**FINAL PROJECT**

**FUTURE SALES PREDICTION USING DIFFERENT REGRESSION MODELS USING MACHINE LEARNING**

By

**Group:**

Ridham Jain (700742608)

Sai Kiran Kalaganti (700744260)

Sai Kiran Reddy Kotha (700746206)

Shruthika Veeravarapu (700737920)



* Initially we have imported all the libraries.
* Then we have filtered all the warning messages that were coming.
* Then we have uploaded the file and then read it.
* Printed the summary of data. (First 5 rows)
* Data Cleaning

\*Checking for missing values.

\*Drop the rows with missing values.

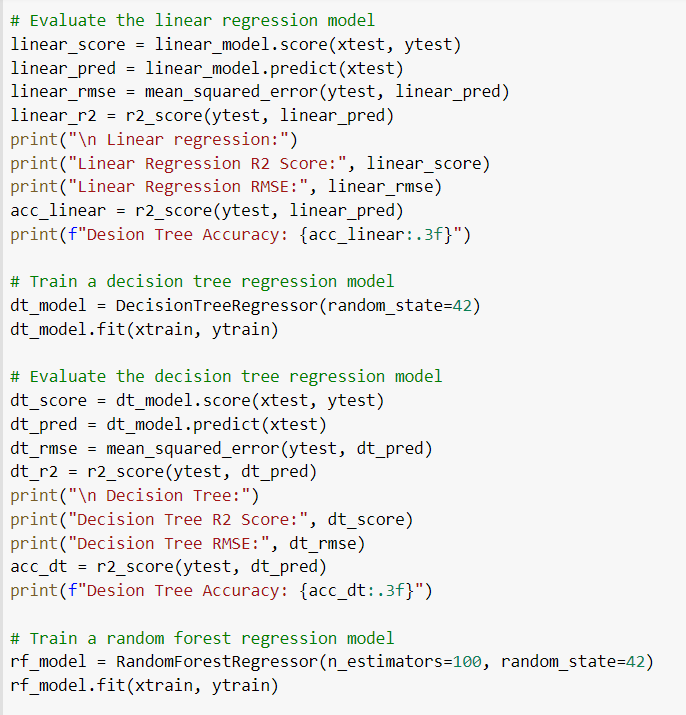
\*Drop the rows with duplicate values.



\*Removed the outliers.

\*Saved clean data to a new file.

* Then we have generated the scatterplots of various modes of advertisement v/s sales.
* Then we have found the correlation coefficient.
* Split the data into training and testing subsets.
* Train the model for linear regression



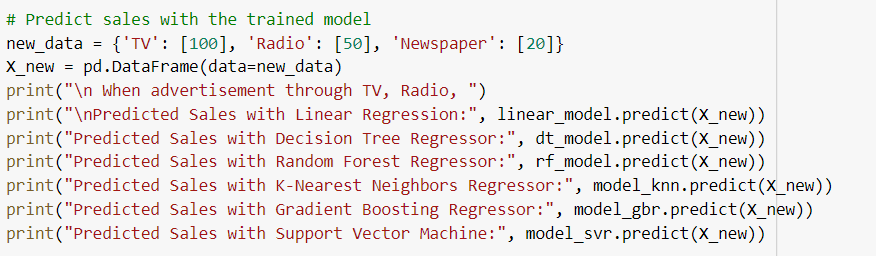
* Evaluated the linear regression model.
* Found the accuracy of linear regression model.
* Train and evaluate the decision tree regression model.
* Found the accuracy of decision tree.
* Train and evaluate the random forest regression model.
* Found the accuracy of random forest.



* Train and evaluated the support vector regression model.
* Found the accuracy of support vector regression model.
* Train and evaluate the K-nearest regression model.
* Found the accuracy of K-nearest regression model.

