In [1]:
```
import pandas as pd
import numpy as np

#Date stuff
from datetime import datetime
from datetime import timedelta

#Library for Nice graphing
import seaborn as sns
import matplotlib.pyplot as plt
import statsmodels.formula.api as sn
%matplotlib inline

#Library for statistics operation
import scipy.stats as stats

# Date Time library
from datetime import datetime

#Machine learning Library
import statsmodels.api as sm
from sklearn import metrics
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import AdaBoostRegressor
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.svm import SVC, LinearSVC
from sklearn.metrics import mean_squared_error as mse
from sklearn.metrics import mean_absolute_error, mean_squared_error
from sklearn.model_selection import train_test_split



# Ignore warnings
import warnings
warnings.filterwarnings('ignore')

# Settings
pd.set_option('display.max_columns', None)
np.set_printoptions(threshold=1000)
np.set_printoptions(precision=3)
sns.set(style="darkgrid")
plt.rcParams['axes.labelsize'] = 14
plt.rcParams['xtick.labelsize'] = 12
plt.rcParams['ytick.labelsize'] = 12
```

Import Data from Csv files

In [2]:
```
train=pd.read_csv('C:/Users/DELL/Downloads/Data/Sales Forecasting/train.csv')
test=pd.read_csv('C:/Users/DELL/Downloads/Data/Sales Forecasting/test.csv')
stores=pd.read_csv('C:/Users/DELL/Downloads/Data/Sales Forecasting/stores.csv')
features=pd.read_csv('C:/Users/DELL/Downloads/Data/Sales Forecasting/features.cs
```

Merge the data sets:

In [3]:
```
# For Train data set
train_bt = pd.merge(train,stores)
train = pd.merge(train_bt,features)

#For test data set
test_bt = pd.merge(test,stores)
test= pd.merge(test_bt,features)
```

In [4]:
```
train.head(2)
```

Out[4]:

	Store	Dept	Date	Weekly_Sales	IsHoliday	Type	Size	Temperature	Fuel_Price
0	1	1	2010-02-05	24924.50	False	A	151315	42.31	2.572
1	1	2	2010-02-05	50605.27	False	A	151315	42.31	2.572

In [5]:
```
test.head(2)
```

Out[5]:

	Store	Dept	Date	IsHoliday	Type	Size	Temperature	Fuel_Price	MarkDown1
0	1	1	2012-11-02	False	A	151315	55.32	3.386	6766.44
1	1	2	2012-11-02	False	A	151315	55.32	3.386	6766.44

In [6]:
```
print (train.info())
print ("########################################")
print (test.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 421570 entries, 0 to 421569
Data columns (total 16 columns):
 #   Column        Non-Null Count   Dtype  
---  ------        --------------   -----  
 0   Store         421570 non-null  int64  
 1   Dept          421570 non-null  int64  
 2   Date          421570 non-null  object 
 3   Weekly_Sales  421570 non-null  float64
 4   IsHoliday     421570 non-null  bool   
 5   Type          421570 non-null  object 
 6   Size          421570 non-null  int64  
 7   Temperature   421570 non-null  float64
 8   Fuel_Price    421570 non-null  float64
 9   MarkDown1     150681 non-null  float64
 10  MarkDown2     111248 non-null  float64
 11  MarkDown3     137091 non-null  float64
 12  MarkDown4     134967 non-null  float64
 13  MarkDown5     151432 non-null  float64
 14  CPI           421570 non-null  float64
 15  Unemployment  421570 non-null  float64
dtypes: bool(1), float64(10), int64(3), object(2)
memory usage: 48.6+ MB
None
****************************************
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 115064 entries, 0 to 115063
Data columns (total 15 columns):
 #   Column        Non-Null Count   Dtype  
---  ------        --------------   -----  
 0   Store         115064 non-null  int64  
 1   Dept          115064 non-null  int64  
 2   Date          115064 non-null  object 
 3   IsHoliday     115064 non-null  bool   
 4   Type          115064 non-null  object 
 5   Size          115064 non-null  int64  
 6   Temperature   115064 non-null  float64
 7   Fuel_Price    115064 non-null  float64
 8   MarkDown1     114915 non-null  float64
 9   MarkDown2     86437 non-null   float64
 10  MarkDown3     105235 non-null  float64
 11  MarkDown4     102176 non-null  float64
 12  MarkDown5     115064 non-null  float64
 13  CPI           76902 non-null   float64
 14  Unemployment  76902 non-null   float64
dtypes: bool(1), float64(9), int64(3), object(2)
memory usage: 12.4+ MB
None
```

In [7]:
```
# take only those values whose sales is positive.
train = train[train['Weekly_Sales']>0]
```

Data Description

1. Training Data

In [8]:
```
numeric_var_train=[key for key in dict(train.dtypes) if dict(train.dtypes)[key]
cat_var_train=[key for key in dict(train.dtypes) if dict(train.dtypes)[key] in [

# Train Numerical Data
train_num=train[numeric_var_train]

# Train Categorical Data
train_cat=train[cat_var_train]

print (numeric_var_train)
print (cat_var_train)
```

```
['Store', 'Dept', 'Weekly_Sales', 'Size', 'Temperature', 'Fuel_Price', 'MarkDown
1', 'MarkDown2', 'MarkDown3', 'MarkDown4', 'MarkDown5', 'CPI', 'Unemployment']
['Date', 'Type']
```

In [9]:
```
# Use a general function that returns multiple values
def var_summary(x):
    return pd.Series([x.count(), x.isnull().sum(), x.sum(), x.mean(), x.median()
```

In [10]:
```
num_summary=train_num.apply(lambda x: var_summary(x)).T
num_summary
```

Out[10]:

	N	NMISS	SUM	MEAN	MEDIAN	
Store	420212.0	0.0	9.326862e+06	22.195611	22.000000	12.
Dept	420212.0	0.0	1.859073e+07	44.241309	37.000000	30.
Weekly_Sales	420212.0	0.0	6.737307e+09	16033.114591	7661.700000	22729.
Size	420212.0	0.0	5.746388e+10	136749.732787	140167.000000	60993.
Temperature	420212.0	0.0	2.525079e+07	60.090599	62.090000	18.
Fuel_Price	420212.0	0.0	1.412286e+06	3.360890	3.452000	0.
MarkDown1	150181.0	270031.0	1.088485e+09	7247.821269	5347.450000	8293.
MarkDown2	110904.0	309308.0	3.693265e+08	3330.146158	192.000000	9460.
MarkDown3	136651.0	283561.0	1.970147e+08	1441.736203	24.600000	9631.
MarkDown4	134518.0	285694.0	4.553141e+08	3384.782267	1481.310000	6295.
MarkDown5	150929.0	269283.0	6.987306e+08	4629.531870	3359.450000	5960.
CPI	420212.0	0.0	7.194555e+07	171.212496	182.350989	39.
Unemployment	420212.0	0.0	3.344888e+06	7.960000	7.866000	1.

In [11]:
```
def cat_summary(x):
    return pd.Series([x.count(), x.isnull().sum(), x.value_counts()],
                     index=['N', 'NMISS', 'ColumnsNames'])

cat_summary=train_cat.apply(lambda x: cat_summary(x))
cat_summary
```

Out[11]:

	Date	Type
N	420212	420212
NMISS	0	0
ColumnsNames	Date 2011-12-23 3018 2011-11-25 3016 201...	Type A 214961 B 162787 C 42464 Name:...

2. Testing Data

In [12]:
```
numeric_var_test=[key for key in dict(test.dtypes) if dict(test.dtypes)[key] in
cat_var_test=[key for key in dict(test.dtypes) if dict(test.dtypes)[key] in ['ob

# Train Numerical Data
test_num=test[numeric_var_test]

# Train Categorical Data
test_cat=test[cat_var_test]

print (numeric_var_test)
print (cat_var_test)
```

```
['Store', 'Dept', 'Size', 'Temperature', 'Fuel_Price', 'MarkDown1', 'MarkDown2',
'MarkDown3', 'MarkDown4', 'MarkDown5', 'CPI', 'Unemployment']
['Date', 'Type']
```

In [13]:
```
# Numerical data summary report
num_summary=test_num.apply(lambda x: var_summary(x)).T

num_summary.head()
```

Out[13]:

	N	NMISS	SUM	MEAN	MEDIAN	STD
Store	115064.0	0.0	2.558817e+06	22.238207	22.000	12.809930
Dept	115064.0	0.0	5.101883e+06	44.339524	37.000	30.656410
Size	115064.0	0.0	1.570597e+10	136497.688921	140167.000	61106.926438
Temperature	115064.0	0.0	6.206760e+06	53.941804	54.470	18.724153
Fuel_Price	115064.0	0.0	4.121070e+05	3.581546	3.606	0.239442

In [14]:
```
# categorical data summary report
def cat_summary(x):
    return pd.Series([x.count(), x.isnull().sum(), x.value_counts()],
                     index=['N', 'NMISS', 'ColumnsNames'])

cat_summary=test_cat.apply(lambda x: cat_summary(x))
cat_summary
```

Out[14]:

	Date	Type
N	115064	115064
NMISS	0	0
ColumnsNames	Date 2012-12-21 3002 2012-12-07 2989 201...	Type A 58713 B 44500 C 11851 Name: co...

