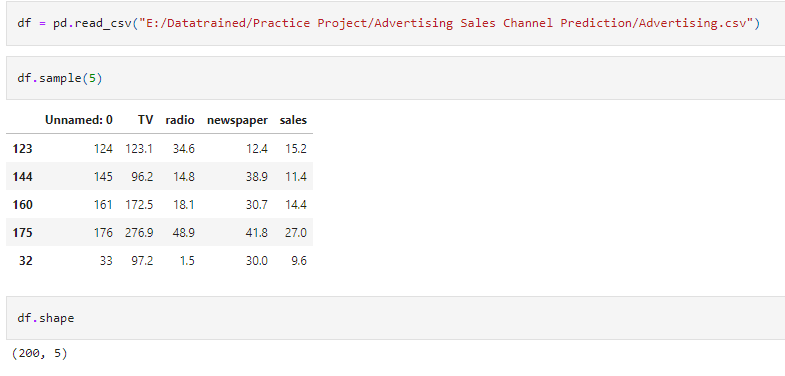
**Advertisement Dataset Analysis**

**Problem Definition:**

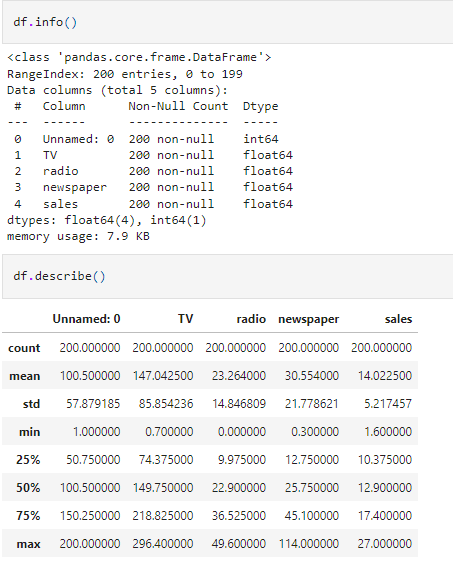
Every channel has a product or service to deliver and to do so they must acquire a good market to generate and leverage good sales. to Reach the customer company uses different platforms to advertise their product for a better reach but a company should have a better understanding about which platform 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.

**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 .

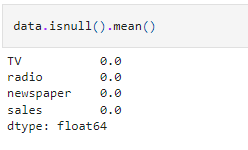


We have 200 rows and 5 columns which includes 200 instances of log entry and all 3 mediums with the sales generation data where one column has been removed due to unwanted variable. So in original we have 200 rows and 4 columns to work on.

