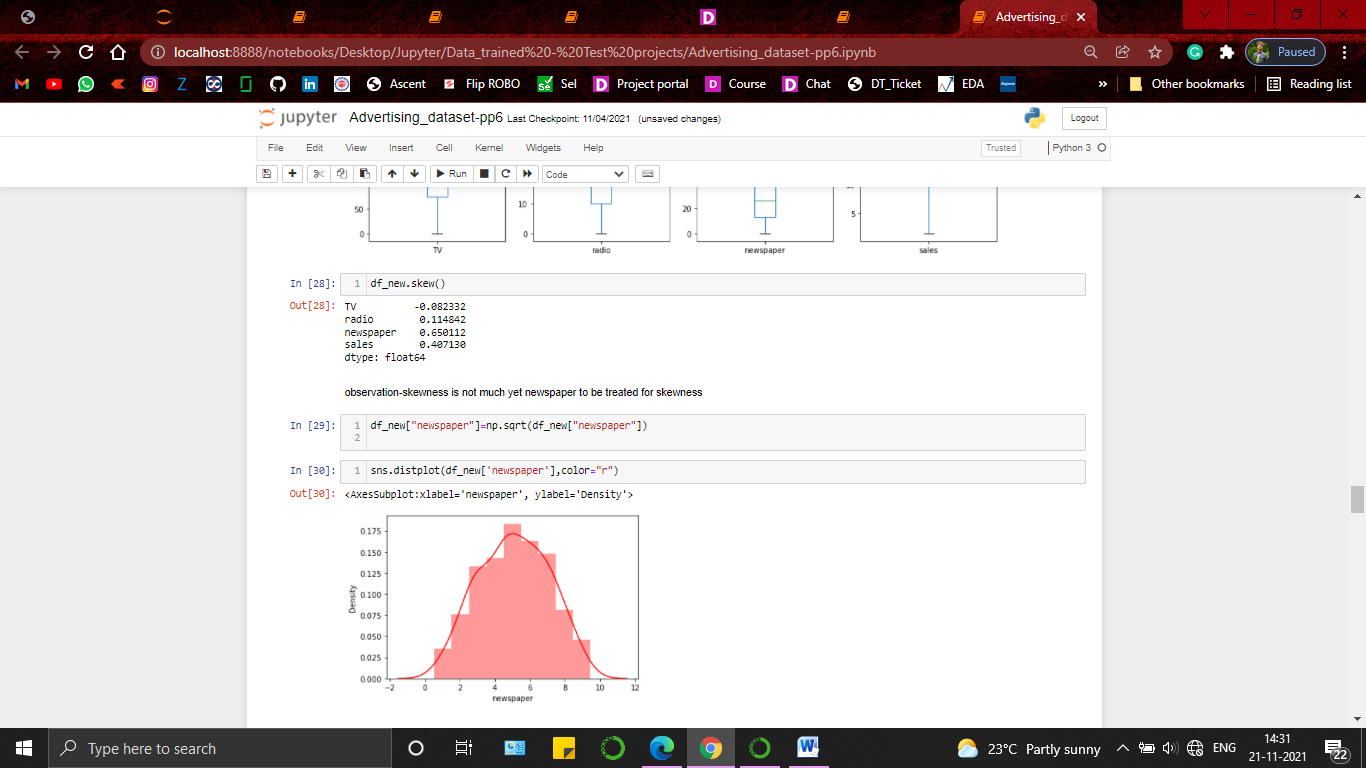
So checked outlier presence in the dataset using boxplot and Found out as assumed , newspaper data have presence of some outliers in the dataset .



