**SALES FORECASTING FOR RETAIL STORES**

**AIM:**

* The objective is predicting store sales using historical markdown data. One challenge of modelling retail data is the need to make decisions based on limited history. If Christmas comes but once a year, so does the chance to see how strategic decisions impacted the bottom line.

**BUSINESS PROBLEM:**

* You are provided with historical sales data for 45 stores located in different regions. Each store contains a number of departments, and you are tasked with predicting the department-wide sales for each store. In addition, Store runs several promotional markdown events throughout the year.
* These markdowns precede prominent holidays, the four largest of which are the Super Bowl, Labour Day, Thanksgiving, and Christmas. The weeks including these holidays are weighted five times higher in the evaluation than non-holiday weeks.
* Part of the challenge presented by this competition is modelling the effects of markdowns on these holiday weeks in the absence of complete/ideal historical data.

**DATA AVAILABILITY:**

* **stores.csv:** This file contains anonymized information about the 45

stores, indicating the type and size of the store.

* **features.csv:** Comprise of information on Markdown and

macroeconomic factors.

* **train.csv**: This is the historical training data.

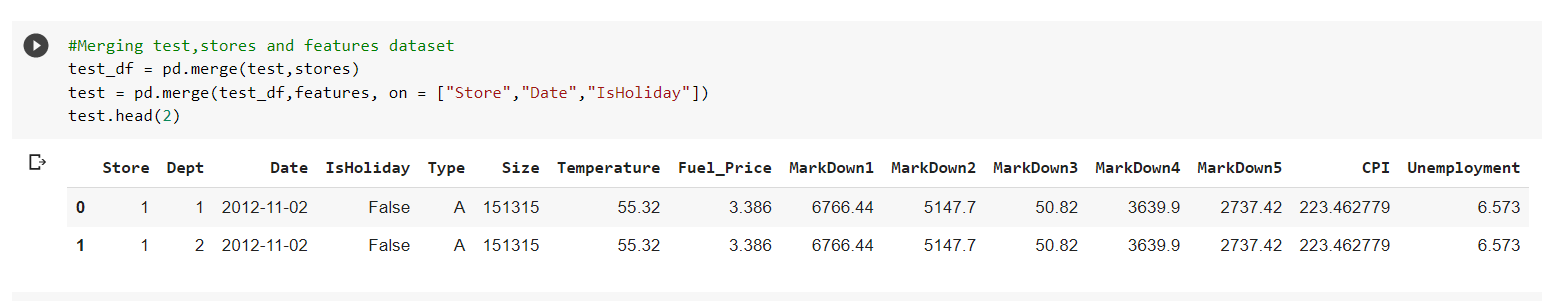
## UNDERSTANDING THE PROBLEM :

* Data from stores across the different regions are given, and it is up to us to forecast their weekly sales. The data is already splitted into a training and a test set, and we want to fit a model to the training data to forecast those weeks sales as accurately as possible.
* Here, we are provided dataset combining of 45 stores including store information and monthly sales. The data is provided on weekly basis. Walmart tries to find the impact of holidays on the sales of store. For which it has included four holidays’ weeks into the dataset which are Christmas, Thanksgiving, Super bowl, Labor Day. Here we are owing to analyze the dataset given.

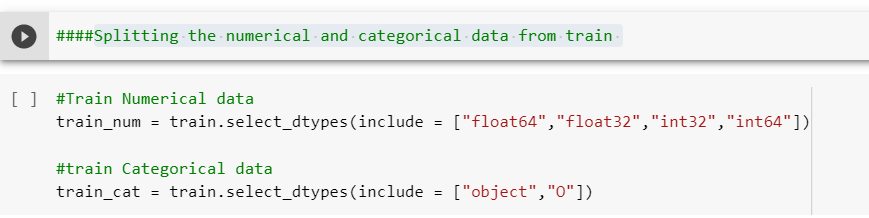
**LOADING THE DATASET:**

* Here we used Google colab for the prediction along with the libraries such as Pandas,Numpy,Matplotlib,Seaborn.
* ****

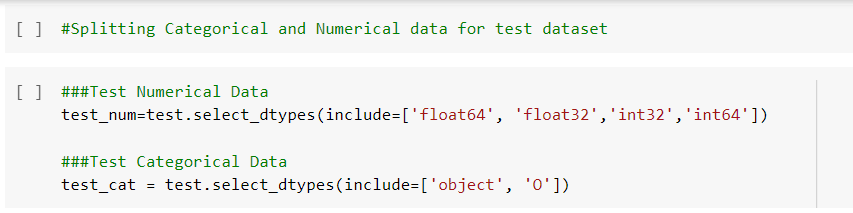
Merging the dataset test ,stores and features.



Splitting the numerical and categorical data from train .