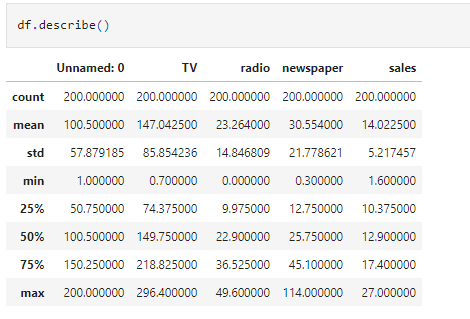


**EDA and Pre-processing:**

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

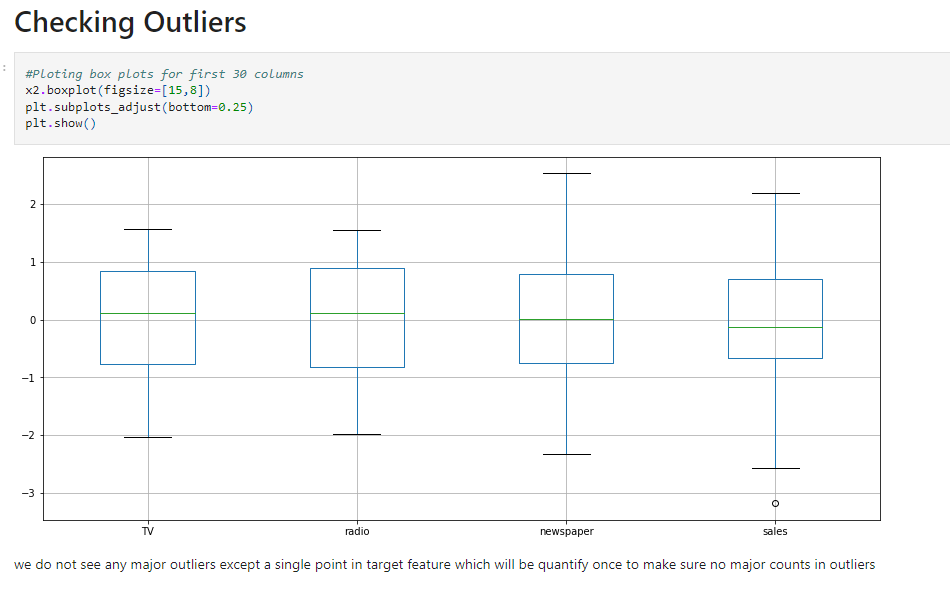


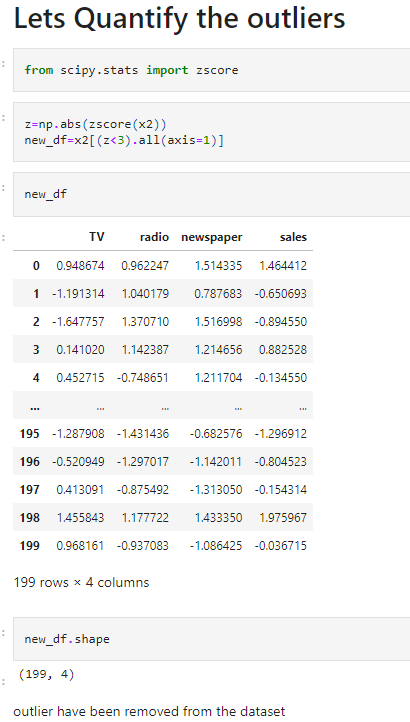
Checked the stats of the dataset . obseved that mean of newpaper and its max value has a huge difference That indicates presence of outliers in the dataset



**Outliers Check :**

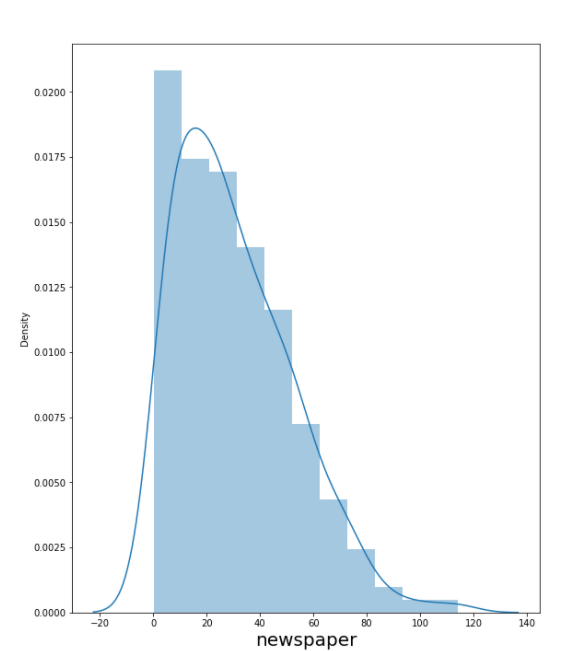
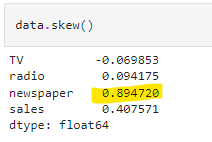
So checked outlier presence in the dataset using boxplot and found we do not see any major outliers except a single point in target feature which will be quantify once to make sure no major counts in outliers. But still I have handled those outliers using ZScore method below