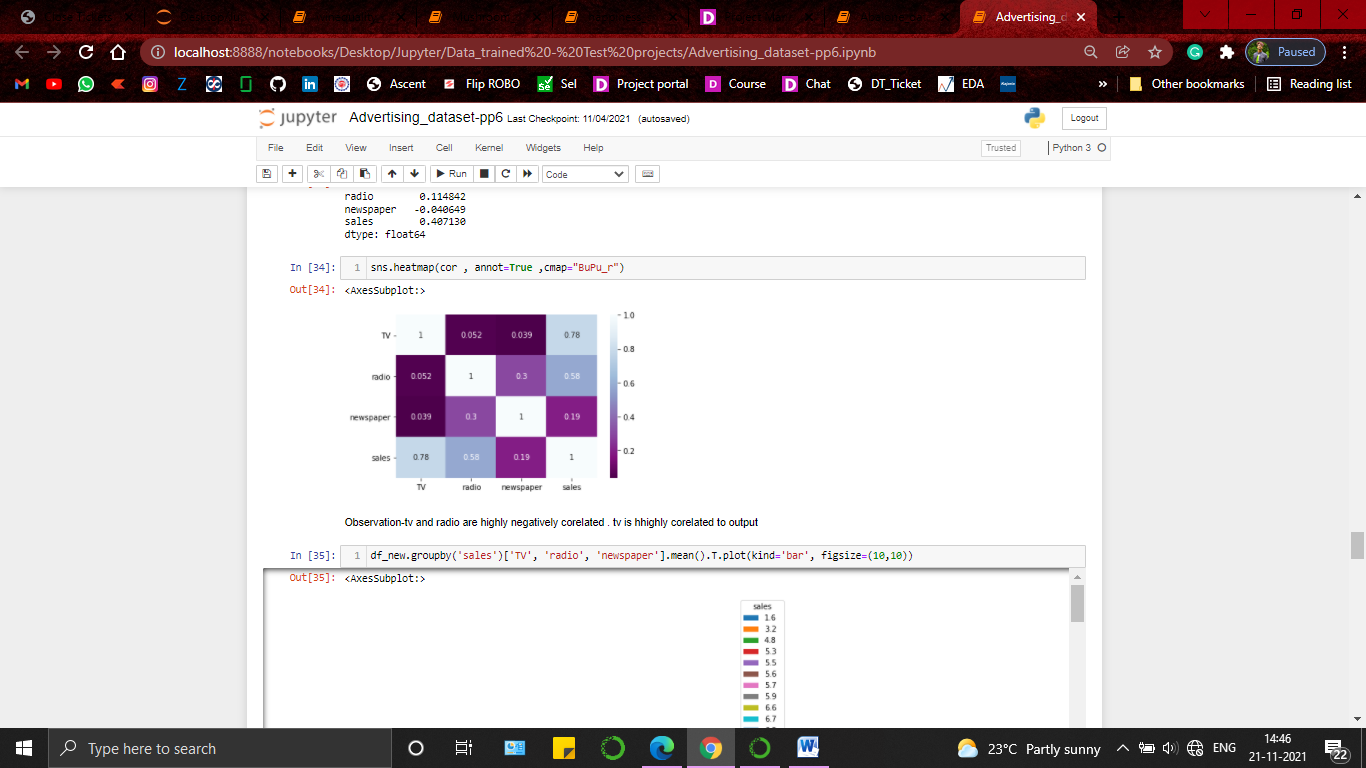
Removed outliers using Zscore method whih resulted as loss of 1% data from the dataset .



