In [1]:

```
import pandas as pd
import numpy as np
```

In [2]:

```
df=pd.read_csv('forestfires.csv')
```

In [3]:

```
df
```

Out[3]:

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
0	7	5	mar	fri	86.2	26.2	94.3	5.1	8.2	51	6.7	0.0	0.00
1	7	4	oct	tue	90.6	35.4	669.1	6.7	18.0	33	0.9	0.0	0.00
2	7	4	oct	sat	90.6	43.7	686.9	6.7	14.6	33	1.3	0.0	0.00
3	8	6	mar	fri	91.7	33.3	77.5	9.0	8.3	97	4.0	0.2	0.00
4	8	6	mar	sun	89.3	51.3	102.2	9.6	11.4	99	1.8	0.0	0.00
...	...	...	...	...	...	...	...	...	...	...	...	...	...
512	4	3	aug	sun	81.6	56.7	665.6	1.9	27.8	32	2.7	0.0	6.44
513	2	4	aug	sun	81.6	56.7	665.6	1.9	21.9	71	5.8	0.0	54.29
514	7	4	aug	sun	81.6	56.7	665.6	1.9	21.2	70	6.7	0.0	11.16
515	1	4	aug	sat	94.4	146.0	614.7	11.3	25.6	42	4.0	0.0	0.00
516	6	3	nov	tue	79.5	3.0	106.7	1.1	11.8	31	4.5	0.0	0.00

517 rows × 13 columns

In [4]:

```
#handle missing value
df.dropna(inplace=True)#drop rows with missing value
```

In [5]:

```
df
```

Out[5]:

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
0	7	5	mar	fri	86.2	26.2	94.3	5.1	8.2	51	6.7	0.0	0.00
1	7	4	oct	tue	90.6	35.4	669.1	6.7	18.0	33	0.9	0.0	0.00
2	7	4	oct	sat	90.6	43.7	686.9	6.7	14.6	33	1.3	0.0	0.00
3	8	6	mar	fri	91.7	33.3	77.5	9.0	8.3	97	4.0	0.2	0.00
4	8	6	mar	sun	89.3	51.3	102.2	9.6	11.4	99	1.8	0.0	0.00
...	...	...	...	...	...	...	...	...	...	...	...	...	...
512	4	3	aug	sun	81.6	56.7	665.6	1.9	27.8	32	2.7	0.0	6.44
513	2	4	aug	sun	81.6	56.7	665.6	1.9	21.9	71	5.8	0.0	54.29
514	7	4	aug	sun	81.6	56.7	665.6	1.9	21.2	70	6.7	0.0	11.16
515	1	4	aug	sat	94.4	146.0	614.7	11.3	25.6	42	4.0	0.0	0.00
516	6	3	nov	tue	79.5	3.0	106.7	1.1	11.8	31	4.5	0.0	0.00

517 rows × 13 columns

In [6]:

```
#remove duplicates
df.drop_duplicates(inplace=True)
```

In [7]:

```
df.dropna()
```

Out[7]:

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
0	7	5	mar	fri	86.2	26.2	94.3	5.1	8.2	51	6.7	0.0	0.00
1	7	4	oct	tue	90.6	35.4	669.1	6.7	18.0	33	0.9	0.0	0.00
2	7	4	oct	sat	90.6	43.7	686.9	6.7	14.6	33	1.3	0.0	0.00
3	8	6	mar	fri	91.7	33.3	77.5	9.0	8.3	97	4.0	0.2	0.00
4	8	6	mar	sun	89.3	51.3	102.2	9.6	11.4	99	1.8	0.0	0.00
...	...	...	...	...	...	...	...	...	...	...	...	...	...
512	4	3	aug	sun	81.6	56.7	665.6	1.9	27.8	32	2.7	0.0	6.44
513	2	4	aug	sun	81.6	56.7	665.6	1.9	21.9	71	5.8	0.0	54.29
514	7	4	aug	sun	81.6	56.7	665.6	1.9	21.2	70	6.7	0.0	11.16
515	1	4	aug	sat	94.4	146.0	614.7	11.3	25.6	42	4.0	0.0	0.00
516	6	3	nov	tue	79.5	3.0	106.7	1.1	11.8	31	4.5	0.0	0.00