In [15]:
```
print(train.describe())
```

```
               Store           Dept   Weekly_Sales           Size  \
count  420212.000000  420212.000000  420212.000000  420212.000000   \
mean       22.195611      44.241309   16033.114591  136749.732787   
std        12.787236      30.508819   22729.492116   60993.084568   
min         1.000000       1.000000       0.010000   34875.000000   
25%        11.000000      18.000000    2120.130000   93638.000000   
50%        22.000000      37.000000    7661.700000  140167.000000   
75%        33.000000      74.000000   20271.265000  202505.000000   
max        45.000000      99.000000  693099.360000  219622.000000   

         Temperature     Fuel_Price      MarkDown1      MarkDown2  \
count  420212.000000  420212.000000  150181.000000  110904.000000   \
mean       60.090599       3.360890    7247.821269    3330.146158   
std        18.447857       0.458519    8293.028741    9460.395025   
min        -2.060000       2.472000       0.270000    -265.760000   
25%        46.680000       2.933000    2240.270000      41.600000   
50%        62.090000       3.452000    5347.450000     192.000000   
75%        74.280000       3.738000    9210.900000    1926.940000   
max       100.140000       4.468000   88646.760000  104519.540000   

          MarkDown3      MarkDown4      MarkDown5            CPI  \
count  136651.000000  134518.000000  150929.000000  420212.000000   \
mean     1441.736203    3384.782267    4629.531870     171.212496   
std      9631.968459    6295.136952    5960.171711      39.162445   
min       -29.100000       0.220000     135.160000     126.064000   
25%         5.100000     504.220000    1878.440000     132.022667   
50%        24.600000    1481.310000    3359.450000     182.350989   
75%       103.990000    3595.040000    5563.800000     212.445487   
max    141630.610000   67474.850000  108519.280000     227.232807   

       Unemployment  
count  420212.000000  
mean        7.960000  
std         1.863879  
min         3.879000  
25%         6.891000  
50%         7.866000  
75%         8.567000  
max        14.313000  
```

In [16]:
```
print(train.columns)
```

```
Index(['Store', 'Dept', 'Date', 'Weekly_Sales', 'IsHoliday', 'Type', 'Size',
       'Temperature', 'Fuel_Price', 'MarkDown1', 'MarkDown2', 'MarkDown3',
       'MarkDown4', 'MarkDown5', 'CPI', 'Unemployment'],
      dtype='object')
```

DATA PREPARATION & ANALYSIS

In [17]:
```
traindf=train.merge(features,how='left',indicator=True).merge(stores,how='left')
```

In [18]:
```
traindf
```

Out[18]:

	Store	Dept	Date	Weekly_Sales	IsHoliday	Type	Size	Temperature	Fuel
0	1	1	2010-02-05	24924.50	False	A	151315	42.31	
1	1	2	2010-02-05	50605.27	False	A	151315	42.31	
2	1	3	2010-02-05	13740.12	False	A	151315	42.31	
3	1	4	2010-02-05	39954.04	False	A	151315	42.31	
4	1	5	2010-02-05	32229.38	False	A	151315	42.31	
...	...	...	...	...	...	...	...	...	
420207	45	93	2012-10-26	2487.80	False	B	118221	58.85	
420208	45	94	2012-10-26	5203.31	False	B	118221	58.85	
420209	45	95	2012-10-26	56017.47	False	B	118221	58.85	
420210	45	97	2012-10-26	6817.48	False	B	118221	58.85	
420211	45	98	2012-10-26	1076.80	False	B	118221	58.85	

420212 rows × 17 columns

In [19]:
```
traindf1=traindf.drop(['MarkDown1','MarkDown2','MarkDown3','MarkDown4','MarkDowr
```

In [20]:
```
traindf1.isna().sum()
```

Out[20]:
```
Store           0
Dept            0
Date            0
Weekly_Sales    0
IsHoliday       0
Type            0
Size            0
Temperature     0
Fuel_Price      0
CPI             0
Unemployment    0
_merge          0
dtype: int64
```

In [21]:
```
traindf1.loc[traindf1['Weekly_Sales']<=0] #outliers
```

Out[21]:

	Store	Dept	Date	Weekly_Sales	IsHoliday	Type	Size	Temperature	Fuel_Price	CPI

In [22]:
```
traindf2=traindf1.loc[traindf1['Weekly_Sales']>0]
traindf3=traindf2.drop(['_merge'],axis=1)
```

In [23]:
```
traindf3.sort_values(by='Date')
```

Out[23]:

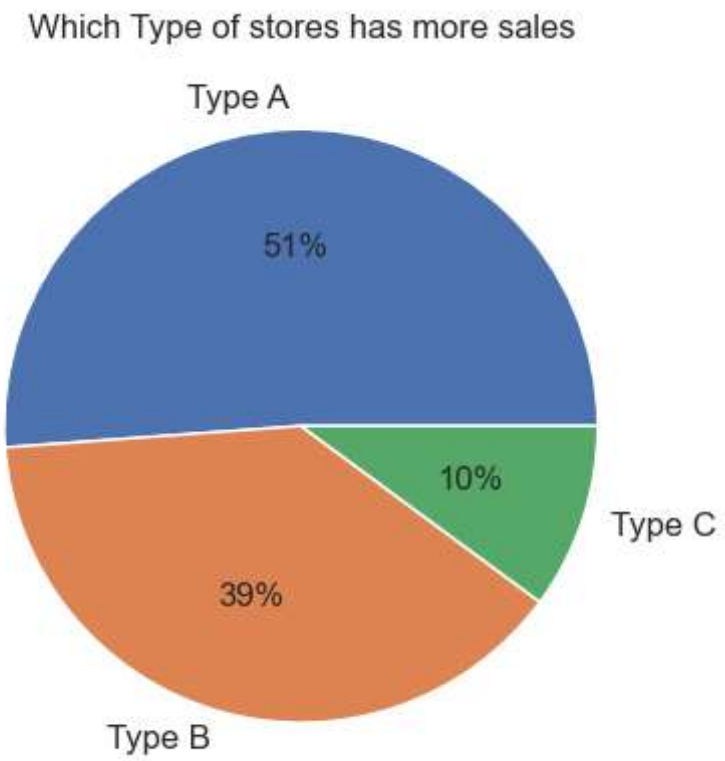
	Store	Dept	Date	Weekly_Sales	IsHoliday	Type	Size	Temperature	Fuel
0	1	1	2010-02-05	24924.50	False	A	151315	42.31	
329781	35	3	2010-02-05	14612.19	False	B	103681	27.19	
329782	35	4	2010-02-05	26323.15	False	B	103681	27.19	
329783	35	5	2010-02-05	36414.63	False	B	103681	27.19	
329784	35	6	2010-02-05	11437.81	False	B	103681	27.19	
...	...	...	...	...	...	...	...	...	
329722	34	14	2012-10-26	8930.71	False	A	158114	57.95	
329723	34	16	2012-10-26	4841.81	False	A	158114	57.95	
329724	34	17	2012-10-26	7035.13	False	A	158114	57.95	
329726	34	20	2012-10-26	2124.60	False	A	158114	57.95	
420211	45	98	2012-10-26	1076.80	False	B	118221	58.85	

420212 rows × 11 columns

In [24]:
```
traindf3['Type'].unique() #Store varities
```

Out[24]:
```
array(['A', 'B', 'C'], dtype=object)
```

In [25]:
```
# Import libraries
import matplotlib.pyplot as plt
import numpy as np


# Creating dataset
stores = ['Type A','Type B','Type C']

data = traindf3['Type'].value_counts()

# Creating plot
fig, ax = plt.subplots()
plt.pie(data, labels = stores,autopct='%.0f%%')
ax.set_title('Which Type of stores has more sales')
# show plot
plt.show()
```



In [26]:
```
traindf3['year'] = pd.DatetimeIndex(traindf3['Date']).year #Separating year data
```

In [27]:
```
data = traindf3.drop(['Date','Type'], axis=1)
```

In [28]:
```
data
```

Out[28]:

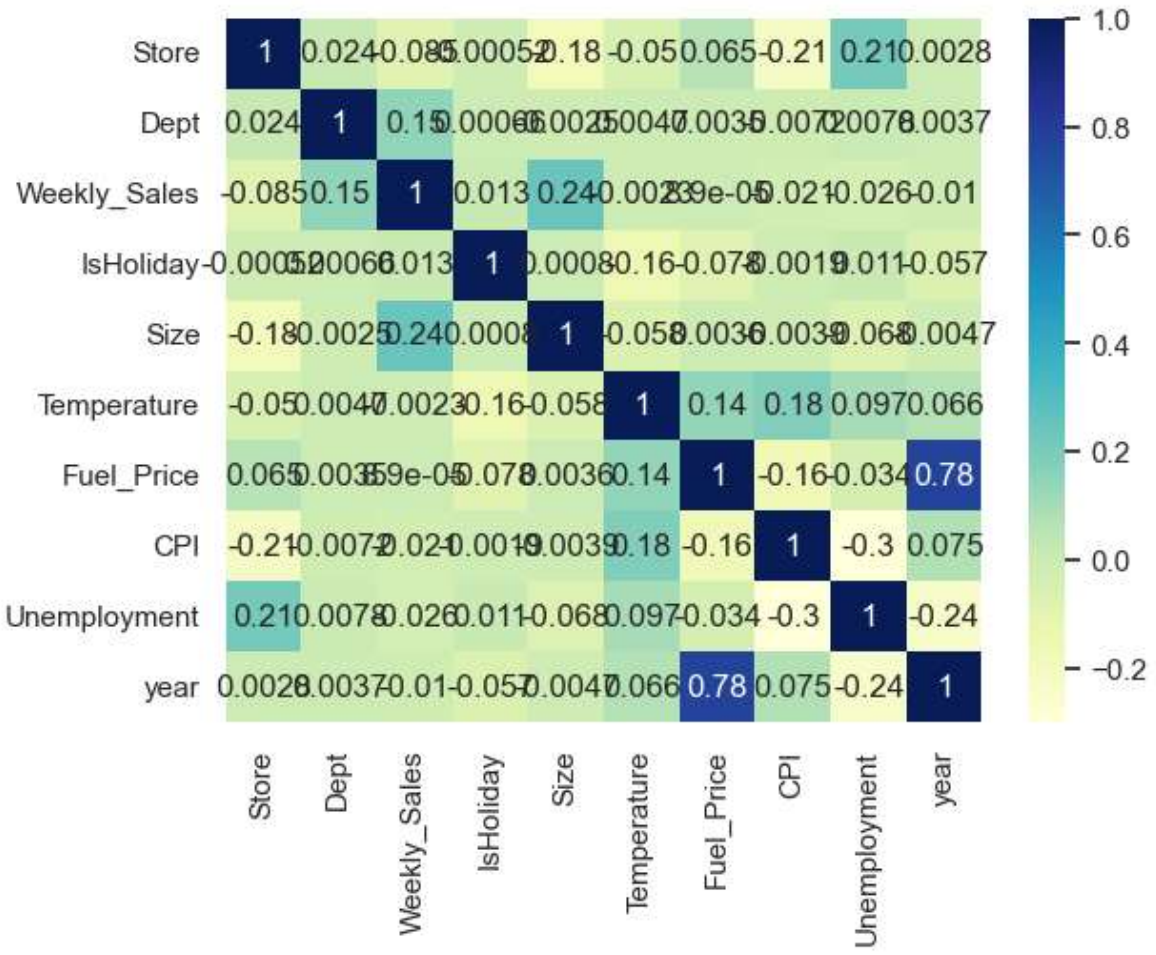
	Store	Dept	Weekly_Sales	IsHoliday	Size	Temperature	Fuel_Price	
0	1	1	24924.50	False	151315	42.31	2.572	211.09
1	1	2	50605.27	False	151315	42.31	2.572	211.09
2	1	3	13740.12	False	151315	42.31	2.572	211.09
3	1	4	39954.04	False	151315	42.31	2.572	211.09
4	1	5	32229.38	False	151315	42.31	2.572	211.09
...	...	...	...	...	...	...	...	
420207	45	93	2487.80	False	118221	58.85	3.882	192.30
420208	45	94	5203.31	False	118221	58.85	3.882	192.30
420209	45	95	56017.47	False	118221	58.85	3.882	192.30
420210	45	97	6817.48	False	118221	58.85	3.882	192.30
420211	45	98	1076.80	False	118221	58.85	3.882	192.30

