

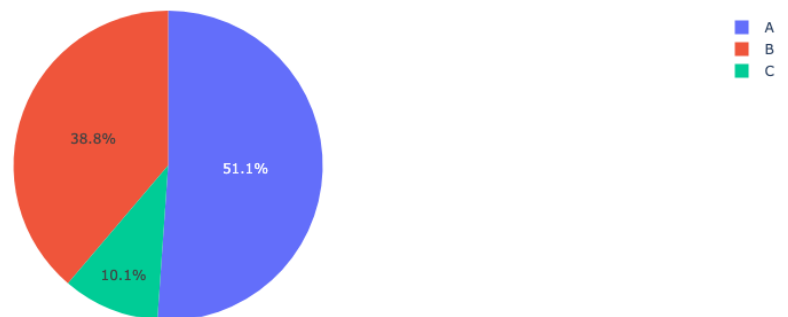
## Walmart Sales Analysis notes

- Using PySpark.
- Dataset- Walmart-recruiting-store-sales-prediction  
<https://www.kaggle.com/competitions/walmart-recruiting-store-sales-forecasting/data>

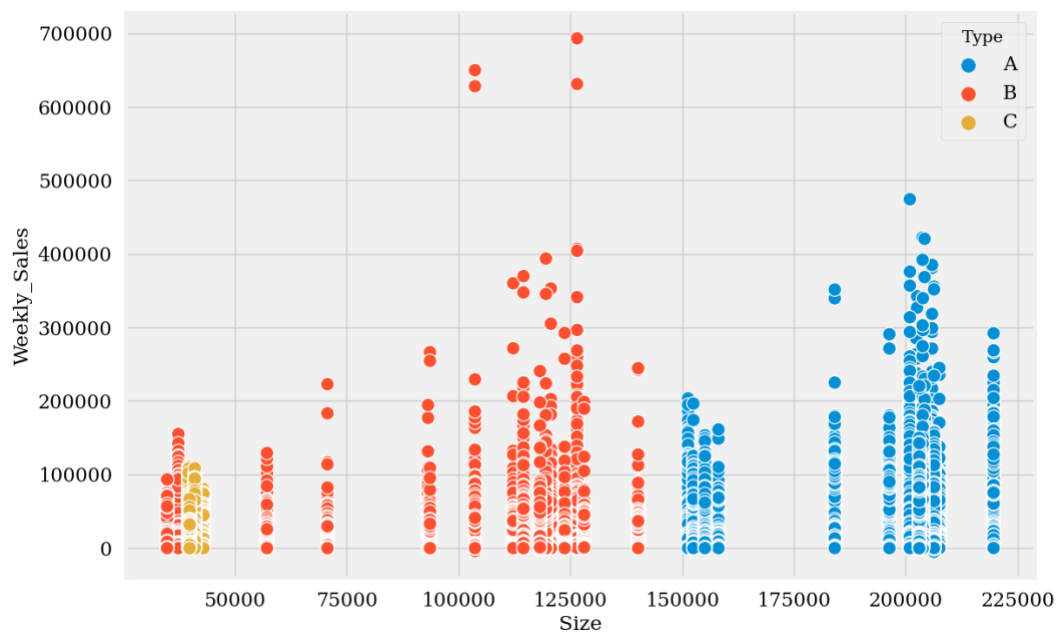
### Preprocessing and Data Analysis:

- Most of the Store types in the data set are TYPE A.