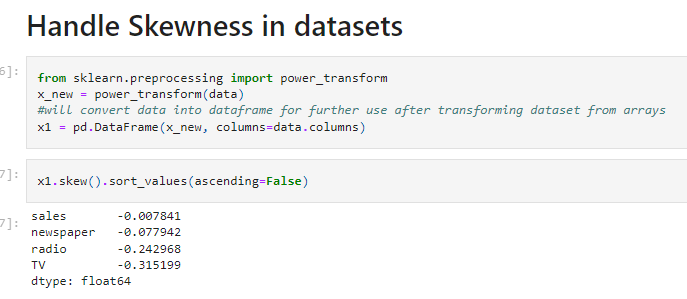




**Skewness Check :**

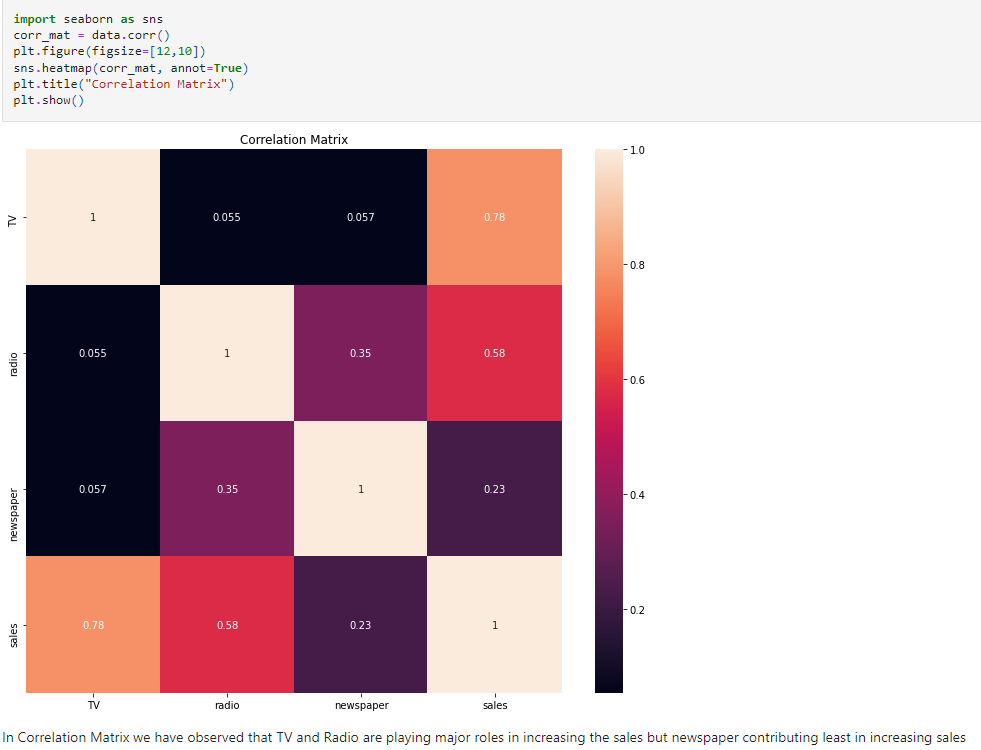
Checked for the skewness present in the dataset. found only news paper variable is having skewness more tha 0.5 so treated only newspaper variable by using Power Transform method to handle skewness.



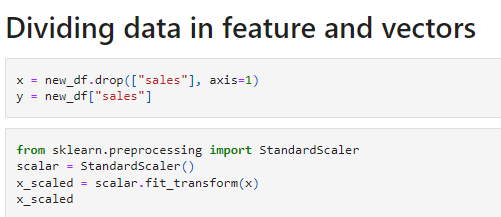


**Correlation Check :**

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 .

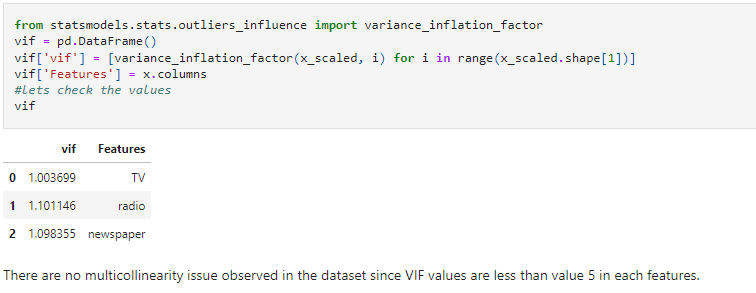


**Performed Scaling:**

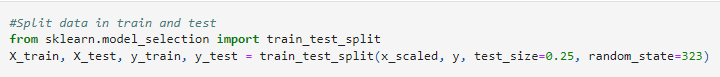
****

**Checking Multicollinearity:**

There is no multicollinearity issue observed in the dataset since VIF values are less than value 5 in each feature.

****

**Split Dataset for training and testing the model:**

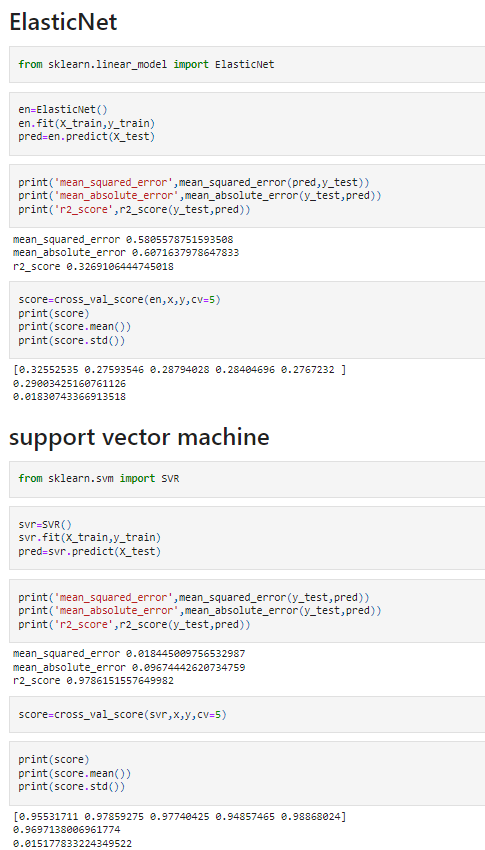
****

**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









Out of all the algorithms we have found Random Forest Regressor performing better

**Hyperparameter Tuning:**

Used Hyperparameter Tuning to find out the best Parameter combination to use those values in our best model.

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



**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 97 % which means its not 100% correct yet it will give the prediction with an 97% 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 Singh**