420212 rows × 10 columns

In [29]:
```
# import modules
import matplotlib.pyplot as mp
```

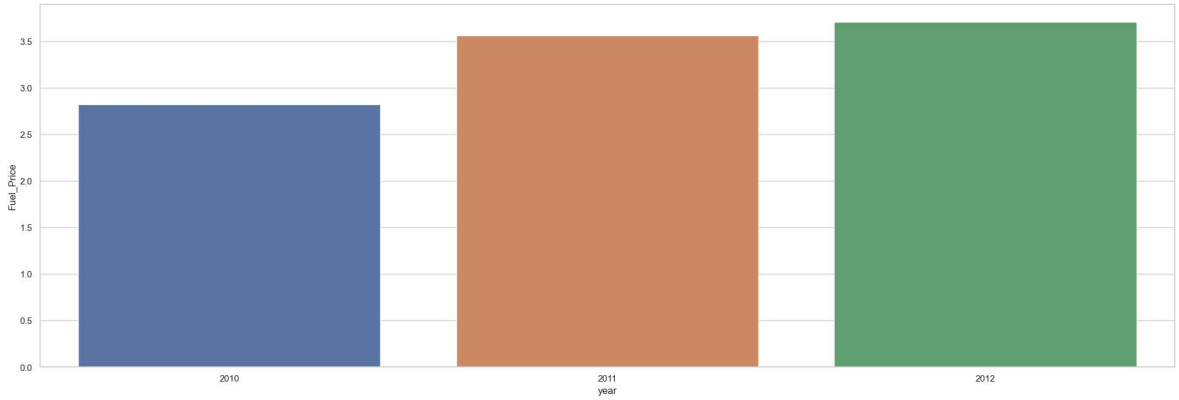
```
import pandas as pd
import seaborn as sns



# prints data that will be plotted
# columns shown here are selected by corr() since
# they are ideal for the plot
print(data.corr())
sns.set_theme(style="whitegrid")
# plotting correlation heatmap
dataplot = sns.heatmap(data.corr(), cmap="YlGnBu", annot=True)
sns.set(rc = {'figure.figsize':(25,8)})

# displaying heatmap
mp.show()
```

```
                Store      Dept  Weekly_Sales  IsHoliday      Size  \
Store        1.000000  0.024258     -0.085117  -0.000522 -0.182763   \
Dept         0.024258  1.000000      0.148749   0.000663 -0.002491
Weekly_Sales -0.085117  0.148749      1.000000   0.012843  0.244117
IsHoliday    -0.000522  0.000663      0.012843   1.000000  0.000797
Size         -0.182763 -0.002491      0.244117   0.000797  1.000000
Temperature  -0.050230  0.004727     -0.002339  -0.155775 -0.058413
Fuel_Price    0.065321  0.003544      0.000089  -0.078155  0.003632
CPI          -0.211261 -0.007178     -0.021162  -0.001933 -0.003903
Unemployment  0.208759  0.007787     -0.025806   0.010555 -0.068335
year          0.002831  0.003716     -0.010015  -0.056572 -0.004716

              Temperature  Fuel_Price       CPI  Unemployment      year
Store           -0.050230    0.065321 -0.211261      0.208759  0.002831
Dept             0.004727    0.003544 -0.007178      0.007787  0.003716
Weekly_Sales    -0.002339    0.000089 -0.021162     -0.025806 -0.010015
IsHoliday       -0.155775   -0.078155 -0.001933      0.010555 -0.056572
Size            -0.058413    0.003632 -0.003903     -0.068335 -0.004716
Temperature      1.000000    0.143700  0.182223      0.096768  0.065712
Fuel_Price       0.143700    1.000000 -0.164199     -0.033915  0.779681
CPI              0.182223   -0.164199  1.000000     -0.299887  0.074547
Unemployment     0.096768   -0.033915 -0.299887      1.000000 -0.237210
year             0.065712    0.779681  0.074547     -0.237210  1.000000
```



Year vs Fuel_price

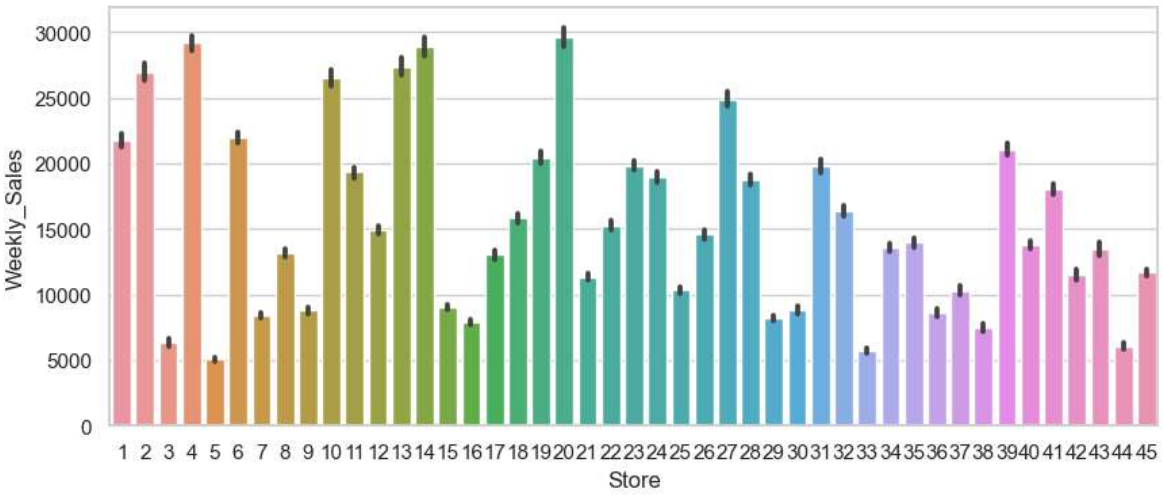
In [30]:
```
import seaborn as sns
sns.set_theme(style="whitegrid")
tips =traindf3
ax = sns.barplot(x="year", y="Fuel_Price", data=tips)
sns.set(rc = {'figure.figsize':(10,4)})
```



Weekly sales vs Store

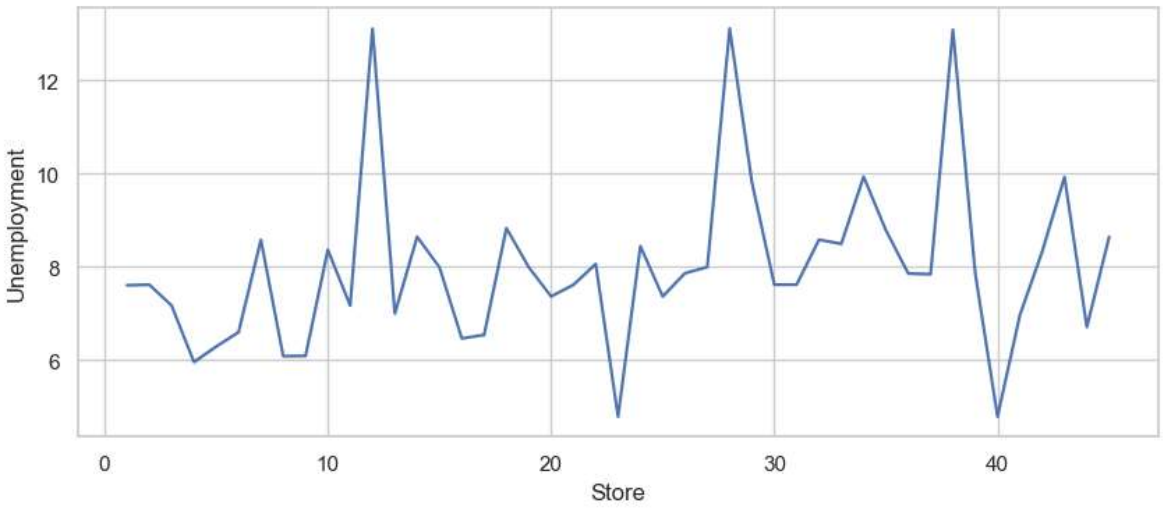
In [31]:
```
import seaborn as sns
sns.set_theme(style="whitegrid")
```

```
tips = traindf3
ax = sns.barplot(x='Store', y="Weekly_Sales", data=tips)
```



Store vs Unemployment

In [32]:
```
# importing packages
import seaborn as sns
import matplotlib.pyplot as plt

# loading dataset
data = traindf3

# draw lineplot
sns.lineplot(x="Store", y="Unemployment", data=data)
plt.show()
```



In [33]:
```
traindf3
```

Out[33]:

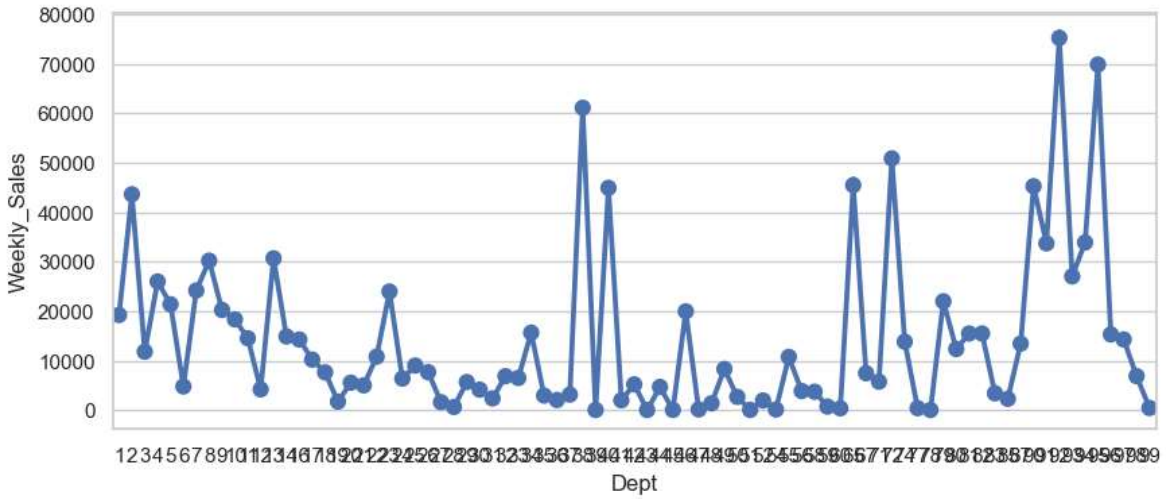
	Store	Dept	Date	Weekly_Sales	IsHoliday	Type	Size	Temperature	Fuel
0	1	1	2010-02-05	24924.50	False	A	151315	42.31	
1	1	2	2010-02-05	50605.27	False	A	151315	42.31	
2	1	3	2010-02-05	13740.12	False	A	151315	42.31	
3	1	4	2010-02-05	39954.04	False	A	151315	42.31	
4	1	5	2010-02-05	32229.38	False	A	151315	42.31	
...	...	...	...	...	...	...	...	...	
420207	45	93	2012-10-26	2487.80	False	B	118221	58.85	
420208	45	94	2012-10-26	5203.31	False	B	118221	58.85	
420209	45	95	2012-10-26	56017.47	False	B	118221	58.85	
420210	45	97	2012-10-26	6817.48	False	B	118221	58.85	
420211	45	98	2012-10-26	1076.80	False	B	118221	58.85	