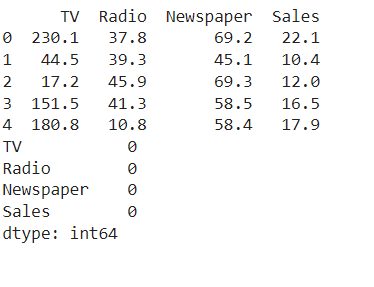


* Train and evaluated the gradient boosting regression model.
* Found the accuracy of gradient boosting regression model.
* Plotted the bar graph to compare the accuracies of all the regression models.

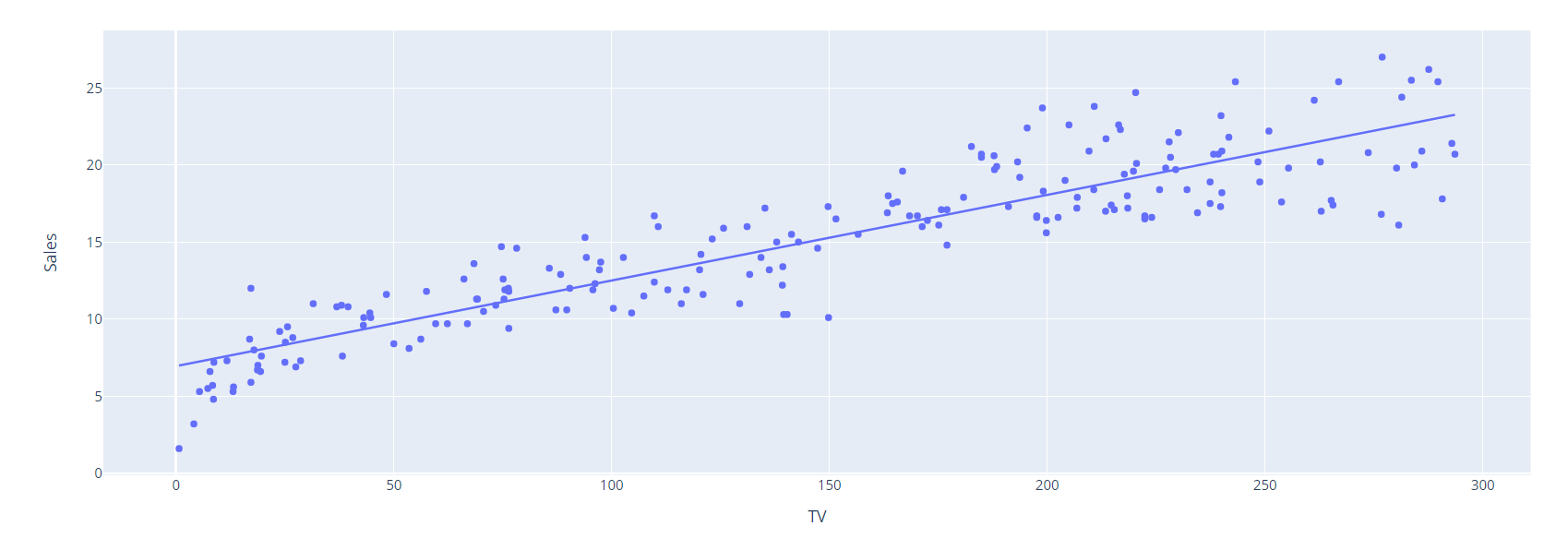


* Predicted the sales using all regression models at advertisement through TV-100, Radio-50 and Newspaper-20.