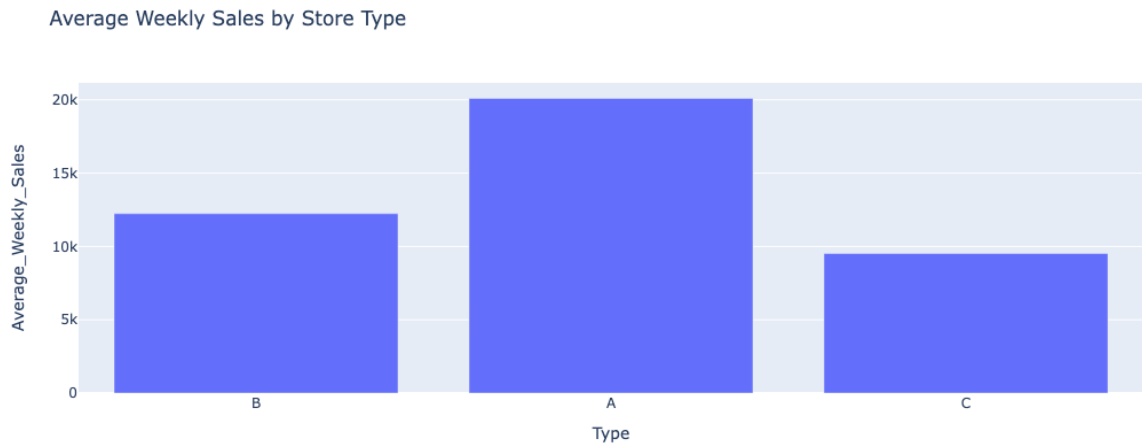
Types of store



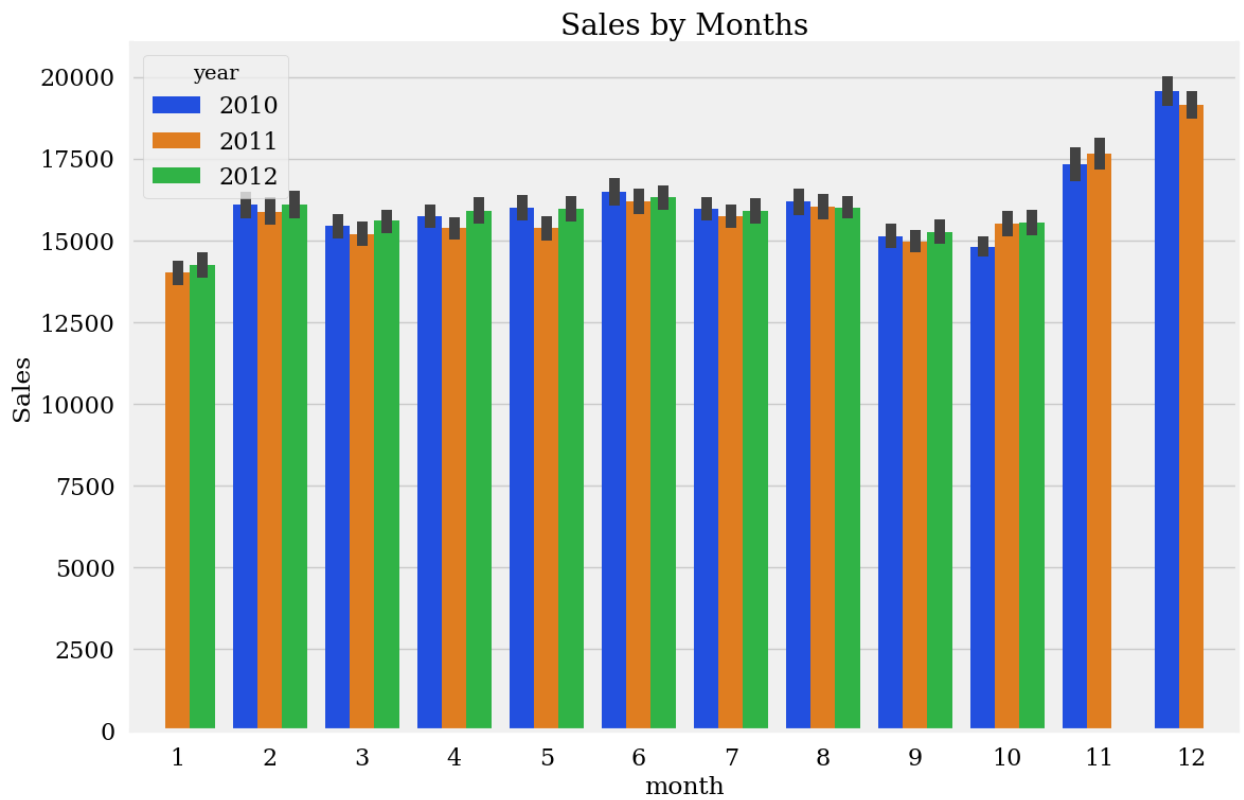
- As the Size of the store increases the sales also increases.