420212 rows × 12 columns

In [34]:
```
traindf3['Dept'].unique()
```

Out[34]:
```
array([ 1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 16, 17, 18,
       19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35,
       36, 37, 38, 40, 41, 42, 44, 45, 46, 47, 48, 49, 51, 52, 54, 55, 56,
       58, 59, 60, 67, 71, 72, 74, 79, 80, 81, 82, 83, 85, 87, 90, 91, 92,
       93, 94, 95, 97, 98, 78, 96, 99, 77, 39, 50, 43, 65], dtype=int64)
```

In [35]:
```
# importing required packages
import seaborn as sns
import matplotlib.pyplot as plt

# loading dataset
data =traindf3

# draw pointplot
sns.pointplot(x ='Dept',
                        y = "Weekly_Sales",
                        data = data)
# show the plot
sns.set(rc = {'figure.figsize':(25,8)})
plt.show()
```

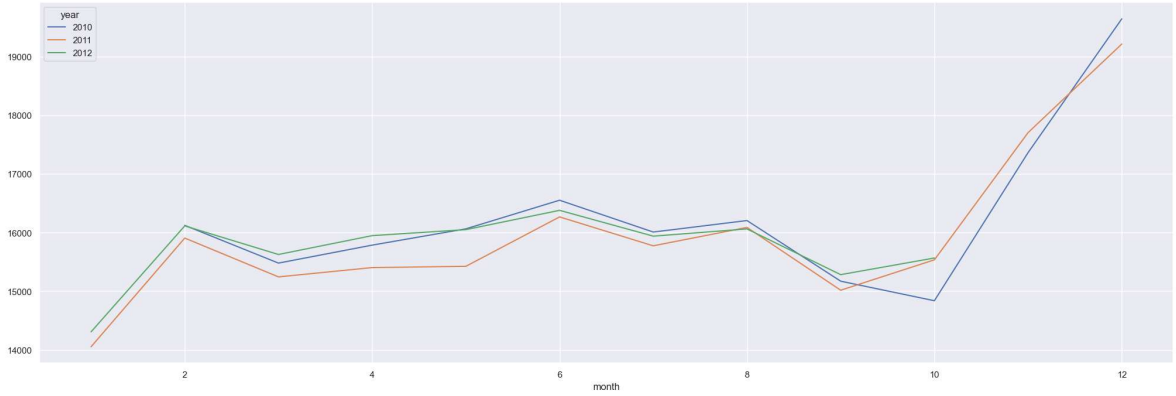


In [36]:
```
traindf3['month'] = pd.DatetimeIndex(traindf3['Date']).month #extract month data
```

In [37]:
```
traindf4=traindf3.drop(['Date'],axis=1)
```

In [38]:
```
month_wise_sales = pd.pivot_table(traindf4, values = "Weekly_Sales", columns = '
month_wise_sales.plot()
```

Out[38]: <Axes: xlabel='month'>



Label encoding for Holiday column and Type

In [39]:
```
# Import label encoder
from sklearn import preprocessing

# label_encoder object knows how to understand word labels.
label_encoder = preprocessing.LabelEncoder()

# Encode labels in column 'species'.
traindf4['IsHoliday']= label_encoder.fit_transform(traindf4['IsHoliday'])
traindf4['Type']= label_encoder.fit_transform(traindf4['Type'])

traindf4
```

Out[39]:

	Store	Dept	Weekly_Sales	IsHoliday	Type	Size	Temperature	Fuel_Price
0	1	1	24924.50	0	0	151315	42.31	2.572
1	1	2	50605.27	0	0	151315	42.31	2.572
2	1	3	13740.12	0	0	151315	42.31	2.572
3	1	4	39954.04	0	0	151315	42.31	2.572
4	1	5	32229.38	0	0	151315	42.31	2.572
...	...	...	...	...	...	...	...	...
420207	45	93	2487.80	0	1	118221	58.85	3.882
420208	45	94	5203.31	0	1	118221	58.85	3.882
420209	45	95	56017.47	0	1	118221	58.85	3.882
420210	45	97	6817.48	0	1	118221	58.85	3.882
420211	45	98	1076.80	0	1	118221	58.85	3.882

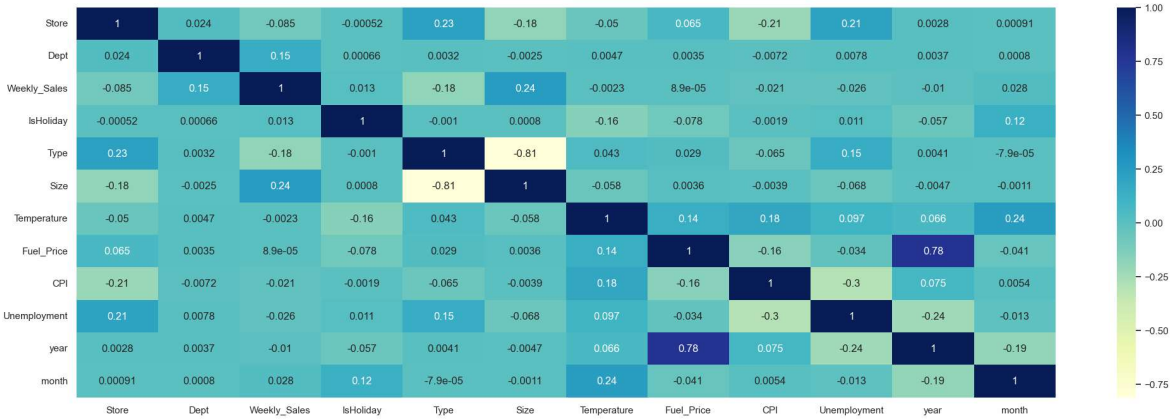
420212 rows × 12 columns

Correlation Map 2

In [40]:

```
data = traindf4

# prints data that will be plotted
# columns shown here are selected by corr() since
# they are ideal for the plot
print(data.corr())
sns.set_theme(style="whitegrid")
# plotting correlation heatmap
dataplot = sns.heatmap(data.corr(), cmap="YlGnBu", annot=True)
sns.set(rc = {'figure.figsize':(25,8)})

# displaying heatmap
mp.show()
```

```
                Store      Dept  Weekly_Sales  IsHoliday      Type      Size  \
Store        1.000000  0.024258     -0.085117  -0.000522  0.226352 -0.182763   
Dept         0.024258  1.000000      0.148749   0.000663  0.003157 -0.002491   
Weekly_Sales -0.085117  0.148749      1.000000   0.012843 -0.182229  0.244117   
IsHoliday    -0.000522  0.000663      0.012843   1.000000 -0.001000  0.000797   
Type          0.226352  0.003157     -0.182229  -0.001000  1.000000 -0.811541   
Size         -0.182763 -0.002491      0.244117   0.000797 -0.811541  1.000000   
Temperature  -0.050230  0.004727     -0.002339  -0.155775  0.043035 -0.058413   
Fuel_Price    0.065321  0.003544      0.000089  -0.078155  0.029483  0.003632   
CPI          -0.211261 -0.007178     -0.021162  -0.001933 -0.065094 -0.003903   
Unemployment  0.208759  0.007787     -0.025806   0.010555  0.148793 -0.068335   
year          0.002831  0.003716     -0.010015  -0.056572  0.004080 -0.004716   
month         0.000907  0.000800      0.028401   0.123058 -0.000079 -0.001051   

              Temperature  Fuel_Price       CPI  Unemployment      year  \
Store           -0.050230    0.065321 -0.211261      0.208759  0.002831   
Dept             0.004727    0.003544 -0.007178      0.007787  0.003716   
Weekly_Sales    -0.002339    0.000089 -0.021162     -0.025806 -0.010015   
IsHoliday       -0.155775   -0.078155 -0.001933      0.010555 -0.056572   
Type             0.043035    0.029483 -0.065094      0.148793  0.004080   
Size            -0.058413    0.003632 -0.003903     -0.068335 -0.004716   
Temperature      1.000000    0.143700  0.182223      0.096768  0.065712   
Fuel_Price       0.143700    1.000000 -0.164199     -0.033915  0.779681   
CPI              0.182223   -0.164199  1.000000     -0.299887  0.074547   
Unemployment     0.096768   -0.033915 -0.299887      1.000000 -0.237210   
year             0.065712    0.779681  0.074547     -0.237210  1.000000   
month            0.235957   -0.040931  0.005366     -0.012562 -0.194295   

                month  
Store        0.000907  
Dept         0.000800  
Weekly_Sales 0.028401  
IsHoliday    0.123058  
Type        -0.000079  
Size        -0.001051  
Temperature  0.235957  
Fuel_Price  -0.040931  
CPI          0.005366  
Unemployment -0.012562  
year        -0.194295  
month        1.000000  
```