513 rows × 13 columns

In [8]:

```
df1=df.iloc[0:50]
```

In [10]:

```
df1
```

Out[10]:

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
0	7	5	mar	fri	86.2	26.2	94.3	5.1	8.2	51	6.7	0.0	0.0
1	7	4	oct	tue	90.6	35.4	669.1	6.7	18.0	33	0.9	0.0	0.0
2	7	4	oct	sat	90.6	43.7	686.9	6.7	14.6	33	1.3	0.0	0.0
3	8	6	mar	fri	91.7	33.3	77.5	9.0	8.3	97	4.0	0.2	0.0
4	8	6	mar	sun	89.3	51.3	102.2	9.6	11.4	99	1.8	0.0	0.0
5	8	6	aug	sun	92.3	85.3	488.0	14.7	22.2	29	5.4	0.0	0.0
6	8	6	aug	mon	92.3	88.9	495.6	8.5	24.1	27	3.1	0.0	0.0
7	8	6	aug	mon	91.5	145.4	608.2	10.7	8.0	86	2.2	0.0	0.0
8	8	6	sep	tue	91.0	129.5	692.6	7.0	13.1	63	5.4	0.0	0.0
9	7	5	sep	sat	92.5	88.0	698.6	7.1	22.8	40	4.0	0.0	0.0
10	7	5	sep	sat	92.5	88.0	698.6	7.1	17.8	51	7.2	0.0	0.0
11	7	5	sep	sat	92.8	73.2	713.0	22.6	19.3	38	4.0	0.0	0.0
12	6	5	aug	fri	63.5	70.8	665.3	0.8	17.0	72	6.7	0.0	0.0
13	6	5	sep	mon	90.9	126.5	686.5	7.0	21.3	42	2.2	0.0	0.0
14	6	5	sep	wed	92.9	133.3	699.6	9.2	26.4	21	4.5	0.0	0.0
15	6	5	sep	fri	93.3	141.2	713.9	13.9	22.9	44	5.4	0.0	0.0
16	5	5	mar	sat	91.7	35.8	80.8	7.8	15.1	27	5.4	0.0	0.0
17	8	5	oct	mon	84.9	32.8	664.2	3.0	16.7	47	4.9	0.0	0.0
18	6	4	mar	wed	89.2	27.9	70.8	6.3	15.9	35	4.0	0.0	0.0
19	6	4	apr	sat	86.3	27.4	97.1	5.1	9.3	44	4.5	0.0	0.0
20	6	4	sep	tue	91.0	129.5	692.6	7.0	18.3	40	2.7	0.0	0.0
21	5	4	sep	mon	91.8	78.5	724.3	9.2	19.1	38	2.7	0.0	0.0
22	7	4	jun	sun	94.3	96.3	200.0	56.1	21.0	44	4.5	0.0	0.0
23	7	4	aug	sat	90.2	110.9	537.4	6.2	19.5	43	5.8	0.0	0.0
24	7	4	aug	sat	93.5	139.4	594.2	20.3	23.7	32	5.8	0.0	0.0
25	7	4	aug	sun	91.4	142.4	601.4	10.6	16.3	60	5.4	0.0	0.0
26	7	4	sep	fri	92.4	117.9	668.0	12.2	19.0	34	5.8	0.0	0.0
27	7	4	sep	mon	90.9	126.5	686.5	7.0	19.4	48	1.3	0.0	0.0
28	6	3	sep	sat	93.4	145.4	721.4	8.1	30.2	24	2.7	0.0	0.0
29	6	3	sep	sun	93.5	149.3	728.6	8.1	22.8	39	3.6	0.0	0.0
30	6	3	sep	fri	94.3	85.1	692.3	15.9	25.4	24	3.6	0.0	0.0
31	6	3	sep	mon	88.6	91.8	709.9	7.1	11.2	78	7.6	0.0	0.0
32	6	3	sep	fri	88.6	69.7	706.8	5.8	20.6	37	1.8	0.0	0.0
33	6	3	sep	sun	91.7	75.6	718.3	7.8	17.7	39	3.6	0.0	0.0
34	6	3	sep	mon	91.8	78.5	724.3	9.2	21.2	32	2.7	0.0	0.0
35	6	3	sep	tue	90.3	80.7	730.2	6.3	18.2	62	4.5	0.0	0.0
36	6	3	oct	tue	90.6	35.4	669.1	6.7	21.7	24	4.5	0.0	0.0