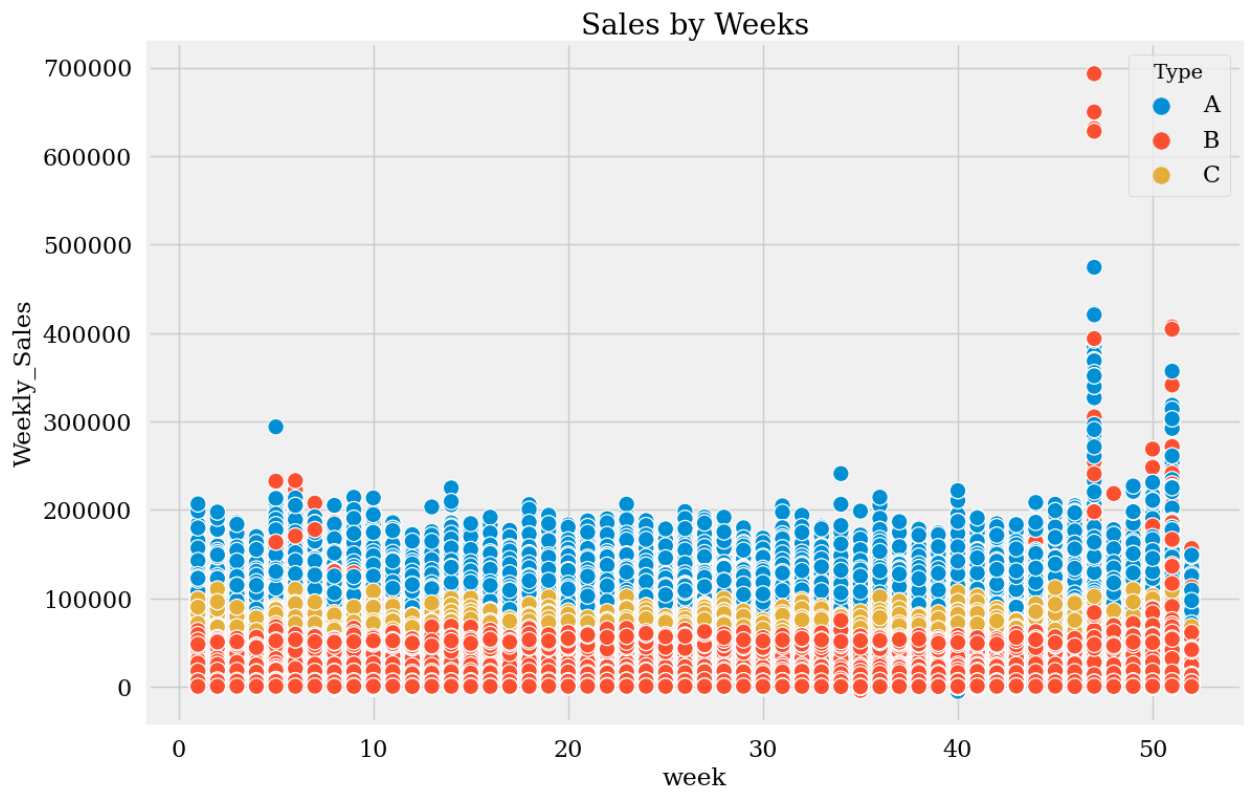
- Type A Store has done most of the sales.