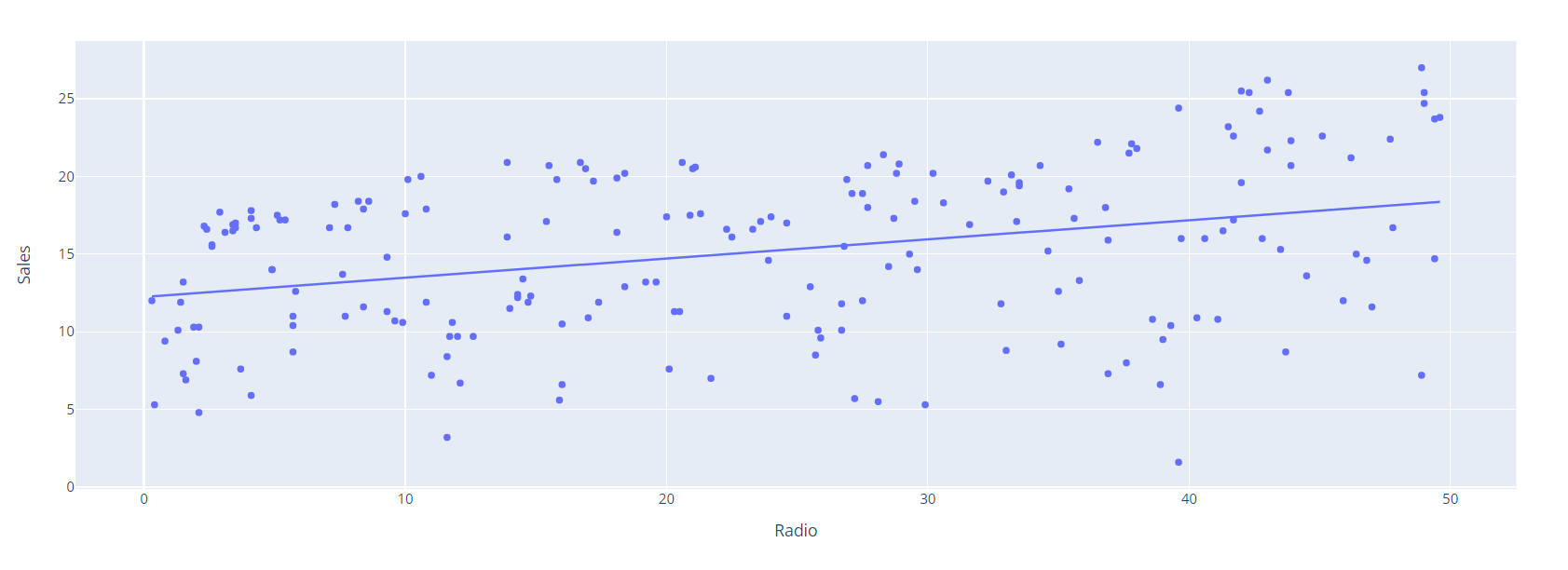
Answer



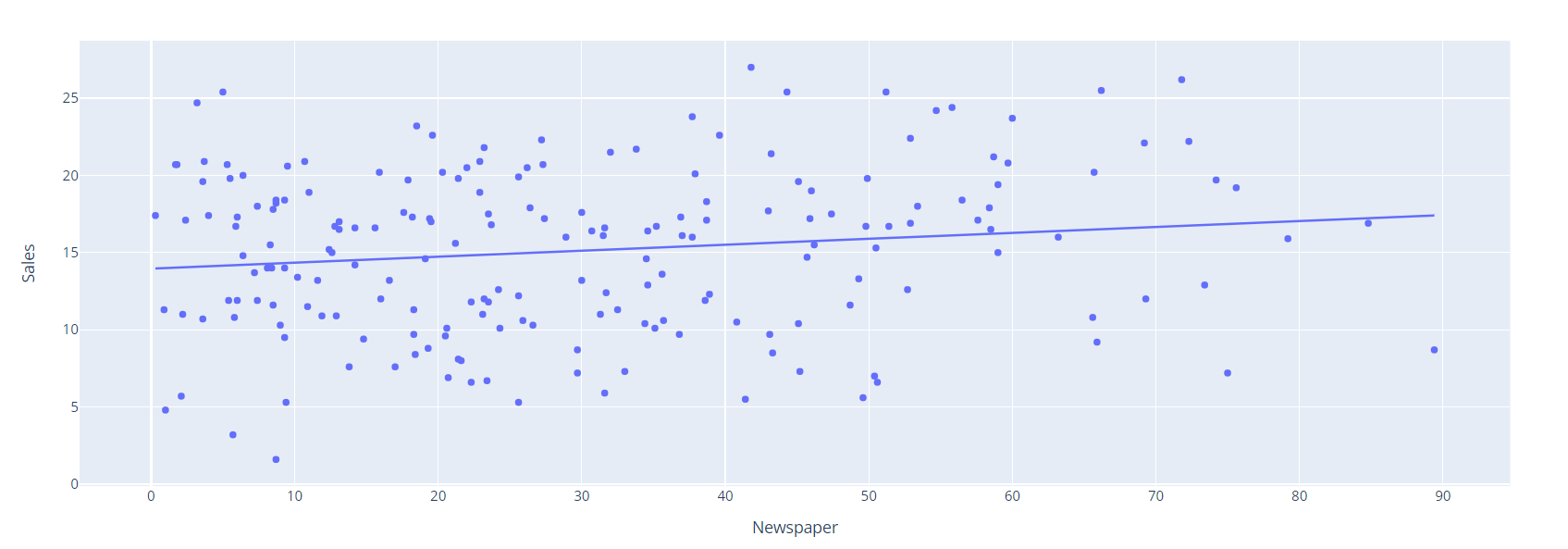
* This is data summary showing first 5 rows.
* Then it shows the number of null values in each column.
* We can see that there is no null value.