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
37	7	4	oct	fri	90.0	41.5	682.6	8.7	11.3	60	5.4	0.0	0.0
38	7	3	oct	sat	90.6	43.7	686.9	6.7	17.8	27	4.0	0.0	0.0
39	4	4	mar	tue	88.1	25.7	67.6	3.8	14.1	43	2.7	0.0	0.0
40	4	4	jul	tue	79.5	60.6	366.7	1.5	23.3	37	3.1	0.0	0.0
41	4	4	aug	sat	90.2	96.9	624.2	8.9	18.4	42	6.7	0.0	0.0
42	4	4	aug	tue	94.8	108.3	647.1	17.0	16.6	54	5.4	0.0	0.0
43	4	4	sep	sat	92.5	88.0	698.6	7.1	19.6	48	2.7	0.0	0.0
44	4	4	sep	wed	90.1	82.9	735.7	6.2	12.9	74	4.9	0.0	0.0
45	5	6	sep	wed	94.3	85.1	692.3	15.9	25.9	24	4.0	0.0	0.0
46	5	6	sep	mon	90.9	126.5	686.5	7.0	14.7	70	3.6	0.0	0.0
47	6	6	jul	mon	94.2	62.3	442.9	11.0	23.0	36	3.1	0.0	0.0
48	4	4	mar	mon	87.2	23.9	64.7	4.1	11.8	35	1.8	0.0	0.0
49	4	4	mar	mon	87.6	52.2	103.8	5.0	11.0	46	5.8	0.0	0.0

In [12]:

```
df1.shape
```

Out[12]:

```
(50, 13)
```

In [13]:

```
df2=df.iloc[51:101]
```

In [14]:

```
df2
```

Out[14]:

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
51	4	3	aug	sun	90.2	99.6	631.2	6.3	21.5	34	2.2	0.0	0.0
52	4	3	aug	wed	92.1	111.2	654.1	9.6	20.4	42	4.9	0.0	0.0
54	4	3	aug	thu	91.7	114.3	661.3	6.3	17.6	45	3.6	0.0	0.0
55	4	3	sep	thu	92.9	137.0	706.4	9.2	27.7	24	2.2	0.0	0.0
56	4	3	sep	tue	90.3	80.7	730.2	6.3	17.8	63	4.9	0.0	0.0
57	4	3	oct	sun	92.6	46.5	691.8	8.8	13.8	50	2.7	0.0	0.0
58	2	2	feb	mon	84.0	9.3	34.0	2.1	13.9	40	5.4	0.0	0.0
59	2	2	feb	fri	86.6	13.2	43.0	5.3	12.3	51	0.9	0.0	0.0
60	2	2	mar	sun	89.3	51.3	102.2	9.6	11.5	39	5.8	0.0	0.0
61	2	2	mar	sun	89.3	51.3	102.2	9.6	5.5	59	6.3	0.0	0.0
62	2	2	aug	thu	93.0	75.3	466.6	7.7	18.8	35	4.9	0.0	0.0
63	2	2	aug	sun	90.2	99.6	631.2	6.3	20.8	33	2.7	0.0	0.0
64	2	2	aug	mon	91.1	103.2	638.8	5.8	23.1	31	3.1	0.0	0.0
65	2	2	aug	thu	91.7	114.3	661.3	6.3	18.6	44	4.5	0.0	0.0
66	2	2	sep	fri	92.4	117.9	668.0	12.2	23.0	37	4.5	0.0	0.0
67	2	2	sep	fri	92.4	117.9	668.0	12.2	19.6	33	5.4	0.0	0.0
68	2	2	sep	fri	92.4	117.9	668.0	12.2	19.6	33	6.3	0.0	0.0
69	4	5	mar	fri	91.7	33.3	77.5	9.0	17.2	26	4.5	0.0	0.0
70	4	5	mar	fri	91.2	48.3	97.8	12.5	15.8	27	7.6	0.0	0.0
71	4	5	sep	fri	94.3	85.1	692.3	15.9	17.7	37	3.6	0.0	0.0
72	5	4	mar	fri	91.7	33.3	77.5	9.0	15.6	25	6.3	0.0	0.0
73	5	4	aug	tue	88.8	147.3	614.5	9.0	17.3	43	4.5	0.0	0.0
74	5	4	sep	fri	93.3	141.2	713.9	13.9	27.6	30	1.3	0.0	0.0
75	9	9	feb	thu	84.2	6.8	26.6	7.7	6.7	79	3.1	0.0	0.0
76	9	9	feb	fri	86.6	13.2	43.0	5.3	15.7	43	3.1	0.0	0.0
77	1	3	mar	mon	87.6	52.2	103.8	5.0	8.3	72	3.1	0.0	0.0
78	1	2	aug	fri	90.1	108.0	529.8	12.5	14.7	66	2.7	0.0	0.0
79	1	2	aug	tue	91.0	121.2	561.6	7.0	21.6	19	6.7	0.0	0.0
80	1	2	aug	sun	91.4	142.4	601.4	10.6	19.5	39	6.3	0.0	0.0
81	1	2	aug	sun	90.2	99.6	631.2	6.3	17.9	44	2.2	0.0	0.0
82	1	2	aug	tue	94.8	108.3	647.1	17.0	18.6	51	4.5	0.0	0.0
83	1	2	aug	wed	92.1	111.2	654.1	9.6	16.6	47	0.9	0.0	0.0
84	1	2	aug	thu	91.7	114.3	661.3	6.3	20.2	45	3.6	0.0	0.0
85	1	2	sep	thu	92.9	137.0	706.4	9.2	21.5	15	0.9	0.0	0.0
86	1	2	sep	thu	92.9	137.0	706.4	9.2	25.4	27	2.2	0.0	0.0
87	1	2	sep	thu	92.9	137.0	706.4	9.2	22.4	34	2.2	0.0	0.0
88	1	2	sep	sun	93.5	149.3	728.6	8.1	25.3	36	3.6	0.0	0.0

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
89	6	5	mar	sat	91.7	35.8	80.8	7.8	17.4	25	4.9	0.0	0.0
90	6	5	aug	sat	90.2	96.9	624.2	8.9	14.7	59	5.8	0.0	0.0
91	8	6	mar	fri	91.7	35.8	80.8	7.8	17.4	24	5.4	0.0	0.0
92	8	6	aug	sun	92.3	85.3	488.0	14.7	20.8	32	6.3	0.0	0.0
93	8	6	aug	sun	91.4	142.4	601.4	10.6	18.2	43	4.9	0.0	0.0
94	8	6	aug	mon	91.1	103.2	638.8	5.8	23.4	22	2.7	0.0	0.0
95	4	4	sep	sun	89.7	90.0	704.4	4.8	17.8	64	1.3	0.0	0.0
96	3	4	feb	sat	83.9	8.0	30.2	2.6	12.7	48	1.8	0.0	0.0
97	3	4	mar	sat	69.0	2.4	15.5	0.7	17.4	24	5.4	0.0	0.0
98	3	4	aug	sun	91.4	142.4	601.4	10.6	11.6	87	4.5	0.0	0.0
99	3	4	aug	sun	91.4	142.4	601.4	10.6	19.8	39	5.4	0.0	0.0
101	3	4	aug	tue	88.8	147.3	614.5	9.0	14.4	66	5.4	0.0	0.0
102	2	4	aug	tue	94.8	108.3	647.1	17.0	20.1	40	4.0	0.0	0.0