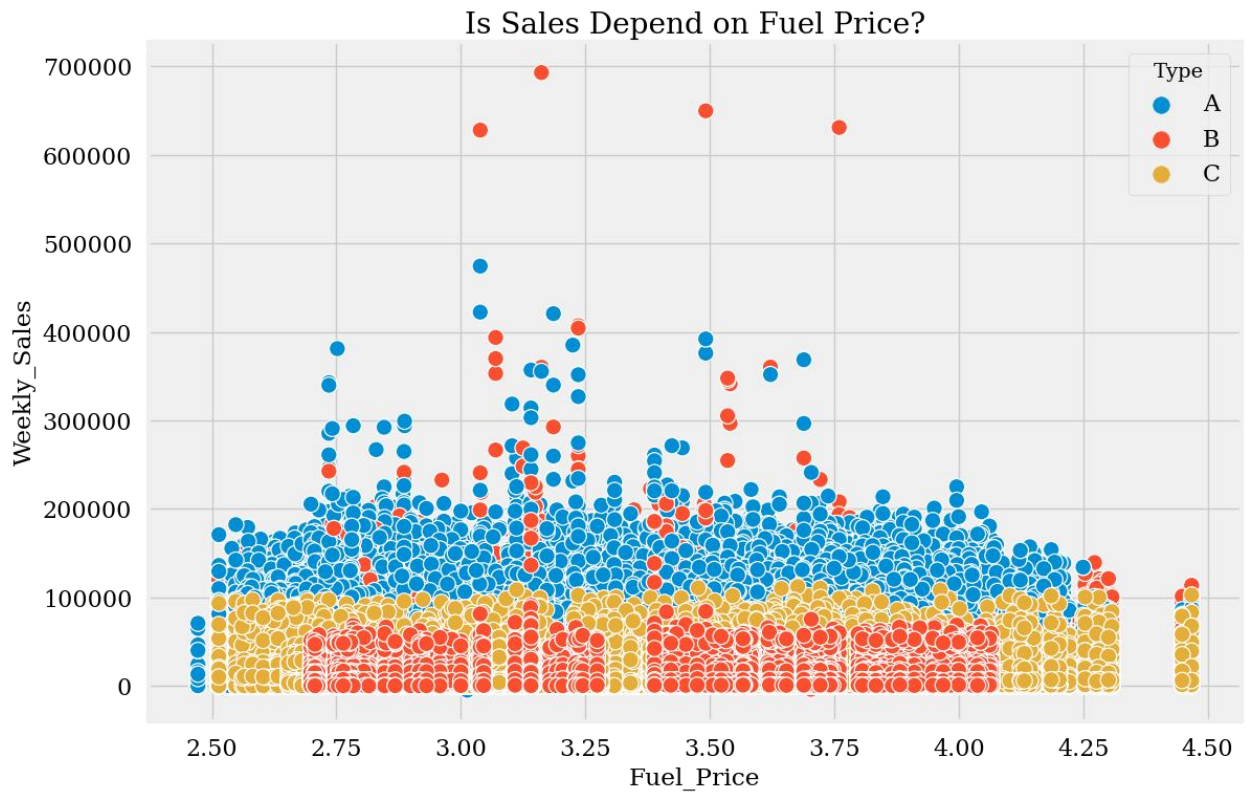
- only slight variation in sales with respect year
- In 2011 sales are relatively lower than both 2010 and 2012
- In first month sales are least and in 12th month sales are higher .May be due to festivals



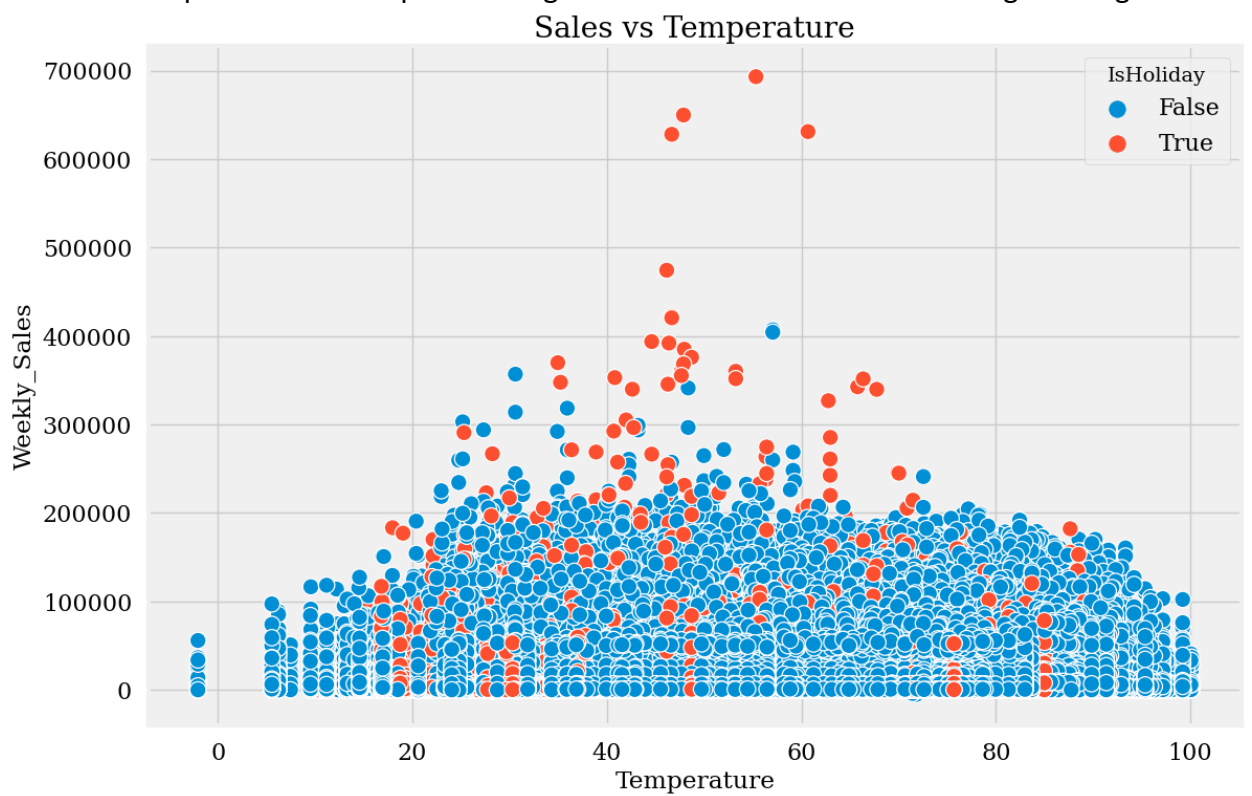
- During the holiday there is a slight increase of sales.



