Exp no: 3 Sales prediction and customer segmentation using linear regression and k-means clustering

Aim: To predict sales based on advertising based on advertising budget (TV) radio and newspaper using linear regression additionally to identify distinct patterns in advertising and sales through k-means clustering

Algorithm:

Step 1: Load the advertising .csv containing TV, radio, news paper, budget and sales.

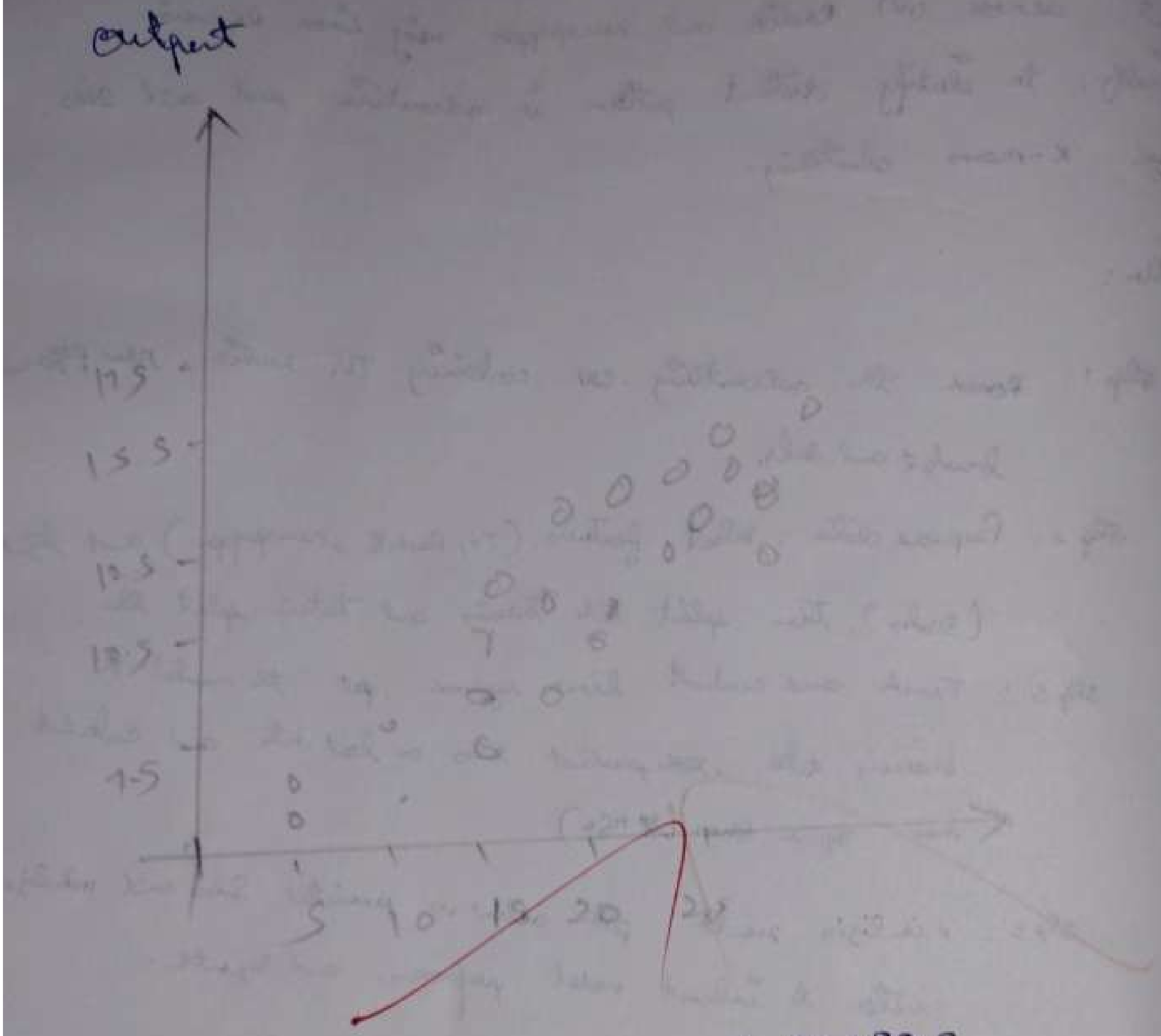
Step 2: prepare data: select features (TV, radio, newspaper) and target (sales), then split into train, splitfit ~~the model~~.

Step 3: Train and evaluate linear regression, fit the model on training data, predict sales on test data and calculate mean squared error

Step 4: interpret result. plot actual vs predicted sales and analyze clusters to interpret model regression and insights

Program

```
import pandas as pd
import matplotlib as plt
import seaborn as sns.
from sklearn.model_selection import train_test_
from sklearn.metrics import split mean_squared_error_split
from sklearn.cluster import k-means.
from sklearn.preprocessing import StandardScaler
```



Liren Regessio MSE : 4.52 25525

```
df = pd-read-csv (advertising_csv)
Print (df. head ()) print(df. info (s))
x = df [['TV', 'radio', 'newspaper']]
y = df ['sales']
x-train , x-test, y-train , y-test = train_test = split (x, y,
                                        test , size = 0.2,
                                        random-state = )

model = linear regression ()
model- fit (x-train , y-train )
y-pred = model - predict (x- test )
mse = mean. predict (e- test )
print ("linear regression MSE:", mse)
plt- figure (figsize =(8-5) ,
sns - scatterplot (x=y , te, y= y pred )
plt- ylabel ("Actual sales",)
plt- title ("linear regression Actual vs predicted sales}
plt- show()
scaler = standard scale ()
plt figure ( fig size =(8 , 16 ) ]
sns. scatter .plot [data= df, x= 'TV', y=' sales', hue='clust',
            pd Cul = ("set" )
plt= show()
```

result-:

The linear regression model with a mean squared ms le) of Aup- show its a good fit between actual and predicted value in mean, clustering grouping market core visible (clust along the relation 3 digits & sales shows patterns , advertising budgets and sales.