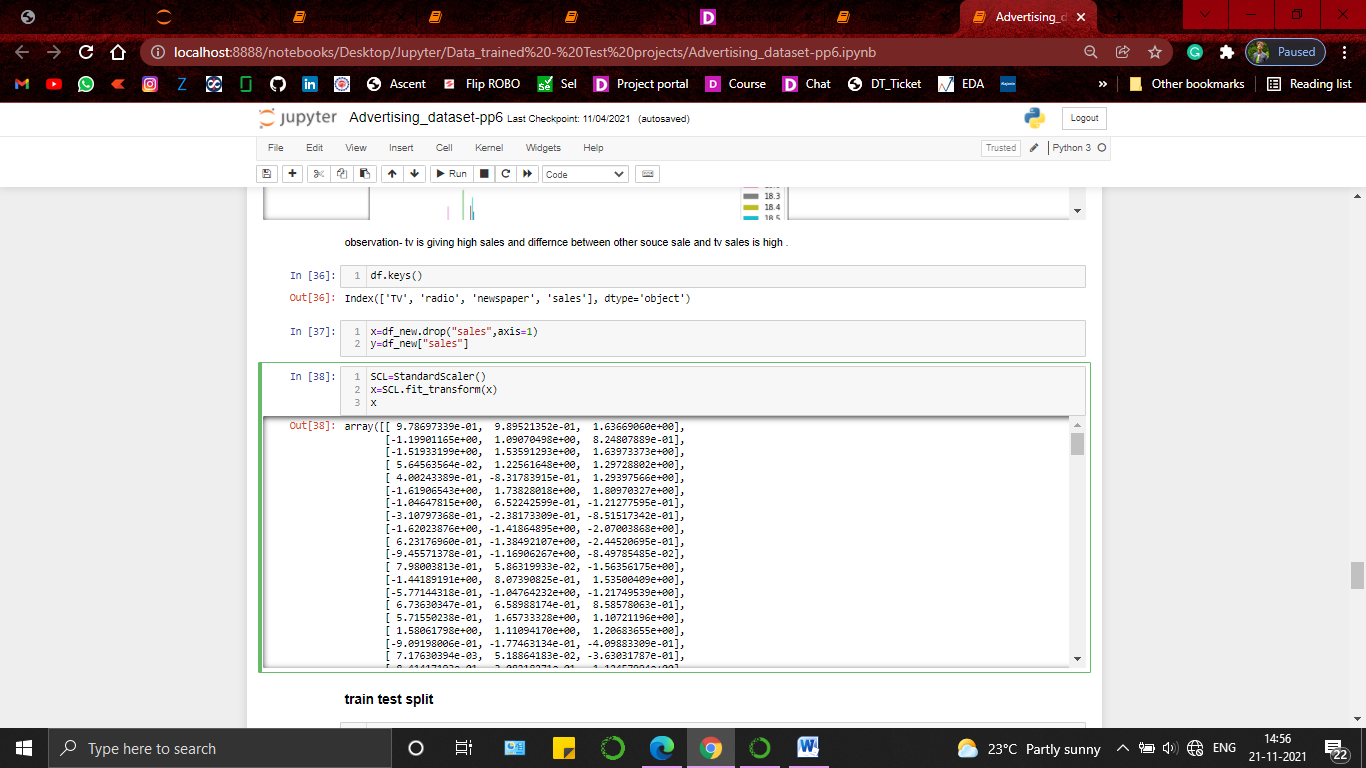
Checked for the skewness present in the dataset . found only news paper variable is having skewness more tha 0.5 so treated only newspaper variable by using Square root function from numpy .



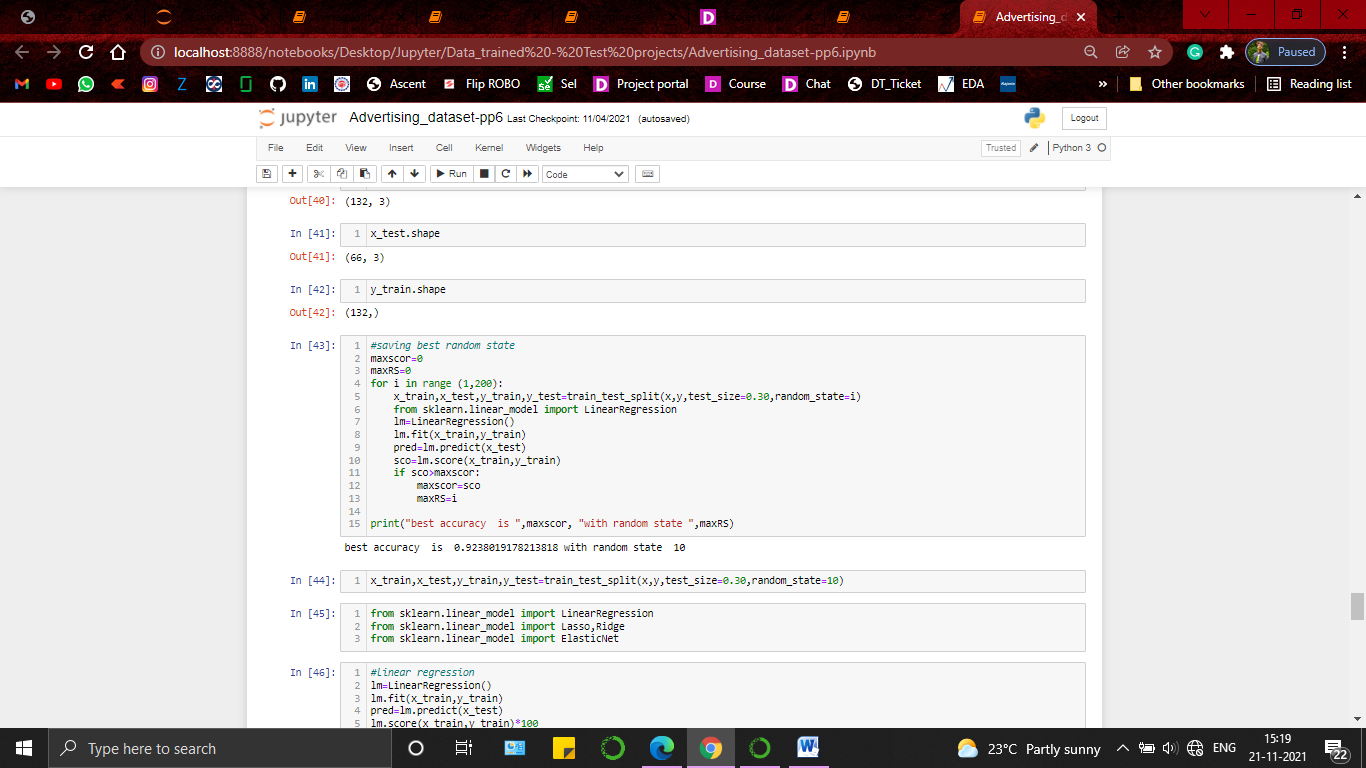
Checked for the corelation between variables and predicting variable to target variable . found no variable is having a quite negative corelation (more than -0.5) towards sales value . so moving forward with this much processed data for the model preparstion .