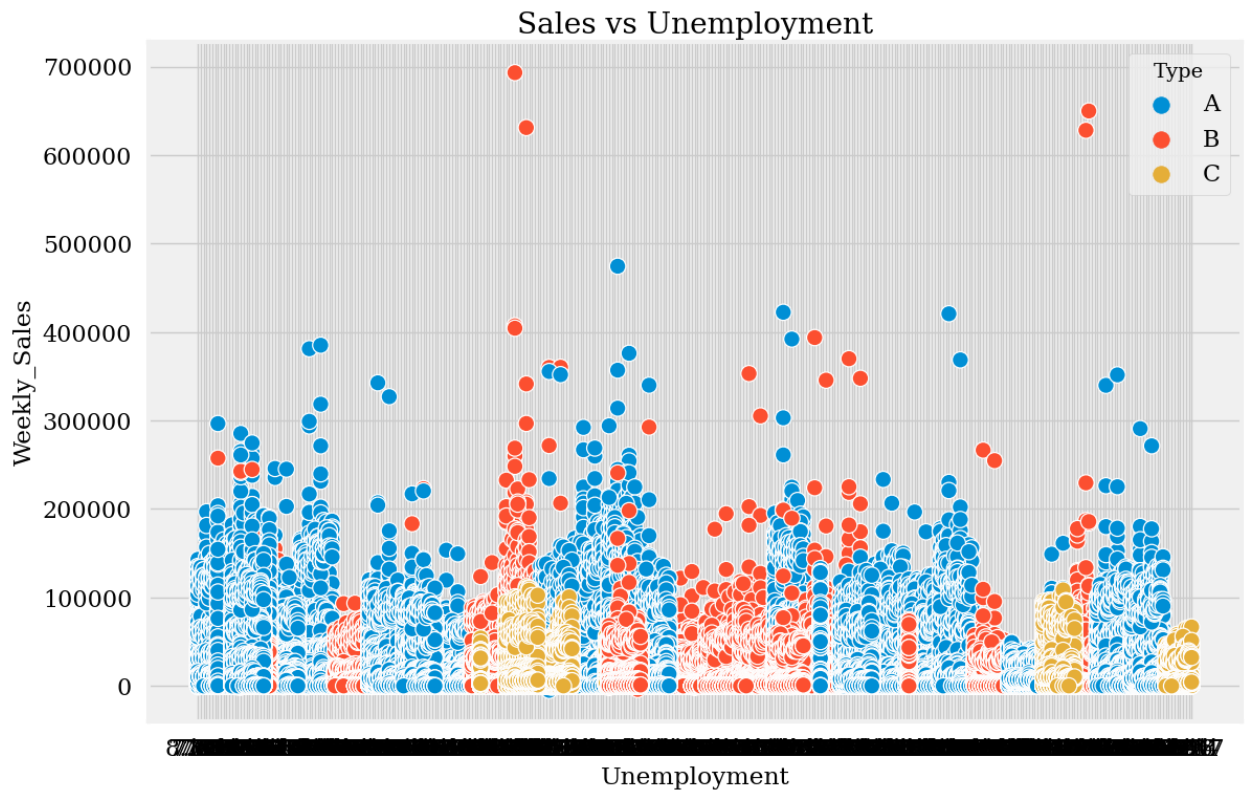
- Sales does not depend on the fuel.