In [15]:

```
df2.shape
```

Out[15]:

```
(50, 13)
```

In [16]:

```
combine_df=pd.concat([df1,df2])
```

In [17]:

```
combine_df
```

Out[17]:

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
0	7	5	mar	fri	86.2	26.2	94.3	5.1	8.2	51	6.7	0.0	0.0
1	7	4	oct	tue	90.6	35.4	669.1	6.7	18.0	33	0.9	0.0	0.0
2	7	4	oct	sat	90.6	43.7	686.9	6.7	14.6	33	1.3	0.0	0.0
3	8	6	mar	fri	91.7	33.3	77.5	9.0	8.3	97	4.0	0.2	0.0
4	8	6	mar	sun	89.3	51.3	102.2	9.6	11.4	99	1.8	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
97	3	4	mar	sat	69.0	2.4	15.5	0.7	17.4	24	5.4	0.0	0.0
98	3	4	aug	sun	91.4	142.4	601.4	10.6	11.6	87	4.5	0.0	0.0
99	3	4	aug	sun	91.4	142.4	601.4	10.6	19.8	39	5.4	0.0	0.0
101	3	4	aug	tue	88.8	147.3	614.5	9.0	14.4	66	5.4	0.0	0.0
102	2	4	aug	tue	94.8	108.3	647.1	17.0	20.1	40	4.0	0.0	0.0

100 rows × 13 columns

In [18]:

```
#data transformation
from sklearn.preprocessing import MinMaxScaler
```

In [19]:

```
scaler=MinMaxScaler()
df['normalized_temp']=scaler.fit_transform(df[['temp']])
```

In [20]:

```
#one hot encoding
df_encoded=pd.get_dummies(df,columns=['day'])
```

In [21]:

```
df_agg=df.groupby('month')['temp'].mean()
```

In [22]:

```
df_agg
```

Out[22]:

```
month
apr    12.044444
aug    21.648352
dec     4.522222
feb     9.635000
jan     5.250000
jul    22.109375
jun    20.575000
mar    13.009434
may    14.650000
nov    11.800000
oct    17.093333
sep    19.612209
Name: temp, dtype: float64
```

In [23]:

```
#ERROR CORRECTING
```

In [24]:

```
import pandas as pd
from scipy import stats
```

In [25]:

```
z_score=stats.zscore(df['temp'])
threshold=20
```

In [29]:

```
df1=df[(z_score<threshold)]
```

In [30]:

```
df1
```

Out[30]:

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area	normalized_
0	7	5	mar	fri	86.2	26.2	94.3	5.1	8.2	51	6.7	0.0	0.00	0.1
1	7	4	oct	tue	90.6	35.4	669.1	6.7	18.0	33	0.9	0.0	0.00	0.5
2	7	4	oct	sat	90.6	43.7	686.9	6.7	14.6	33	1.3	0.0	0.00	0.3
3	8	6	mar	fri	91.7	33.3	77.5	9.0	8.3	97	4.0	0.2	0.00	0.1
4	8	6	mar	sun	89.3	51.3	102.2	9.6	11.4	99	1.8	0.0	0.00	0.2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
512	4	3	aug	sun	81.6	56.7	665.6	1.9	27.8	32	2.7	0.0	6.44	0.8
513	2	4	aug	sun	81.6	56.7	665.6	1.9	21.9	71	5.8	0.0	54.29	0.6
514	7	4	aug	sun	81.6	56.7	665.6	1.9	21.2	70	6.7	0.0	11.16	0.6
515	1	4	aug	sat	94.4	146.0	614.7	11.3	25.6	42	4.0	0.0	0.00	0.7
516	6	3	nov	tue	79.5	3.0	106.7	1.1	11.8	31	4.5	0.0	0.00	0.3

513 rows × 14 columns

In [28]:

```
#model building
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report
```

In [31]:

```
X=df.drop('RH',axis=1)
y=df['RH']
```

In [32]:

```
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2,random_state=0)
```