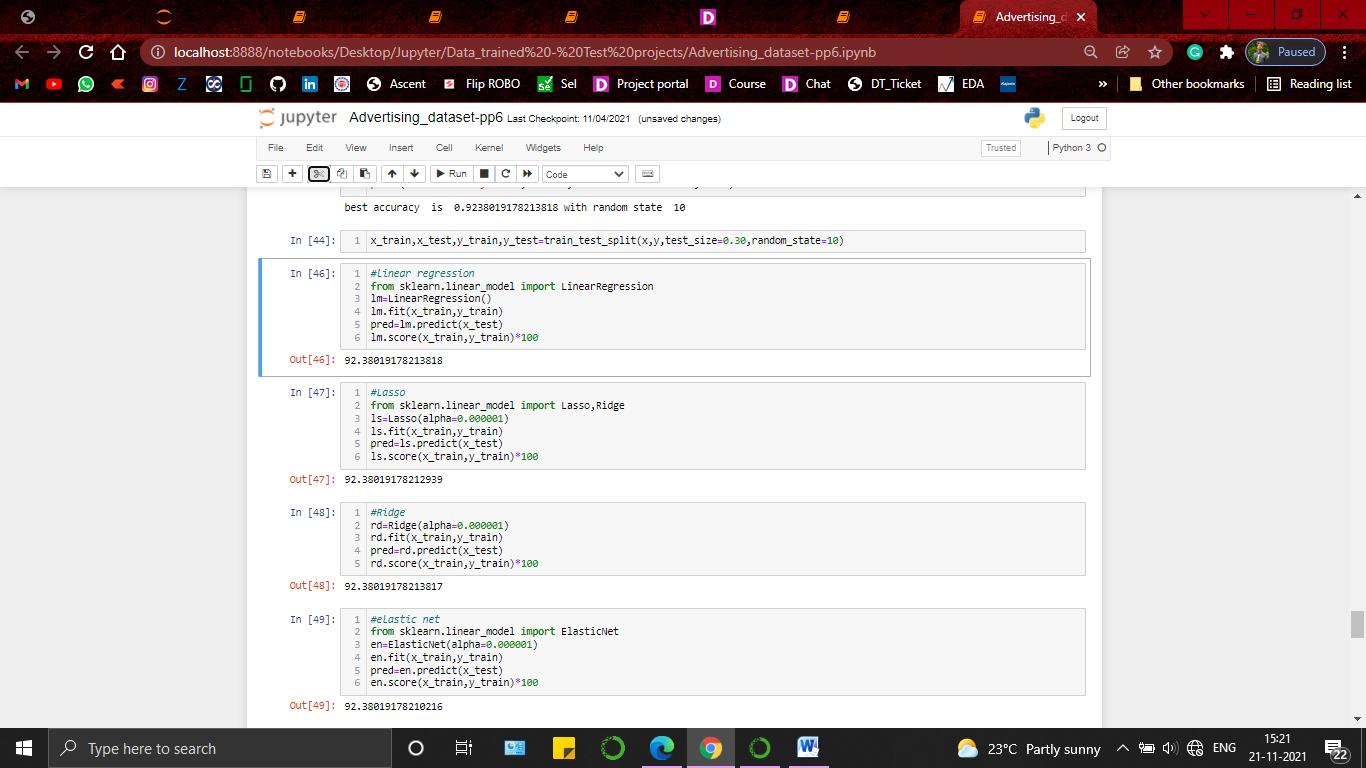
4.Building Machine Learning Models.