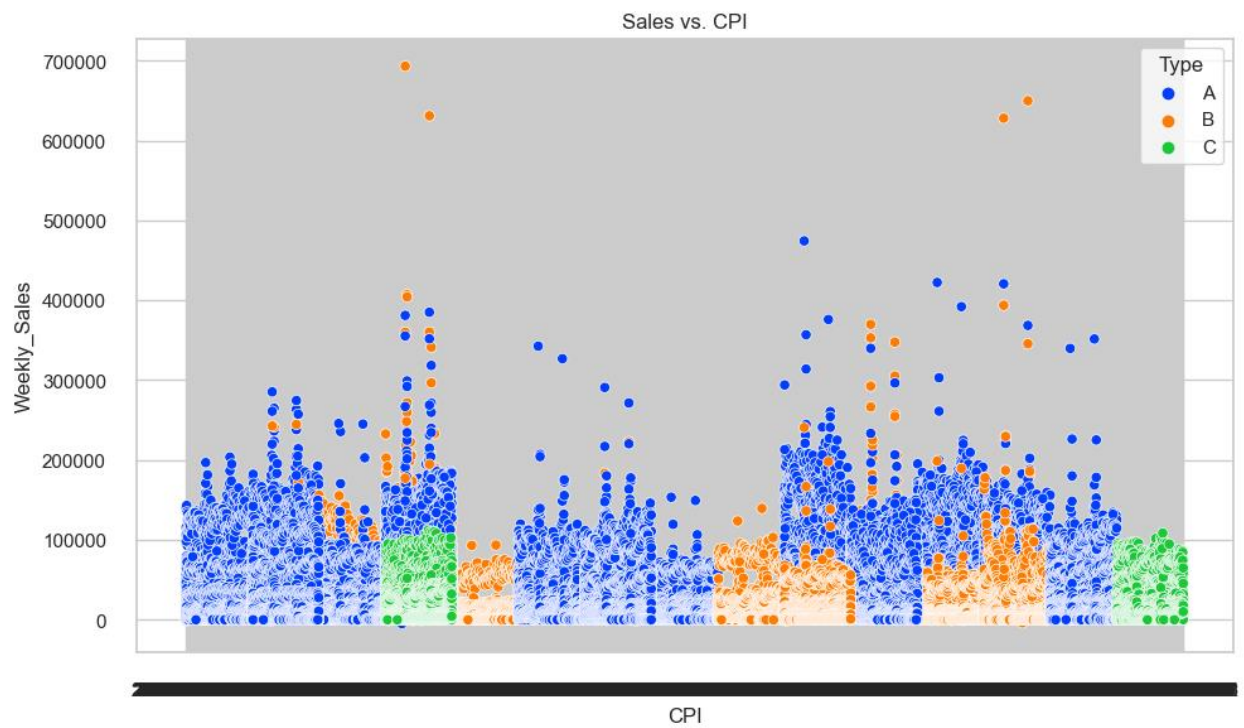
- Sales don't depend on the temp but during too cold and too hold there is a slight change.



