

```
#importing the lib
import pandas as pd
import numpy as np

dataset=pd.read_csv("supplement.csv")
dataset
```

	ID	Store_id	Store_Type	Location_Type	Region_Code	Date	Holiday	Discount	#Order	Sales
0	T1000001	1	S1	L3	R1	2018-01-01	1	Yes	9	1
1	T1000002	253	S4	L2	R1	2018-01-01	1	Yes	60	5
2	T1000003	252	S3	L2	R1	2018-01-01	1	Yes	42	36
3	T1000004	251	S2	L3	R1	2018-01-01	1	Yes	23	19
4	T1000005	250	S2	L3	R4	2018-01-01	1	Yes	62	49
...
188335	T1188336	149	S2	L3	R2	2019-05-31	1	Yes	51	31
188336	T1188337	153	S4	L2	R1	2019-05-31	1	No	90	54
...

```
dataset.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 188340 entries, 0 to 188339
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  ---
0    ID                    188340 non-null object
1    Store_id              188340 non-null int64
2    Store_Type            188340 non-null object
3    Location_Type         188340 non-null object
4    Region_Code           188340 non-null object
5    Date                  188340 non-null object
6    Holiday               188340 non-null int64
7    Discount              188340 non-null object
8    #Order                188340 non-null int64
9    Sales                 188340 non-null float64
dtypes: float64(1), int64(3), object(6)
memory usage: 14.4+ MB
```

```
dataset.isnull()

ID      Store_id  Store_Type  Location_Type  Region_Code  Date  Holiday  Discount  #Order  Sales
0      False     False      False          False        False False     False     False   False
1      False     False      False          False        False False     False     False   False
2      False     False      False          False        False False     False     False   False
3      False     False      False          False        False False     False     False   False
4      False     False      False          False        False False     False     False   False
...     ...       ...       ...           ...          ...   ...      ...      ...     ...     ...
188335  False     False      False          False        False False     False     False   False
188336  False     False      False          False        False False     False     False   False
188337  False     False      False          False        False False     False     False   False
188338  False     False      False          False        False False     False     False   False
188339  False     False      False          False        False False     False     False   False
188340 rows x 10 columns
```

```
dataset.isnull().sum()

ID      0
Store_id 0
Store_Type 0
Location_Type 0
```

```
Region_Code      0
Date             0
Holiday          0
Discount         0
#Order           0
Sales           0
dtype: int64

dataset.describe()
```

	Store_id	Holiday	#Order	Sales
count	188340.000000	188340.000000	188340.000000	188340.000000
mean	183.000000	0.131783	68.205692	42784.327982
std	105.366308	0.338256	30.467415	18456.708302
min	1.000000	0.000000	0.000000	0.000000
25%	92.000000	0.000000	48.000000	30426.000000
50%	183.000000	0.000000	63.000000	39678.000000
75%	274.000000	0.000000	82.000000	51909.000000
max	365.000000	1.000000	371.000000	247215.000000

```
!pip install plotly

Requirement already satisfied: plotly in c:\users\dell\anaconda3\lib\site-packages (5.9.0)
Requirement already satisfied: tenacity>=6.2.0 in c:\users\dell\anaconda3\lib\site-packages (from plotly) (8.0.1)

import plotly.express as px

pie=dataset["Store_Type"].value_counts()
store=pie.index
orders=pie.values

fig=px.pie(dataset,values=orders,names=store)
fig.show()
```

```
pie=dataset["Location_Type"].value_counts()
location=pie.index
orders=pie.values

fig=px.pie(dataset,values=orders,names=location)
fig.show()
```



```
pie=dataset["Discount"].value_counts()
discount=pie.index
orders=pie.values

fig=px.pie(dataset,values=orders,names=discount)
fig.show()
```

```
pie4=dataset["Holiday"].value_counts()
holiday=pie4.index
orders=pie4.values

fig=px.pie(dataset,values=orders,names=holiday)
fig.show()
```

```
dataset["Discount"]=dataset["Discount"].map({"No":0,"Yes":1})
```

dataset

	ID	Store_id	Store_Type	Location_Type	Region_Code	Date	Holiday	Discount	#Order	Sales
0	T1000001	1	S1	L3	R1	2018-01-01	1	1	9	7011.84
1	T1000002	253	S4	L2	R1	2018-01-01	1	1	60	51789.12
2	T1000003	252	S3	L2	R1	2018-01-01	1	1	42	36868.20
3	T1000004	251	S2	L3	R1	2018-01-01	1	1	23	19715.16
4	T1000005	250	S2	L3	R4	2018-01-01	1	1	62	45614.52
...
188335	T1188336	149	S2	L3	R2	2019-05-31	1	1	51	37272.00
188336	T1188337	153	S4	L2	R1	2019-05-31	1	0	90	54572.64
188337	T1188338	154	S1	L3	R2	2019-05-31	1	0	56	31624.56
188338	T1188339	155	S3	L1	R2	2019-05-31	1	1	70	49162.41
188339	T1188340	152	S2	L1	R1	2019-05-31	1	0	47	37977.00

