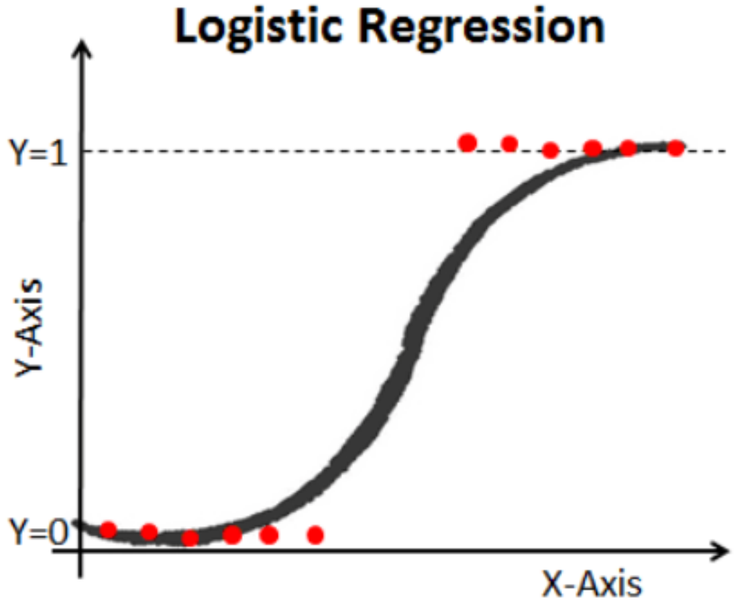
LOGISTIC REGRESSION

PRATHAPANI SATWIKA
20BCD7160

What is logistic regression?

A statistical analysis method called logistic regression uses previous observations from a data set to predict a binary outcome, such as yes or no. By examining the correlation between one or more already present independent variables, a logistic regression model forecasts a dependent data variable.



TYPES OF LOGISTIC REGRESSION :

Three types of logistic regression

Binary logistic regression

Multinomial logistic regression

Ordinal logistic regression

Advantages :

Because it is simple and effective, data analysts and scientists frequently utilise it because it doesn't take a lot of computing power, is simple to apply, and is simple to comprehend. Additionally, scaling of features is not necessary. A probability score for observations is provided by logistic regression.

Disadvantages :

Large numbers of categorical features or variables cannot be handled using logistic regression. It is susceptible to being overfit. Additionally, logistic regression cannot handle the non-linear problem; for this reason, non-linear features must be transformed. When independent variables are substantially similar to one another or are associated to one another but are not correlated to the target variable, logistic regression will not work well.

DATA SET

	A	B	C	D	E
1	company	budget(in K)	sales(in K)	Revenue growth	
2	1	50	90	1	
3	2	75	80	1	
4	3	25	28	1	
5	4	35	45	1	
6	5	100	95	0	
7	6	90	100	1	
8	7	45	46	1	
9	8	12	10	0	
10	9	80	86	1	
11	10	15	10	0	
12	11	65	72	1	
13	12	96	98	1	
14	13	36	32	0	
15	14	87	82	0	
16	15	34	35	1	
17	16	54	67	1	
18	17	34	56	1	
19	18	98	102	1	
20	19	54	50	0	
21	20	37	43	1	
22	21	23	26	1	
23	22	90	92	1	
24	23	66	68	1	
25	24	18	16	0	
26	25	99	105	1	

CODE

```
[108] import numpy as np
      import matplotlib.pyplot as plt
      import pandas as pd
```

```
      dataset=pd.read_csv('sales.csv')
      dataset.head(5)
```

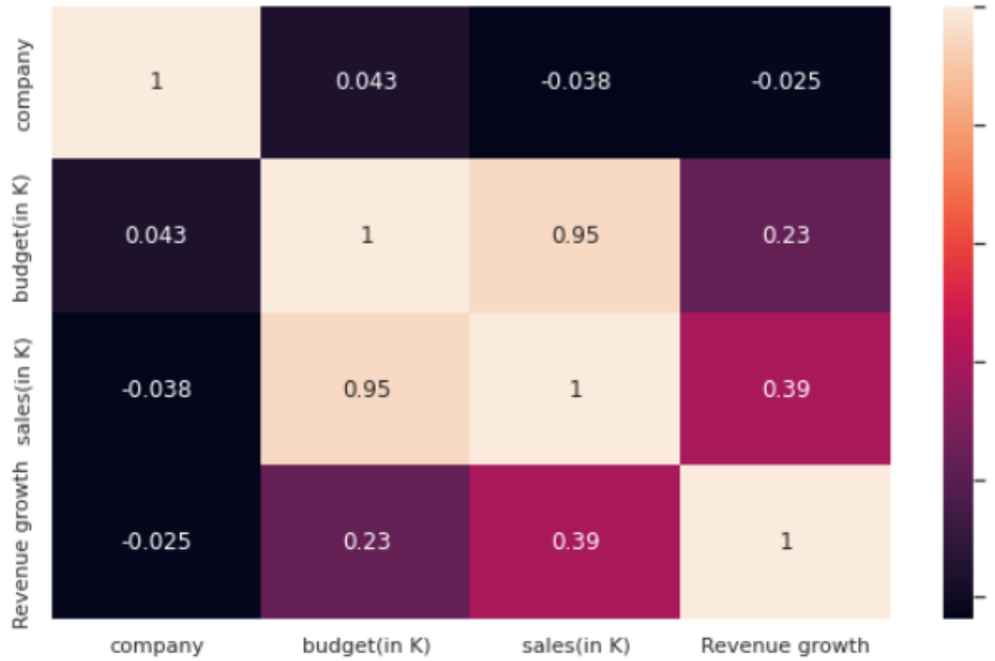
	company	budget(in K)	sales(in K)	Revenue growth
0	1	50	90	1
1	2	75	80	1
2	3	25	28	1
3	4	35	45	1
4	5	100	95	0

```
[136] X = dataset.iloc[:, 0:3].values
      x = dataset.iloc[:, 0:3]
      y = dataset.iloc[:, 3:4].values
      y_name = dataset.iloc[:, 3:4]
      print("X shape:",X.shape)
      print("y shape:",y_name.shape)
```

```
X shape: (25, 3)
y shape: (25, 1)
```

CODE

```
import seaborn as sn
import matplotlib.pyplot as plt
sn.set(font_scale=1)
sn.set_style("darkgrid")
fig_dims = (10, 6)
fig, ax = plt.subplots(figsize=fig_dims)
sn.heatmap(dataset.corr(),annot=True, ax=ax)
plt.show()
```



CODE

```
[144] from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
      ac  = accuracy_score(y_test, y_pred_lr)
      pre = precision_score(y_test, y_pred_lr)
      re  = recall_score(y_test, y_pred_lr)
      f1  = f1_score(y_test, y_pred_lr)
```

```
[145] print("Summary of Logistic regression")
      print("Accuracy    =", ac)
      print("Precison    =", pre)
      print("Recall      =", re)
      print("f1 score    =", f1)
```

```
Summary of Logistic regression
Accuracy    = 0.8
Precison    = 1.0
Recall      = 0.8
f1 score    = 0.888888888888889
```

Thank You