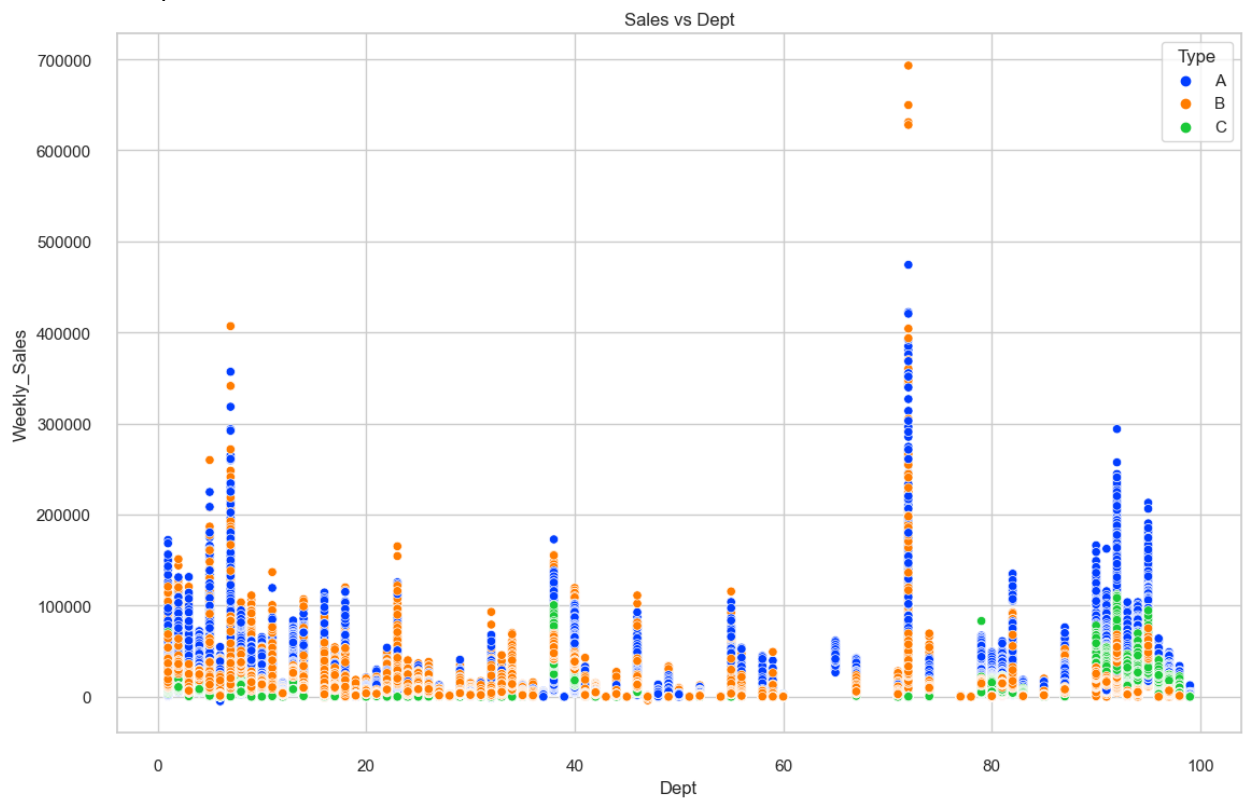
- Only slight changes in the sales when unemployment increases.



- CPI(Consumer Price Index) does not effect sales.



- Dept does affect sales
- Different department have different sales



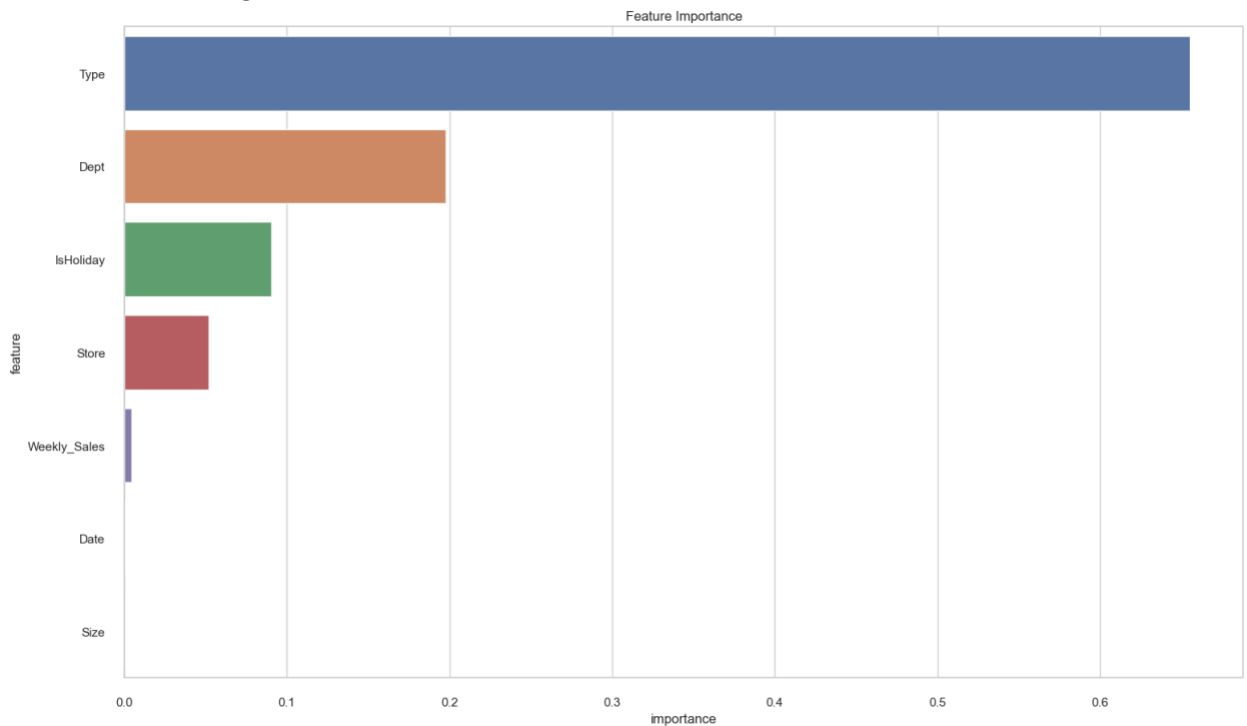
- Based on the features we have removed some of the features that are not important and kept 'Store', 'IsHoliday', 'Type', 'Size', 'week', 'Dept', 'year'.