In [41]:
```
pip install numba
```

```
Collecting numba
  Using cached numba-0.56.4.tar.gz (2.4 MB)
  Installing build dependencies: started
  Installing build dependencies: finished with status 'done'
  Getting requirements to build wheel: started
  Getting requirements to build wheel: finished with status 'error'
Note: you may need to restart the kernel to use updated packages.
  error: subprocess-exited-with-error

  Getting requirements to build wheel did not run successfully.
  exit code: 1

  [22 lines of output]
  Traceback (most recent call last):
    File "C:\Users\DELL\AppData\Local\Programs\Python\Python311\Lib\site-packages
\pip\_vendor\pyproject_hooks\_in_process\_in_process.py", line 353, in <module>
      main()
    File "C:\Users\DELL\AppData\Local\Programs\Python\Python311\Lib\site-packages
\pip\_vendor\pyproject_hooks\_in_process\_in_process.py", line 335, in main
      json_out['return_val'] = hook(**hook_input['kwargs'])
                               ^^^^^^^^^^^^^^^^^^^^^^^^^^^^
    File "C:\Users\DELL\AppData\Local\Programs\Python\Python311\Lib\site-packages
\pip\_vendor\pyproject_hooks\_in_process\_in_process.py", line 118, in get_requir
es_for_build_wheel
      return hook(config_settings)
             ^^^^^^^^^^^^^^^^^^^^^
    File "C:\Users\DELL\AppData\Local\Temp\pip-build-env-za9mgqa3\overlay\Lib\sit
e-packages\setuptools\build_meta.py", line 341, in get_requires_for_build_wheel
      return self._get_build_requires(config_settings, requirements=['wheel'])
             ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
    File "C:\Users\DELL\AppData\Local\Temp\pip-build-env-za9mgqa3\overlay\Lib\sit
e-packages\setuptools\build_meta.py", line 323, in _get_build_requires
      self.run_setup()
    File "C:\Users\DELL\AppData\Local\Temp\pip-build-env-za9mgqa3\overlay\Lib\sit
e-packages\setuptools\build_meta.py", line 488, in run_setup
      self).run_setup(setup_script=setup_script)
            ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
    File "C:\Users\DELL\AppData\Local\Temp\pip-build-env-za9mgqa3\overlay\Lib\sit
e-packages\setuptools\build_meta.py", line 338, in run_setup
      exec(code, locals())
    File "<string>", line 51, in <module>
    File "<string>", line 48, in _guard_py_ver
  RuntimeError: Cannot install on Python version 3.11.3; only versions >=3.7,<3.1
1 are supported.
  [end of output]

  note: This error originates from a subprocess, and is likely not a problem with
pip.
error: subprocess-exited-with-error

Getting requirements to build wheel did not run successfully.
exit code: 1

See above for output.

note: This error originates from a subprocess, and is likely not a problem with p
ip.
```

In [42]:
```
pip install shap
```

```
Note: you may need to restart the kernel to use updated packages.Collecting shap
  Using cached shap-0.41.0.tar.gz (380 kB)
  Installing build dependencies: started
  Installing build dependencies: finished with status 'done'
  Getting requirements to build wheel: started
  Getting requirements to build wheel: finished with status 'done'
  Preparing metadata (pyproject.toml): started
  Preparing metadata (pyproject.toml): finished with status 'done'
Requirement already satisfied: numpy in c:\users\dell\appdata\local\programs\pyth
on\python311\lib\site-packages (from shap) (1.24.2)
Requirement already satisfied: scipy in c:\users\dell\appdata\local\programs\pyth
on\python311\lib\site-packages (from shap) (1.10.1)
Requirement already satisfied: scikit-learn in c:\users\dell\appdata\local\progra
ms\python\python311\lib\site-packages (from shap) (1.2.2)
Requirement already satisfied: pandas in c:\users\dell\appdata\local\programs\pyt
hon\python311\lib\site-packages (from shap) (2.0.0)
Requirement already satisfied: tqdm>4.25.0 in c:\users\dell\appdata\local\program
s\python\python311\lib\site-packages (from shap) (4.65.0)
Requirement already satisfied: packaging>20.9 in c:\users\dell\appdata\local\prog
rams\python\python311\lib\site-packages (from shap) (23.1)
Collecting slicer==0.0.7 (from shap)
  Using cached slicer-0.0.7-py3-none-any.whl (14 kB)
Collecting numba (from shap)
  Using cached numba-0.56.4.tar.gz (2.4 MB)
  Installing build dependencies: started
  Installing build dependencies: finished with status 'done'
  Getting requirements to build wheel: started
  Getting requirements to build wheel: finished with status 'error'
```

```
  error: subprocess-exited-with-error

  Getting requirements to build wheel did not run successfully.
  exit code: 1

  [22 lines of output]
  Traceback (most recent call last):
    File "C:\Users\DELL\AppData\Local\Programs\Python\Python311\Lib\site-packages
\pip\_vendor\pyproject_hooks\_in_process\_in_process.py", line 353, in <module>
      main()
    File "C:\Users\DELL\AppData\Local\Programs\Python\Python311\Lib\site-packages
\pip\_vendor\pyproject_hooks\_in_process\_in_process.py", line 335, in main
      json_out['return_val'] = hook(**hook_input['kwargs'])
                               ^^^^^^^^^^^^^^^^^^^^^^^^^^^^
    File "C:\Users\DELL\AppData\Local\Programs\Python\Python311\Lib\site-packages
\pip\_vendor\pyproject_hooks\_in_process\_in_process.py", line 118, in get_requir
es_for_build_wheel
      return hook(config_settings)
             ^^^^^^^^^^^^^^^^^^^^^
    File "C:\Users\DELL\AppData\Local\Temp\pip-build-env-bshsqihp\overlay\Lib\sit
e-packages\setuptools\build_meta.py", line 341, in get_requires_for_build_wheel
      return self._get_build_requires(config_settings, requirements=['wheel'])
             ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
    File "C:\Users\DELL\AppData\Local\Temp\pip-build-env-bshsqihp\overlay\Lib\sit
e-packages\setuptools\build_meta.py", line 323, in _get_build_requires
      self.run_setup()
    File "C:\Users\DELL\AppData\Local\Temp\pip-build-env-bshsqihp\overlay\Lib\sit
e-packages\setuptools\build_meta.py", line 488, in run_setup
      self).run_setup(setup_script=setup_script)
            ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
    File "C:\Users\DELL\AppData\Local\Temp\pip-build-env-bshsqihp\overlay\Lib\sit
e-packages\setuptools\build_meta.py", line 338, in run_setup
      exec(code, locals())
    File "<string>", line 51, in <module>
    File "<string>", line 48, in _guard_py_ver
  RuntimeError: Cannot install on Python version 3.11.3; only versions >=3.7,<3.1
1 are supported.
  [end of output]

  note: This error originates from a subprocess, and is likely not a problem with
pip.
error: subprocess-exited-with-error

Getting requirements to build wheel did not run successfully.
exit code: 1

See above for output.

note: This error originates from a subprocess, and is likely not a problem with p
ip.
```

Feature Importance Test using various techniques

In [45]:
```
Features=traindf4.drop(['Weekly_Sales'],axis=1)
Target=traindf4['Weekly_Sales']
```

In [46]:
```
rf = RandomForestRegressor(n_estimators=100)
rf.fit(Features,Target)
```

Out[46]:
```
RandomForestRegressor
RandomForestRegressor()
```

In [47]:
```
Features
```

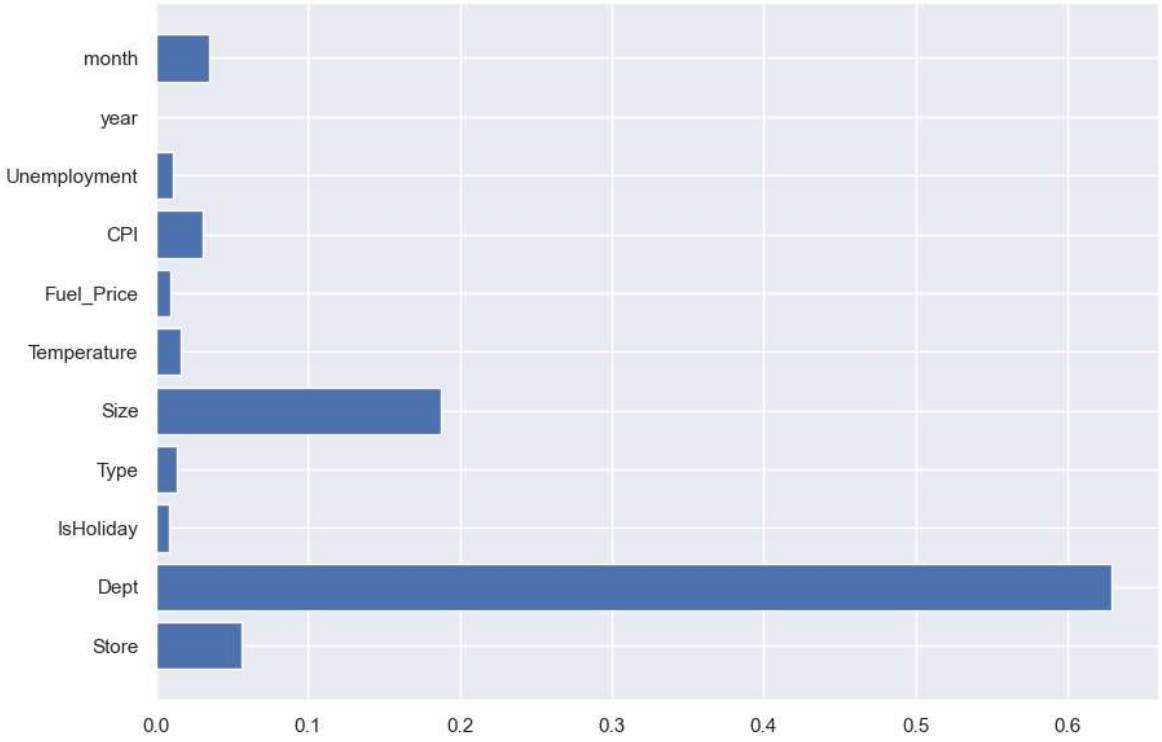
Out[47]:

	Store	Dept	IsHoliday	Type	Size	Temperature	Fuel_Price	CPI	Ur
0	1	1	0	0	151315	42.31	2.572	211.096358	
1	1	2	0	0	151315	42.31	2.572	211.096358	
2	1	3	0	0	151315	42.31	2.572	211.096358	
3	1	4	0	0	151315	42.31	2.572	211.096358	
4	1	5	0	0	151315	42.31	2.572	211.096358	
...	...	...	...	...	...	...	...	...	
420207	45	93	0	1	118221	58.85	3.882	192.308899	
420208	45	94	0	1	118221	58.85	3.882	192.308899	
420209	45	95	0	1	118221	58.85	3.882	192.308899	
420210	45	97	0	1	118221	58.85	3.882	192.308899	
420211	45	98	0	1	118221	58.85	3.882	192.308899	

420212 rows × 11 columns

In [48]:
```
f = plt.figure()
f.set_figwidth(10)
f.set_figheight(7)
plt.barh(Features.columns, rf.feature_importances_)
```

Out[48]:
```
<BarContainer object of 11 artists>
```



In [49]:
```
F=Features.drop(["IsHoliday",'year'],axis=1)
```

In [50]:
```
F
```

Out[50]:

	Store	Dept	Type	Size	Temperature	Fuel_Price	CPI	Unemployme
0	1	1	0	151315	42.31	2.572	211.096358	8.1
1	1	2	0	151315	42.31	2.572	211.096358	8.1
2	1	3	0	151315	42.31	2.572	211.096358	8.1
3	1	4	0	151315	42.31	2.572	211.096358	8.1
4	1	5	0	151315	42.31	2.572	211.096358	8.1
...	...	...	...	...	...	...	...	
420207	45	93	1	118221	58.85	3.882	192.308899	8.6
420208	45	94	1	118221	58.85	3.882	192.308899	8.6
420209	45	95	1	118221	58.85	3.882	192.308899	8.6
420210	45	97	1	118221	58.85	3.882	192.308899	8.6
420211	45	98	1	118221	58.85	3.882	192.308899	8.6

420212 rows × 9 columns

In [51]:
```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test= train_test_split(F, Target, test_size= 0.25, r
```

In [52]:
```
from sklearn.ensemble import RandomForestRegressor
from sklearn.tree import DecisionTreeRegressor
```

```
from sklearn.metrics import r2_score,mean_squared_error
from math import sqrt
```

In [53]:
```
DTRmodel = DecisionTreeRegressor(max_depth=3,random_state=0)
DTRmodel.fit(x_train,y_train)
y_pred = DTRmodel.predict(x_test)
```

In [54]:
```
print("R2 score  :",r2_score(y_test, y_pred))
print("MSE score  :",mean_squared_error(y_test, y_pred))
print("RMSE: ",sqrt(mean_squared_error(y_test, y_pred)))
```

```
R2 score  : 0.3749492429348211
MSE score  : 326619231.0073728
RMSE:  18072.609966669806
```

In [55]:
```
rf1 = RandomForestRegressor(n_estimators=50, random_state=42, n_jobs=-1, max_dep
                            max_features = 'sqrt',min_samples_split = 10)
rf1.fit(x_train,y_train)
y_pred1 = rf1.predict(x_test)
```

In [56]:
```
print("R2 score  :",r2_score(y_test, y_pred))
print("MSE score  :",mean_squared_error(y_test, y_pred1))
print("RMSE: ",sqrt(mean_squared_error(y_test, y_pred1)))
```

```
R2 score  : 0.3749492429348211
MSE score  : 55770391.91594352
RMSE:  7467.957680379792
```