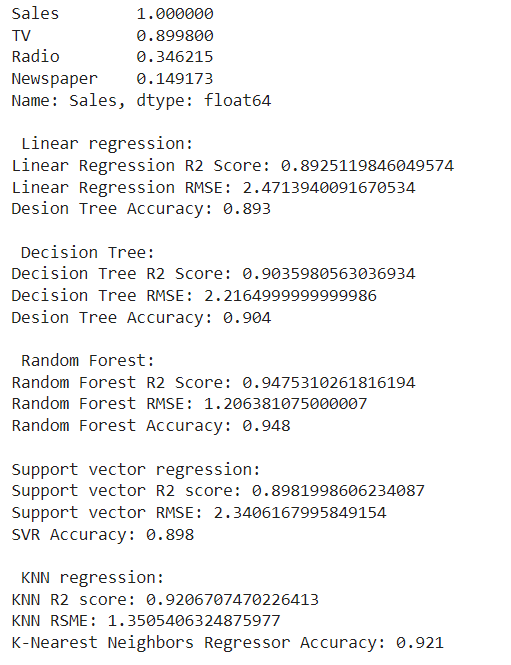
* Scatterplot showing TV advertisement v/s sales.
* We can see that as the advertisement through TV increases the sales also increases. There is positive relation between both.



* Scatterplot showing Radio advertisement v/s sales.
* We can see that as the advertisement through Radio increases the sales also increases but the slope is quite low as compared to the TV advertisement. Still there is positive relation between both.

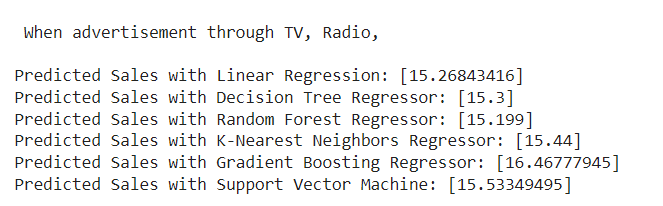


* Scatterplot showing Newspaper advertisement v/s sales.
* We can see that as the advertisement through Newspaper increases the sales also increases but the slope is quite flat as compared to other two. Still there is positive relation between both.



* Here, we can see the results of the correlation coefficient. Here the results are calculated keeping the sales as 1.
* Correlation coefficient of TV-0.8998
* Correlation coefficient of Radio-0.3462
* Correlation coefficient of Radio-0.1492
* Linear regression accuracy:89.3%
* Decision Tree regression accuracy:90.4%
* Random Forest regression accuracy:94.8%
* Support Vector regression accuracy:89.8%
* K-NN regression accuracy:92.1%





* Gradient boosting regression accuracy:93%
* Plotted bar graphs of all regression models and accuracies.
* Predicted the sales using different regression models at same value of advertisement through TV, Radio, and Newspaper.