#### Models Analysis:

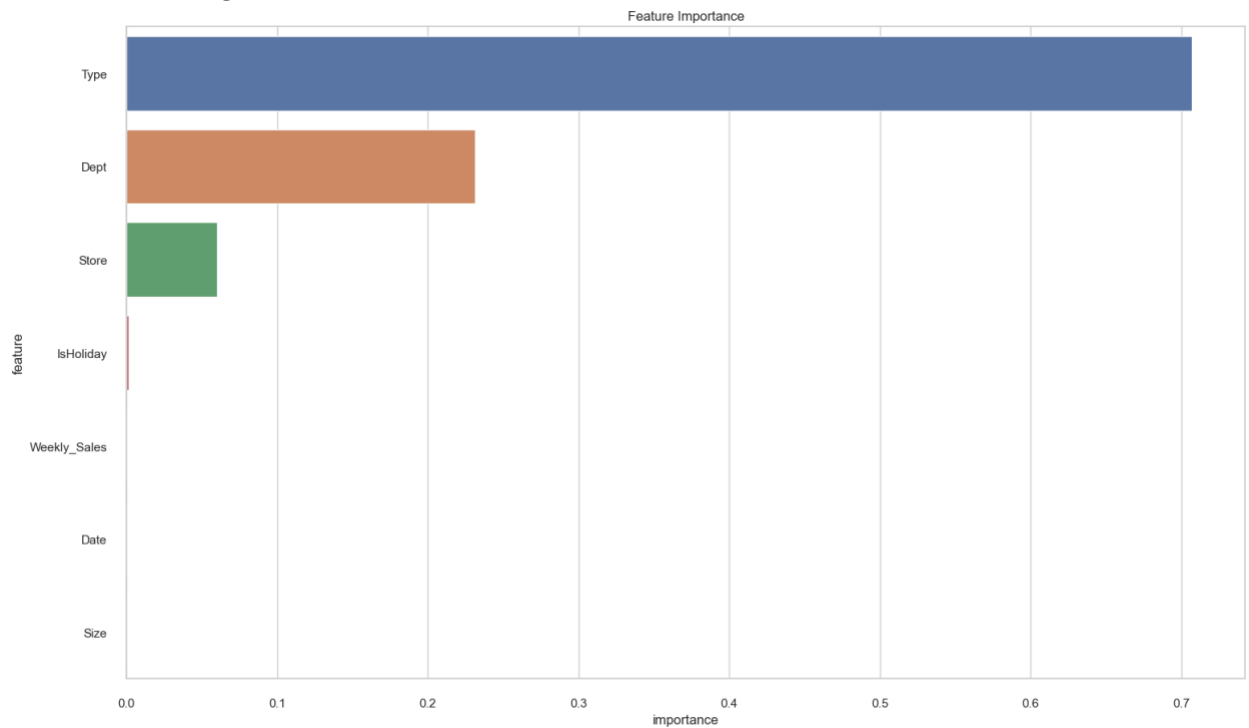
- Random Forest Regressor:**  
The WMAE loss for the training set is 10986.69728484865.  
The WMAE loss for the validation set is 10842.144435508004.
- Linear Regression:**  
RMSE on training data: 14569.467490828793  
RMSE on validation data: 14583.551300042831
- Decision Tree Regression:**  
The WMAE loss for the training set is 9468.314207525373.  
The WMAE loss for the validation set is 9183.5112872512.

#### Feature Importance:

- Random Forest Regressor:



- Decision Tree Regressor:



Summary:

1. Sales heavily depend on Type and Department.
2. Size of store also play major role in sales big store usually have more weekly\_sales.
3. A type store did more sales than b and c.
4. weekly sales also depend on week of the year holidays weeks are good for sales.