In [58]:
```
pip install xgboost
```

```
Collecting xgboostNote: you may need to restart the kernel to use updated package
s.

  Downloading xgboost-1.7.5-py3-none-win_amd64.whl (70.9 MB)
                                              0.0/70.9 MB ? eta -:--:--
                                             0.0/70.9 MB 653.6 kB/s eta 0:01:49
                                              0.2/70.9 MB 1.5 MB/s eta 0:00:47
                                              0.3/70.9 MB 1.8 MB/s eta 0:00:40
                                              0.3/70.9 MB 1.5 MB/s eta 0:00:47
                                              0.5/70.9 MB 2.1 MB/s eta 0:00:34
                                              0.8/70.9 MB 2.4 MB/s eta 0:00:29
                                              1.0/70.9 MB 2.8 MB/s eta 0:00:25
                                              1.4/70.9 MB 3.3 MB/s eta 0:00:22
                                              1.6/70.9 MB 3.6 MB/s eta 0:00:20
     -                                        1.9/70.9 MB 3.7 MB/s eta 0:00:19
     -                                        2.2/70.9 MB 4.0 MB/s eta 0:00:18
     -                                        2.5/70.9 MB 4.0 MB/s eta 0:00:18
     -                                        2.8/70.9 MB 4.2 MB/s eta 0:00:17
     -                                        3.1/70.9 MB 4.4 MB/s eta 0:00:16
     -                                        3.1/70.9 MB 4.5 MB/s eta 0:00:16
     --                                       3.6/70.9 MB 4.5 MB/s eta 0:00:15
     --                                       3.8/70.9 MB 4.6 MB/s eta 0:00:15
     --                                       4.1/70.9 MB 4.6 MB/s eta 0:00:15
     --                                       4.5/70.9 MB 4.8 MB/s eta 0:00:14
     --                                       4.9/70.9 MB 4.9 MB/s eta 0:00:14
     --                                       5.2/70.9 MB 5.0 MB/s eta 0:00:14
     ---                                      5.4/70.9 MB 5.0 MB/s eta 0:00:13
     ---                                      5.9/70.9 MB 5.0 MB/s eta 0:00:14
     ---                                      6.1/70.9 MB 4.9 MB/s eta 0:00:14
     ---                                      6.6/70.9 MB 5.1 MB/s eta 0:00:13
     ---                                      6.7/70.9 MB 5.1 MB/s eta 0:00:13
     ----                                     7.2/70.9 MB 5.2 MB/s eta 0:00:13
     ----                                     7.3/70.9 MB 5.1 MB/s eta 0:00:13
     ----                                     7.8/70.9 MB 5.3 MB/s eta 0:00:12
     ----                                     8.0/70.9 MB 5.2 MB/s eta 0:00:13
     ----                                     8.4/70.9 MB 5.3 MB/s eta 0:00:12
     ----                                     8.6/70.9 MB 5.2 MB/s eta 0:00:12
     -----                                    9.1/70.9 MB 5.3 MB/s eta 0:00:12
     -----                                    9.5/70.9 MB 5.4 MB/s eta 0:00:12
     -----                                    9.8/70.9 MB 5.4 MB/s eta 0:00:12
     -----                                    10.4/70.9 MB 5.7 MB/s eta 0:00:11
     -----                                    10.5/70.9 MB 5.8 MB/s eta 0:00:11
     ------                                   10.8/70.9 MB 5.9 MB/s eta 0:00:11
     ------                                   11.1/70.9 MB 6.1 MB/s eta 0:00:10
     ------                                   11.2/70.9 MB 6.1 MB/s eta 0:00:10
     ------                                   11.4/70.9 MB 6.0 MB/s eta 0:00:10
     ------                                   11.9/70.9 MB 6.1 MB/s eta 0:00:10
     ------                                   12.2/70.9 MB 6.0 MB/s eta 0:00:10
     -------                                  12.7/70.9 MB 6.2 MB/s eta 0:00:10
     -------                                  12.9/70.9 MB 6.0 MB/s eta 0:00:10
     -------                                  13.1/70.9 MB 6.1 MB/s eta 0:00:10
     -------                                  13.6/70.9 MB 6.1 MB/s eta 0:00:10
     -------                                  13.8/70.9 MB 6.0 MB/s eta 0:00:10
     -------                                  14.2/70.9 MB 6.1 MB/s eta 0:00:10
     --------                                 14.5/70.9 MB 6.1 MB/s eta 0:00:10
     --------                                 14.8/70.9 MB 6.0 MB/s eta 0:00:10
     --------                                 15.3/70.9 MB 6.0 MB/s eta 0:00:10
     --------                                 15.5/70.9 MB 6.0 MB/s eta 0:00:10
     --------                                 15.7/70.9 MB 6.0 MB/s eta 0:00:10
     ---------                                16.1/70.9 MB 6.0 MB/s eta 0:00:10
```

```
--------- 16.3/70.9 MB 6.1 MB/s eta 0:00:09
--------- 16.5/70.9 MB 6.1 MB/s eta 0:00:09
--------- 16.7/70.9 MB 6.0 MB/s eta 0:00:10
--------- 17.2/70.9 MB 6.1 MB/s eta 0:00:09
--------- 17.5/70.9 MB 6.1 MB/s eta 0:00:09
--------- 17.7/70.9 MB 6.0 MB/s eta 0:00:09
---------- 18.2/70.9 MB 6.1 MB/s eta 0:00:09
---------- 18.5/70.9 MB 6.1 MB/s eta 0:00:09
---------- 18.7/70.9 MB 6.1 MB/s eta 0:00:09
---------- 19.0/70.9 MB 6.0 MB/s eta 0:00:09
---------- 19.5/70.9 MB 6.1 MB/s eta 0:00:09
----------- 19.8/70.9 MB 6.1 MB/s eta 0:00:09
----------- 20.0/70.9 MB 5.9 MB/s eta 0:00:09
----------- 20.4/70.9 MB 6.0 MB/s eta 0:00:09
----------- 20.9/70.9 MB 6.1 MB/s eta 0:00:09
----------- 21.0/70.9 MB 6.1 MB/s eta 0:00:09
----------- 21.2/70.9 MB 5.9 MB/s eta 0:00:09
------------ 21.5/70.9 MB 6.1 MB/s eta 0:00:09
------------ 21.9/70.9 MB 6.1 MB/s eta 0:00:09
------------ 22.2/70.9 MB 6.1 MB/s eta 0:00:09
------------ 22.3/70.9 MB 6.2 MB/s eta 0:00:08
------------ 22.5/70.9 MB 6.0 MB/s eta 0:00:09
------------ 23.0/70.9 MB 6.1 MB/s eta 0:00:08
------------- 23.4/70.9 MB 6.2 MB/s eta 0:00:08
------------- 23.8/70.9 MB 6.1 MB/s eta 0:00:08
------------- 24.0/70.9 MB 6.2 MB/s eta 0:00:08
------------- 24.4/70.9 MB 6.1 MB/s eta 0:00:08
------------- 24.7/70.9 MB 6.1 MB/s eta 0:00:08
-------------- 25.1/70.9 MB 6.3 MB/s eta 0:00:08
-------------- 25.2/70.9 MB 6.2 MB/s eta 0:00:08
-------------- 25.5/70.9 MB 6.1 MB/s eta 0:00:08
-------------- 25.9/70.9 MB 6.2 MB/s eta 0:00:08
-------------- 26.3/70.9 MB 6.2 MB/s eta 0:00:08
--------------- 26.6/70.9 MB 6.2 MB/s eta 0:00:08
--------------- 26.8/70.9 MB 6.2 MB/s eta 0:00:08
--------------- 27.2/70.9 MB 6.2 MB/s eta 0:00:08
--------------- 27.7/70.9 MB 6.2 MB/s eta 0:00:07
--------------- 28.1/70.9 MB 6.3 MB/s eta 0:00:07
--------------- 28.3/70.9 MB 6.2 MB/s eta 0:00:07
---------------- 28.7/70.9 MB 6.2 MB/s eta 0:00:07
---------------- 29.2/70.9 MB 6.4 MB/s eta 0:00:07
---------------- 29.6/70.9 MB 6.3 MB/s eta 0:00:07
---------------- 29.7/70.9 MB 6.2 MB/s eta 0:00:07
---------------- 29.9/70.9 MB 6.2 MB/s eta 0:00:07
----------------- 30.3/70.9 MB 6.3 MB/s eta 0:00:07
----------------- 30.6/70.9 MB 6.4 MB/s eta 0:00:07
----------------- 30.9/70.9 MB 6.4 MB/s eta 0:00:07
----------------- 31.2/70.9 MB 6.4 MB/s eta 0:00:07
----------------- 31.3/70.9 MB 6.5 MB/s eta 0:00:07
----------------- 31.5/70.9 MB 6.3 MB/s eta 0:00:07
----------------- 31.8/70.9 MB 6.3 MB/s eta 0:00:07
------------------ 32.1/70.9 MB 6.3 MB/s eta 0:00:07
------------------ 32.4/70.9 MB 6.3 MB/s eta 0:00:07
------------------ 32.6/70.9 MB 6.5 MB/s eta 0:00:06
------------------ 32.6/70.9 MB 6.5 MB/s eta 0:00:06
------------------ 32.6/70.9 MB 6.5 MB/s eta 0:00:06
------------------ 32.6/70.9 MB 6.5 MB/s eta 0:00:06
------------------ 32.6/70.9 MB 6.5 MB/s eta 0:00:06
------------------ 32.6/70.9 MB 6.5 MB/s eta 0:00:06
------------------ 32.6/70.9 MB 6.5 MB/s eta 0:00:06
```

```
                         ------------------                       32.6/70.9 MB 6.5 MB/s eta 0:00:06
                         ------------------                       32.6/70.9 MB 6.5 MB/s eta 0:00:06
                         ------------------                       32.6/70.9 MB 6.5 MB/s eta 0:00:06
                         ------------------                       32.6/70.9 MB 6.5 MB/s eta 0:00:06
                         ------------------                       32.6/70.9 MB 6.5 MB/s eta 0:00:06
                         ------------------                       32.6/70.9 MB 6.5 MB/s eta 0:00:06
                         ------------------                       32.6/70.9 MB 6.5 MB/s eta 0:00:06
                         ------------------                       32.6/70.9 MB 6.5 MB/s eta 0:00:06
                         ------------------                       32.8/70.9 MB 4.6 MB/s eta 0:00:09
                         ------------------                       32.8/70.9 MB 4.6 MB/s eta 0:00:09
                         ------------------                       32.8/70.9 MB 4.6 MB/s eta 0:00:09
                         ------------------                       32.8/70.9 MB 4.6 MB/s eta 0:00:09
                         ------------------                       32.8/70.9 MB 4.6 MB/s eta 0:00:09
                         ------------------                       32.8/70.9 MB 4.6 MB/s eta 0:00:09
                         ------------------                       32.8/70.9 MB 4.6 MB/s eta 0:00:09
                         ------------------                       32.8/70.9 MB 4.6 MB/s eta 0:00:09
                         ------------------                       32.8/70.9 MB 4.6 MB/s eta 0:00:09
                         ------------------                       32.9/70.9 MB 4.0 MB/s eta 0:00:10
                         ------------------                       32.9/70.9 MB 4.0 MB/s eta 0:00:10
                         ------------------                       32.9/70.9 MB 4.0 MB/s eta 0:00:10
                         ------------------                       32.9/70.9 MB 4.0 MB/s eta 0:00:10
                         ------------------                       32.9/70.9 MB 4.0 MB/s eta 0:00:10
                         ------------------                       32.9/70.9 MB 4.0 MB/s eta 0:00:10
                         ------------------                       32.9/70.9 MB 4.0 MB/s eta 0:00:10
                         ------------------                       33.0/70.9 MB 3.6 MB/s eta 0:00:11
                         ------------------                       33.0/70.9 MB 3.6 MB/s eta 0:00:11
                         ------------------                       33.0/70.9 MB 3.6 MB/s eta 0:00:11
                         ------------------                       33.0/70.9 MB 3.6 MB/s eta 0:00:11
                         ------------------                       33.0/70.9 MB 3.6 MB/s eta 0:00:11
                         ------------------                       33.0/70.9 MB 3.6 MB/s eta 0:00:11
                         ------------------                       33.0/70.9 MB 3.6 MB/s eta 0:00:11
                         ------------------                       33.0/70.9 MB 3.2 MB/s eta 0:00:12
                         ------------------                       33.0/70.9 MB 3.2 MB/s eta 0:00:12
                         ------------------                       33.0/70.9 MB 3.2 MB/s eta 0:00:12
                         ------------------                       33.0/70.9 MB 3.2 MB/s eta 0:00:12
                         ------------------                       33.0/70.9 MB 3.2 MB/s eta 0:00:12
                         ------------------                       33.0/70.9 MB 3.2 MB/s eta 0:00:12
                         ------------------                       33.0/70.9 MB 3.2 MB/s eta 0:00:12
                         ------------------                       33.1/70.9 MB 2.9 MB/s eta 0:00:13
                         ------------------                       33.1/70.9 MB 2.9 MB/s eta 0:00:13
                         ------------------                       33.1/70.9 MB 2.9 MB/s eta 0:00:13
                         ------------------                       33.1/70.9 MB 2.9 MB/s eta 0:00:13
                         ------------------                       33.1/70.9 MB 2.9 MB/s eta 0:00:13
                         ------------------                       33.1/70.9 MB 2.9 MB/s eta 0:00:13
                         ------------------                       33.1/70.9 MB 2.9 MB/s eta 0:00:13
                         ------------------                       33.1/70.9 MB 2.9 MB/s eta 0:00:13
                         ------------------                       33.1/70.9 MB 2.9 MB/s eta 0:00:13
                         ------------------                       33.1/70.9 MB 2.9 MB/s eta 0:00:13
                         ------------------                       33.1/70.9 MB 2.9 MB/s eta 0:00:13
                         ------------------                       33.1/70.9 MB 2.9 MB/s eta 0:00:13
                         ------------------                       33.1/70.9 MB 2.9 MB/s eta 0:00:13
                         ------------------                       33.1/70.9 MB 2.9 MB/s eta 0:00:13
                         ------------------                       33.1/70.9 MB 2.9 MB/s eta 0:00:13
                         ------------------                       33.2/70.9 MB 2.5 MB/s eta 0:00:16
                         ------------------                       33.2/70.9 MB 2.5 MB/s eta 0:00:16
                         ------------------                       33.2/70.9 MB 2.5 MB/s eta 0:00:16
                         ------------------                       33.2/70.9 MB 2.5 MB/s eta 0:00:16
                         ------------------                       33.2/70.9 MB 2.5 MB/s eta 0:00:16
                         ------------------                       33.2/70.9 MB 2.5 MB/s eta 0:00:16
                         ------------------                       33.2/70.9 MB 2.5 MB/s eta 0:00:16
```

```
     ------------------                       33.3/70.9 MB 2.3 MB/s eta 0:00:17
     ------------------                       33.3/70.9 MB 2.3 MB/s eta 0:00:17
     ------------------                       33.3/70.9 MB 2.3 MB/s eta 0:00:17
     ------------------                       33.3/70.9 MB 2.3 MB/s eta 0:00:17
     ------------------                       33.3/70.9 MB 2.3 MB/s eta 0:00:17
     ------------------                       33.3/70.9 MB 2.3 MB/s eta 0:00:17
     ------------------                       33.3/70.9 MB 2.3 MB/s eta 0:00:17
     ------------------                       33.4/70.9 MB 2.2 MB/s eta 0:00:18
     ------------------                       33.5/70.9 MB 2.2 MB/s eta 0:00:18
     ------------------                       33.5/70.9 MB 2.2 MB/s eta 0:00:18
     ------------------                       33.6/70.9 MB 2.2 MB/s eta 0:00:18
     -------------------                      33.7/70.9 MB 2.1 MB/s eta 0:00:18
     -------------------                      34.0/70.9 MB 2.1 MB/s eta 0:00:18
     -------------------                      34.1/70.9 MB 2.1 MB/s eta 0:00:18
     -------------------                      34.4/70.9 MB 2.1 MB/s eta 0:00:18
     -------------------                      34.5/70.9 MB 2.1 MB/s eta 0:00:18
     -------------------                      34.6/70.9 MB 2.1 MB/s eta 0:00:18
     -------------------                      35.0/70.9 MB 2.1 MB/s eta 0:00:18
     -------------------                      35.3/70.9 MB 2.1 MB/s eta 0:00:17
     --------------------                     35.6/70.9 MB 2.1 MB/s eta 0:00:17
     --------------------                     35.9/70.9 MB 2.1 MB/s eta 0:00:17
     --------------------                     36.0/70.9 MB 2.1 MB/s eta 0:00:17
     --------------------                     36.2/70.9 MB 2.1 MB/s eta 0:00:17
     --------------------                     36.7/70.9 MB 2.1 MB/s eta 0:00:17
     --------------------                     36.8/70.9 MB 2.1 MB/s eta 0:00:17
     --------------------                     37.1/70.9 MB 2.1 MB/s eta 0:00:16
     ---------------------                    37.6/70.9 MB 2.1 MB/s eta 0:00:16
     ---------------------                    37.7/70.9 MB 2.1 MB/s eta 0:00:16
     ---------------------                    38.1/70.9 MB 2.1 MB/s eta 0:00:16
     ---------------------                    38.6/70.9 MB 2.1 MB/s eta 0:00:16
     ---------------------                    38.7/70.9 MB 2.1 MB/s eta 0:00:16
     ----------------------                   39.1/70.9 MB 2.1 MB/s eta 0:00:16
     ----------------------                   39.5/70.9 MB 2.1 MB/s eta 0:00:15
     ----------------------                   39.6/70.9 MB 2.1 MB/s eta 0:00:15
     ----------------------                   40.0/70.9 MB 2.1 MB/s eta 0:00:15
     ----------------------                   40.3/70.9 MB 2.1 MB/s eta 0:00:15
     ----------------------                   40.4/70.9 MB 2.1 MB/s eta 0:00:15
     ----------------------                   40.7/70.9 MB 2.1 MB/s eta 0:00:15
     -----------------------                  41.1/70.9 MB 2.1 MB/s eta 0:00:15
     -----------------------                  41.1/70.9 MB 2.0 MB/s eta 0:00:15
     -----------------------                  41.4/70.9 MB 2.0 MB/s eta 0:00:15
     -----------------------                  41.7/70.9 MB 2.1 MB/s eta 0:00:15
     -----------------------                  42.0/70.9 MB 2.1 MB/s eta 0:00:15
     -----------------------                  42.1/70.9 MB 2.0 MB/s eta 0:00:15
     -----------------------                  42.5/70.9 MB 2.0 MB/s eta 0:00:14
     ------------------------                 42.8/70.9 MB 2.0 MB/s eta 0:00:14
     ------------------------                 42.9/70.9 MB 2.3 MB/s eta 0:00:13
     ------------------------                 43.1/70.9 MB 2.5 MB/s eta 0:00:12
     ------------------------                 43.6/70.9 MB 5.0 MB/s eta 0:00:06
     ------------------------                 43.8/70.9 MB 5.1 MB/s eta 0:00:06
     ------------------------                 44.2/70.9 MB 5.4 MB/s eta 0:00:05
     -------------------------                44.4/70.9 MB 5.4 MB/s eta 0:00:05
     -------------------------                44.7/70.9 MB 5.4 MB/s eta 0:00:05
     -------------------------                45.1/70.9 MB 5.5 MB/s eta 0:00:05
     -------------------------                45.4/70.9 MB 5.5 MB/s eta 0:00:05
     -------------------------                45.6/70.9 MB 5.4 MB/s eta 0:00:05
     -------------------------                45.9/70.9 MB 5.4 MB/s eta 0:00:05
     --------------------------               46.2/70.9 MB 5.4 MB/s eta 0:00:05
     --------------------------               46.3/70.9 MB 5.4 MB/s eta 0:00:05
     --------------------------               46.6/70.9 MB 5.4 MB/s eta 0:00:05
```

```
------------------------              46.9/70.9 MB 5.4 MB/s eta 0:00:05
------------------------              47.2/70.9 MB 5.5 MB/s eta 0:00:05
------------------------              47.3/70.9 MB 5.5 MB/s eta 0:00:05
------------------------              47.7/70.9 MB 5.4 MB/s eta 0:00:05
-------------------------             48.1/70.9 MB 5.5 MB/s eta 0:00:05
-------------------------             48.2/70.9 MB 5.3 MB/s eta 0:00:05
-------------------------             48.6/70.9 MB 5.3 MB/s eta 0:00:05
-------------------------             48.8/70.9 MB 5.4 MB/s eta 0:00:05
-------------------------             49.0/70.9 MB 5.4 MB/s eta 0:00:05
-------------------------             49.2/70.9 MB 5.3 MB/s eta 0:00:05
-------------------------             49.6/70.9 MB 5.3 MB/s eta 0:00:05
--------------------------            49.9/70.9 MB 5.4 MB/s eta 0:00:04
--------------------------            50.1/70.9 MB 5.3 MB/s eta 0:00:04
--------------------------            50.5/70.9 MB 5.4 MB/s eta 0:00:04
--------------------------            50.8/70.9 MB 5.5 MB/s eta 0:00:04
--------------------------            50.8/70.9 MB 5.5 MB/s eta 0:00:04
--------------------------            51.4/70.9 MB 5.5 MB/s eta 0:00:04
---------------------------           51.8/70.9 MB 5.7 MB/s eta 0:00:04
---------------------------           51.9/70.9 MB 5.6 MB/s eta 0:00:04
---------------------------           52.2/70.9 MB 5.7 MB/s eta 0:00:04
---------------------------           52.6/70.9 MB 5.8 MB/s eta 0:00:04
---------------------------           52.8/70.9 MB 5.8 MB/s eta 0:00:04
---------------------------           52.9/70.9 MB 5.7 MB/s eta 0:00:04
----------------------------          53.4/70.9 MB 6.0 MB/s eta 0:00:03
----------------------------          53.7/70.9 MB 5.9 MB/s eta 0:00:03
----------------------------          53.8/70.9 MB 5.7 MB/s eta 0:00:03
----------------------------          54.2/70.9 MB 5.9 MB/s eta 0:00:03
----------------------------          54.5/70.9 MB 6.0 MB/s eta 0:00:03
----------------------------          54.8/70.9 MB 5.8 MB/s eta 0:00:03
-----------------------------         55.1/70.9 MB 5.9 MB/s eta 0:00:03
-----------------------------         55.4/70.9 MB 6.0 MB/s eta 0:00:03
-----------------------------         55.5/70.9 MB 5.9 MB/s eta 0:00:03
-----------------------------         55.7/70.9 MB 6.0 MB/s eta 0:00:03
-----------------------------         56.1/70.9 MB 6.0 MB/s eta 0:00:03
-----------------------------         56.4/70.9 MB 6.0 MB/s eta 0:00:03
-----------------------------         56.7/70.9 MB 6.1 MB/s eta 0:00:03
------------------------------        56.9/70.9 MB 6.0 MB/s eta 0:00:03
------------------------------        57.3/70.9 MB 6.0 MB/s eta 0:00:03
------------------------------        57.7/70.9 MB 6.1 MB/s eta 0:00:03
------------------------------        57.9/70.9 MB 6.0 MB/s eta 0:00:03
------------------------------        58.4/70.9 MB 6.1 MB/s eta 0:00:03
-------------------------------       58.6/70.9 MB 6.1 MB/s eta 0:00:03
-------------------------------       58.7/70.9 MB 6.1 MB/s eta 0:00:03
-------------------------------       59.0/70.9 MB 6.0 MB/s eta 0:00:02
-------------------------------       59.5/70.9 MB 6.2 MB/s eta 0:00:02
-------------------------------       59.6/70.9 MB 6.1 MB/s eta 0:00:02
-------------------------------       60.0/70.9 MB 6.1 MB/s eta 0:00:02
--------------------------------      60.4/70.9 MB 6.2 MB/s eta 0:00:02
--------------------------------      60.5/70.9 MB 6.1 MB/s eta 0:00:02
--------------------------------      61.0/70.9 MB 6.1 MB/s eta 0:00:02
--------------------------------      61.3/70.9 MB 6.2 MB/s eta 0:00:02
--------------------------------      61.5/70.9 MB 6.0 MB/s eta 0:00:02
--------------------------------      62.0/70.9 MB 6.0 MB/s eta 0:00:02
---------------------------------     62.3/70.9 MB 6.1 MB/s eta 0:00:02
---------------------------------     62.8/70.9 MB 6.1 MB/s eta 0:00:02
---------------------------------     63.2/70.9 MB 6.2 MB/s eta 0:00:02
---------------------------------     63.4/70.9 MB 6.0 MB/s eta 0:00:02
---------------------------------     63.7/70.9 MB 6.0 MB/s eta 0:00:02
----------------------------------    64.2/70.9 MB 6.2 MB/s eta 0:00:02
----------------------------------    64.2/70.9 MB 6.2 MB/s eta 0:00:02
```

```
   ------------------------------------    64.4/70.9 MB 6.1 MB/s eta 0:00:02
   ------------------------------------    64.9/70.9 MB 6.1 MB/s eta 0:00:01
   ------------------------------------    65.3/70.9 MB 6.1 MB/s eta 0:00:01
   -------------------------------------   65.7/70.9 MB 6.2 MB/s eta 0:00:01
   -------------------------------------   65.8/70.9 MB 6.1 MB/s eta 0:00:01
   -------------------------------------   66.3/70.9 MB 6.1 MB/s eta 0:00:01
   -------------------------------------   66.6/70.9 MB 6.1 MB/s eta 0:00:01
   -------------------------------------   66.9/70.9 MB 6.0 MB/s eta 0:00:01
   -------------------------------------   67.3/70.9 MB 6.1 MB/s eta 0:00:01
   --------------------------------------  67.7/70.9 MB 6.1 MB/s eta 0:00:01
   --------------------------------------  67.8/70.9 MB 6.0 MB/s eta 0:00:01
   --------------------------------------  68.2/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------  68.4/70.9 MB 6.1 MB/s eta 0:00:01
   --------------------------------------  68.8/70.9 MB 6.1 MB/s eta 0:00:01
   --------------------------------------  68.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 69.2/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 69.5/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 69.8/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 69.9/70.9 MB 6.1 MB/s eta 0:00:01
   --------------------------------------- 70.1/70.9 MB 6.1 MB/s eta 0:00:01
   --------------------------------------- 70.5/70.9 MB 6.1 MB/s eta 0:00:01
   --------------------------------------- 70.8/70.9 MB 6.3 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
   --------------------------------------- 70.9/70.9 MB 6.2 MB/s eta 0:00:01
```

```
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------  70.9/70.9 MB 6.2 MB/s eta 0:00:01
     ---------------------------------------- 70.9/70.9 MB 2.3 MB/s eta 0:00:00
Requirement already satisfied: numpy in c:\users\dell\appdata\local\programs\pyth
on\python311\lib\site-packages (from xgboost) (1.24.2)
Requirement already satisfied: scipy in c:\users\dell\appdata\local\programs\pyth
on\python311\lib\site-packages (from xgboost) (1.10.1)
Installing collected packages: xgboost
Successfully installed xgboost-1.7.5
```