Splitted the variables and target value into x and y for training . used standard scaler on the predicting variables as all the values were in different ranges 

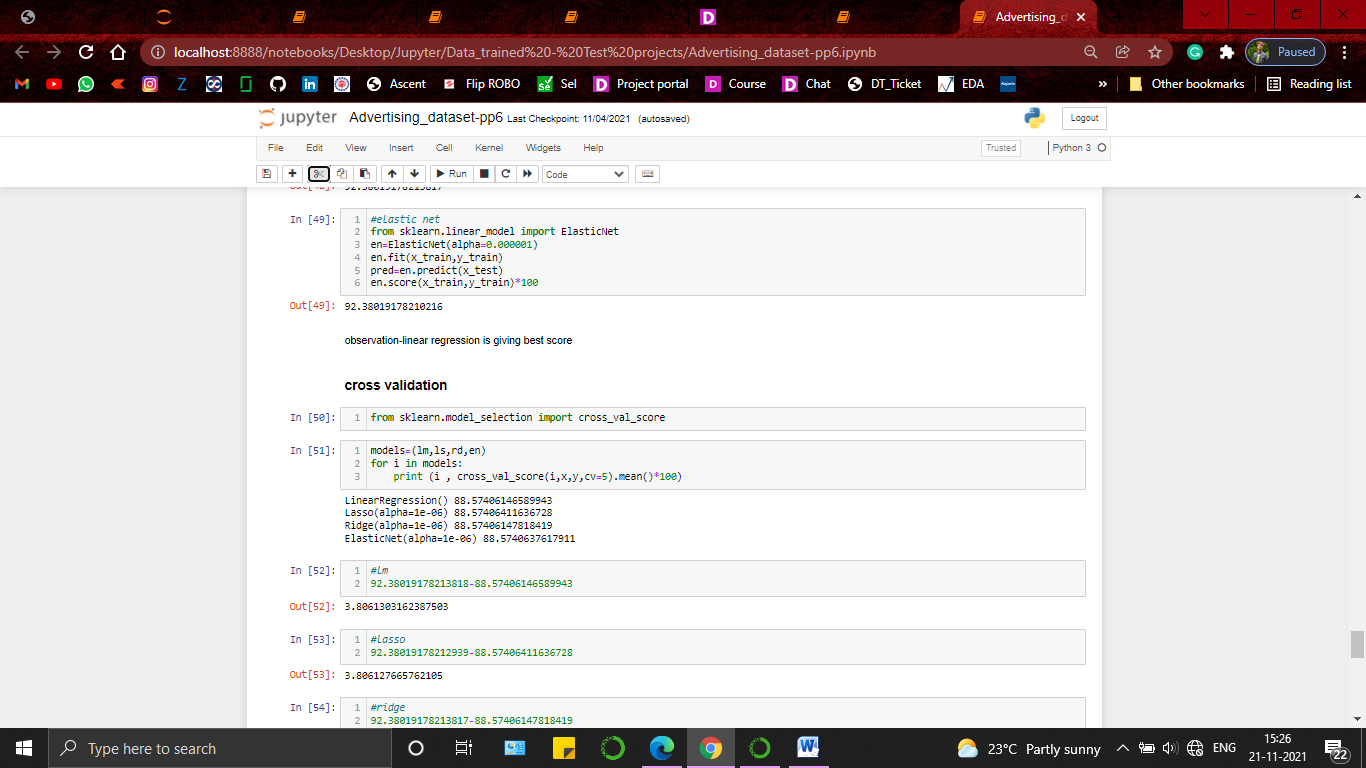
Checked for the best random state which would helpme to provide the right kind of data to make the module learn best . Best random state happens to be 10 so splitted x and y with random state 10 for 70:30 ration of split .