In [33]:

```
model=LogisticRegression()
```

In [34]:

```
model.fit(X_train,y_train)
```

```
---------------------------------------------------------------------------
-
ValueError                                Traceback (most recent call las
t)
Cell In[34], line 1
----> 1 model.fit(X_train,y_train)

File ~\AppData\Roaming\Python\Python311\site-packages\sklearn\linear_model
\_logistic.py:1196, in LogisticRegression.fit(self, X, y, sample_weight)
   1193 else:
   1194     _dtype = [np.float64, np.float32]
-> 1196 X, y = self._validate_data(
   1197     X,
   1198     y,
   1199     accept_sparse="csr",
   1200     dtype=_dtype,
   1201     order="C",
   1202     accept_large_sparse=solver not in ["liblinear", "sag", "sag
a"],
   1203 )
   1204 check_classification_targets(y)
   1205 self.classes_ = np.unique(y)

File ~\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:584,
in BaseEstimator._validate_data(self, X, y, reset, validate_separately, **
check_params)
    582         y = check_array(y, input_name="y", **check_y_params)
    583     else:
--> 584         X, y = check_X_y(X, y, **check_params)
    585     out = X, y
    587 if not no_val_X and check_params.get("ensure_2d", True):

File ~\AppData\Roaming\Python\Python311\site-packages\sklearn\utils\valida
tion.py:1106, in check_X_y(X, y, accept_sparse, accept_large_sparse, dtyp
e, order, copy, force_all_finite, ensure_2d, allow_nd, multi_output, ensur
e_min_samples, ensure_min_features, y_numeric, estimator)
   1101         estimator_name = _check_estimator_name(estimator)
   1102     raise ValueError(
   1103         f"{estimator_name} requires y to be passed, but the target
y is None"
   1104     )
-> 1106 X = check_array(
   1107     X,
   1108     accept_sparse=accept_sparse,
   1109     accept_large_sparse=accept_large_sparse,
   1110     dtype=dtype,
   1111     order=order,
   1112     copy=copy,
   1113     force_all_finite=force_all_finite,
   1114     ensure_2d=ensure_2d,
   1115     allow_nd=allow_nd,
   1116     ensure_min_samples=ensure_min_samples,
   1117     ensure_min_features=ensure_min_features,
   1118     estimator=estimator,
   1119     input_name="X",
   1120 )
   1122 y = _check_y(y, multi_output=multi_output, y_numeric=y_numeric, es
timator=estimator)
   1124 check_consistent_length(X, y)

File ~\AppData\Roaming\Python\Python311\site-packages\sklearn\utils\valida
```

```
tion.py:879, in check_array(array, accept_sparse, accept_large_sparse, dty
pe, order, copy, force_all_finite, ensure_2d, allow_nd, ensure_min_sample
s, ensure_min_features, estimator, input_name)
    877         array = xp.astype(array, dtype, copy=False)
    878     else:
--> 879         array = _asarray_with_order(array, order=order, dtype=dtyp
e, xp=xp)
    880 except ComplexWarning as complex_warning:
    881     raise ValueError(
    882         "Complex data not supported\n{}\n".format(array)
    883     ) from complex_warning

File ~\AppData\Roaming\Python\Python311\site-packages\sklearn\utils\_array
_api.py:185, in _asarray_with_order(array, dtype, order, copy, xp)
    182     xp, _ = get_namespace(array)
    183 if xp.__name__ in {"numpy", "numpy.array_api"}:
    184     # Use NumPy API to support order
--> 185     array = numpy.asarray(array, order=order, dtype=dtype)
    186     return xp.asarray(array, copy=copy)
    187 else:

File ~\AppData\Roaming\Python\Python311\site-packages\pandas\core\generic.
py:2070, in NDFrame.__array__(self, dtype)
   2069 def __array__(self, dtype: npt.DTypeLike | None = None) -> np.ndar
ray:
-> 2070     return np.asarray(self._values, dtype=dtype)

ValueError: could not convert string to float: 'jun'
```

```
In [ ]:
```