In [59]:

```python
from xgboost import XGBRegressor
model = XGBRegressor()
model.fit(x_train,y_train)
```

Out[59]:

```
▾                          XGBRegressor
XGBRegressor(base_score=None, booster=None, callbacks=None,
             colsample_bylevel=None, colsample_bynode=None,
             colsample_bytree=None, early_stopping_rounds=None,
             enable_categorical=False, eval_metric=None, feature_typ
es=None,
             gamma=None, gpu_id=None, grow_policy=None, importance_t
ype=None,
             interaction_constraints=None, learning_rate=None, max_b
in=None,
```

In [60]:

```python
y_pred2 = model.predict(x_test)
```

In [61]:

```python
print("R2 score  :",r2_score(y_test, y_pred2))
print("MSE score  :",mean_squared_error(y_test, y_pred2))
```

```
print("RMSE: ",sqrt(mean_squared_error(y_test, y_pred2)))
```

```
R2 score  : 0.9267737558896152
MSE score  : 38264251.77563304
RMSE:  6185.810518891848
```

In [62]:
```
y_pred2
```

Out[62]:
```
array([ 2374.166, 17729.592, 10288.253, ..., 55534.836, 43356.68 ,
       -1123.841], dtype=float32)
```

In [63]:
```
#Regularization
from sklearn.linear_model import Ridge
rr_model = Ridge(alpha=0.5)
rr_model.fit(x_train,y_train)
```

Out[63]:
```
Ridge
Ridge(alpha=0.5)
```

In [64]:
```
y_pred3 = model.predict(x_test)
```

In [65]:
```
y_pred3
```

Out[65]:
```
array([ 2374.166, 17729.592, 10288.253, ..., 55534.836, 43356.68 ,
       -1123.841], dtype=float32)
```

In [66]:
```
print("R2 score  :",r2_score(y_test, y_pred3))
print("MSE score  :",mean_squared_error(y_test, y_pred3))
print("RMSE: ",sqrt(mean_squared_error(y_test, y_pred3)))
```

```
R2 score  : 0.9267737558896152
MSE score  : 38264251.77563304
RMSE:  6185.810518891848
```

In [67]:
```
y_test
```

Out[67]:
```
197943     4988.57
341383    23214.66
266827    13078.92
168560    18423.55
45041       937.16
             ...
322326      470.08
21999        20.96
260496    70459.25
211383    43923.35
405197       57.04
Name: Weekly_Sales, Length: 105053, dtype: float64
```

In [ ]:
```

```