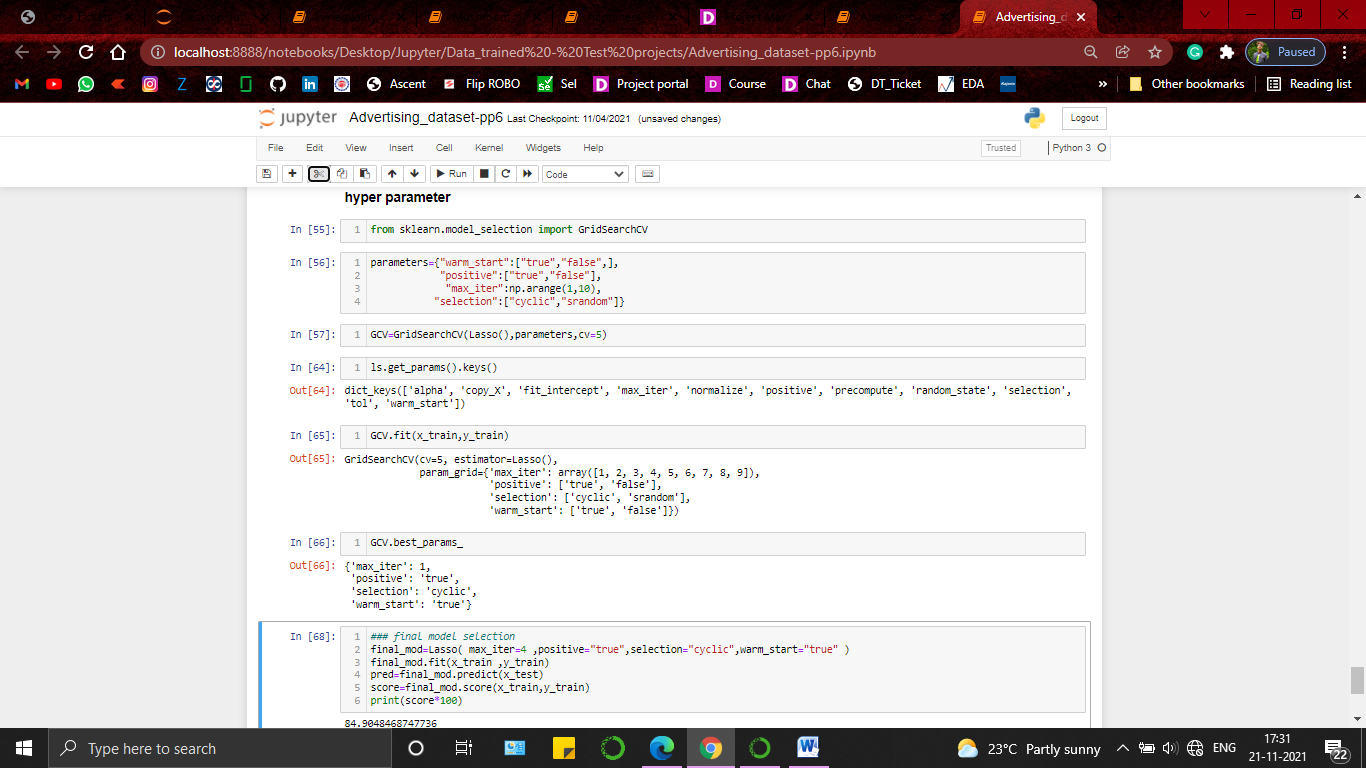
And trained LinearRegression , Lasso , Ridge and ElasticNet to comape the scores of accuracy .



Used cross validation score and comapred the gap between cross validation score and and the accuracy score to find out the model learning best from the given data . Happens to be Elastic net is the model with best score and minimum difference . so chose LinearRegression as the best model .



Prepared the final model by using Lasso and parameters from the grid search cv for a better accuracy .



5. conclusion.

The model is al setup to predict the sales generation based on the asset or effort the company puts into different platform . Model accuracy is 84.90 % which means its not 100% correct yet it will give the prediction with an 84% of accuracy which will help the investors or company capital investment team to take a decision on which platform to focus to increase the sales generation .

Prepared by-Amit patel .