ASSIGNMENT 2 ASTROS REPORT

1. Dataset:-

The dataset contains advertising expenditures of the company on various modes of marketing methods. It also contains the total sales too. The independent variables are the cost on TV, radio and newspaper modes. The dependent variable is the total sales.

1. Data profiling:-

Using ydata\_profiling library to find correlation between features and target variable. It summarises the data in a clean manner using graphs and plots. Using the data profile it seems only TV has some sort of correlation with sales.

1. Feature Extraction and Splitting:-

Extracting values for features.

sel1 = sel.sort\_values(by=['TV','Sales'])

x = sel1[['TV']]

y = sel1['Sales']

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.20,random\_state=255)

1. Feature Engineering:-

Changing the values of x to make the graph linear and to find the better R2 value.

x=x\*\*(1/1.5)

1. Linear Regression:-

Training a linear regression model on the values of feature and target.

lm = LinearRegression()

lm.fit(x\_train,y\_train)

1. Predictions:-

Predicting sales using the model build.

predic = lm.predict(x\_test)

plt.scatter(predic,y\_test)

plt.show()

1. Results:-

r2=r2\_score(y\_test,predic)

Intercept:- 4.578521143597522

slope:- [0.39918043]

R2:- 0.8699300855680505

1. Conclusion:-

The linear regression model shows good results on the test dataset. It has a R2 score of 0.8699. Overall this shows that the model trained is good for predicting total sales with sufficient data.