188340 rows × 10 columns

```
dataset["Store_Type"]=dataset["Store_Type"].map({"S1":1,"S2":2,"S3":3,"S4":4})
```

dataset

	ID	Store_id	Store_Type	Location_Type	Region_Code	Date	Holiday	Discount	#Order	Sales
0	T1000001	1	1	L3	R1	2018-01-01	1	1	9	7011.84
1	T1000002	253	4	L2	R1	2018-01-01	1	1	60	51789.12
2	T1000003	252	3	L2	R1	2018-01-01	1	1	42	36868.20
3	T1000004	251	2	L3	R1	2018-01-01	1	1	23	19715.16
4	T1000005	250	2	L3	R4	2018-01-01	1	1	62	45614.52
...
188335	T1188336	149	2	L3	R2	2019-05-31	1	1	51	37272.00
188336	T1188337	153	4	L2	R1	2019-05-31	1	0	90	54572.64
188337	T1188338	154	1	L3	R2	2019-05-31	1	0	56	31624.56
188338	T1188339	155	3	L1	R2	2019-05-31	1	1	70	49162.41
188339	T1188340	152	2	L1	R1	2019-05-31	1	0	47	37977.00

188340 rows × 10 columns

```
dataset["Location_Type"]=dataset["Location_Type"].map({"L1":1,"L2":2,"L3":3,"L4":4,"L5":5})
```

dataset

	ID	Store_id	Store_Type	Location_Type	Region_Code	Date	Holiday	Discount	#Order	Sales
0	T1000001	1	1	3	R1	2018-01-01	1	1	9	7011.84
1	T1000002	253	4	2	R1	2018-01-01	1	1	60	51789.12
2	T1000003	252	3	2	R1	2018-01-01	1	1	42	36868.20
3	T1000004	251	2	3	R1	2018-01-01	1	1	23	19715.16
4	T1000005	250	2	3	R4	2018-01-01	1	1	62	45614.52
...
188335	T1188336	149	2	3	R2	2019-05-31	1	1	51	37272.00
188336	T1188337	153	4	2	R1	2019-05-31	1	0	90	54572.64

```
x=np.array(dataset[["Store_Type","Location_Type","Holiday","Discount"]])
```

```
y=np.array(dataset[["#Order"]])
```

```
188335 T1188336 149 2 3 R2 2019-05-31 1 1 51 37272.00
```

```
x
```

```
array([[1, 3, 1, 1],
       [4, 2, 1, 1],
       [3, 2, 1, 1],
       ...,
       [1, 3, 1, 0],
       [3, 1, 1, 1],
       [2, 1, 1, 0]], dtype=int64)
```

```
y
```

```
array([[ 9],
       [60],
       [42],
       ...,
       [56],
       [70],
       [47]], dtype=int64)
```

```
#building the ML model
```

```
from sklearn.model_selection import train_test_split
```

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=42)
```

```
len(x_train)
```

```
150672
```

```
!pip install lightgbm
```

```
Collecting lightgbm
  Downloading lightgbm-3.3.5-py3-none-win_amd64.whl (1.0 MB)
    ----- 1.0/1.0 MB 3.6 MB/s eta 0:00:00
Requirement already satisfied: scikit-learn!=0.22.0 in c:\users\dell\anaconda3\lib\site-packages (from lightgbm) (1.2.1)
Requirement already satisfied: numpy in c:\users\dell\anaconda3\lib\site-packages (from lightgbm) (1.23.5)
Requirement already satisfied: wheel in c:\users\dell\anaconda3\lib\site-packages (from lightgbm) (0.38.4)
Requirement already satisfied: scipy in c:\users\dell\anaconda3\lib\site-packages (from lightgbm) (1.10.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\dell\anaconda3\lib\site-packages (from scikit-learn!=0.22.0->lightg
Requirement already satisfied: joblib>=1.1.1 in c:\users\dell\anaconda3\lib\site-packages (from scikit-learn!=0.22.0->lightgbm) (1.
Installing collected packages: lightgbm
Successfully installed lightgbm-3.3.5
```

```
import lightgbm as ltb
model=ltb.LGBMRegressor()
```

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

```
C:\Users\Dell\anaconda3\lib\site-packages\sklearn\utils\validation.py:1143: DataConversionWarning:
```

```
A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples, ), for example using ravel()
```

```
▼ LGBMRegressor
LGBMRegressor()
```

```
y_pred=model.predict(x_test)
```

```
y_pred

array([47.35189701, 97.06871721, 66.57778822, ..., 47.35189701,
       61.74938636, 85.34103853])

y_test

array([[ 54],
       [111],
       [ 59],
       ...,
       [ 40],
       [ 69],
       [ 68]], dtype=int64)

data =pd.DataFrame(data={"Predicted Orders ":y_pred.flatten()})
```

data

Predicted Orders	
0	47.351897
1	97.068717
2	66.577788
3	85.143083
4	54.451098
...	...
37663	66.577788
37664	47.722874
37665	47.351897
37666	61.749386
37667	85.341039

37668 rows × 1 columns