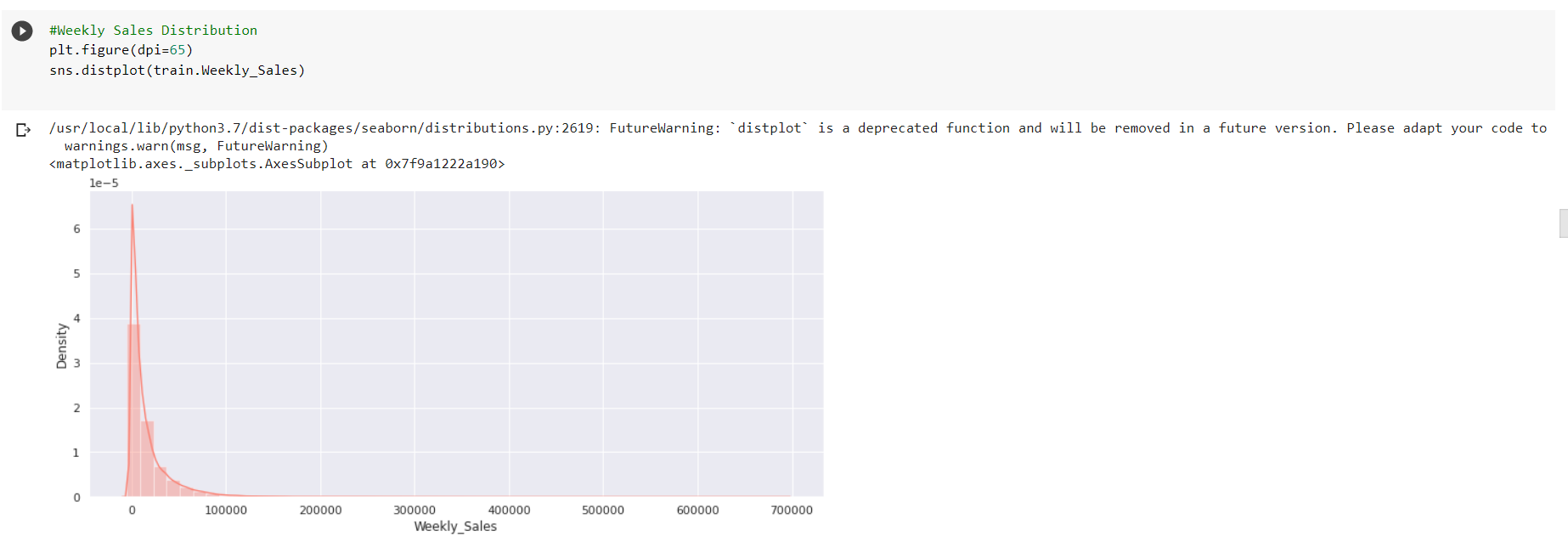
**Do the same for the test data:**

****

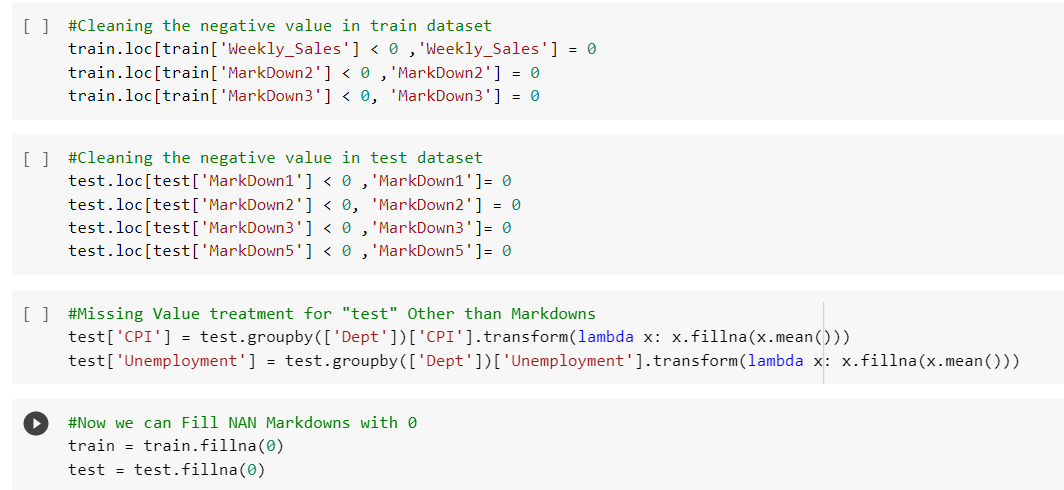
**Data Analaysis of the stores:**

****

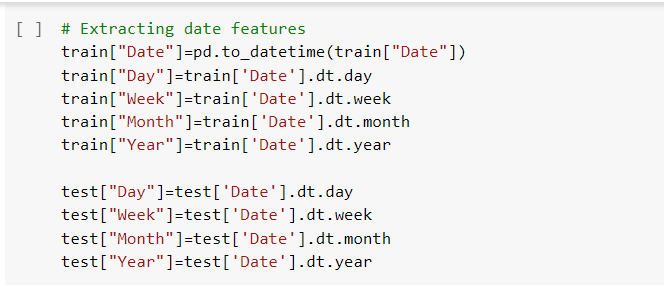
Weekly sales Distribution:



**Cleaning the negative values in the dataset:**

****

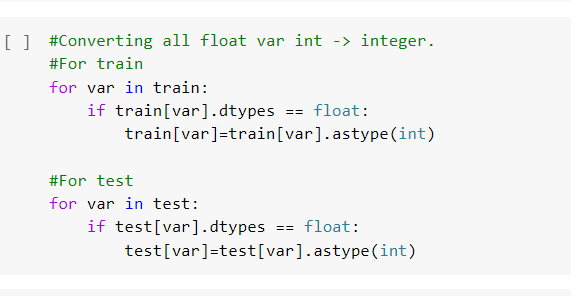
**Extracting the date features:**

****

**Here we categorized dates :**

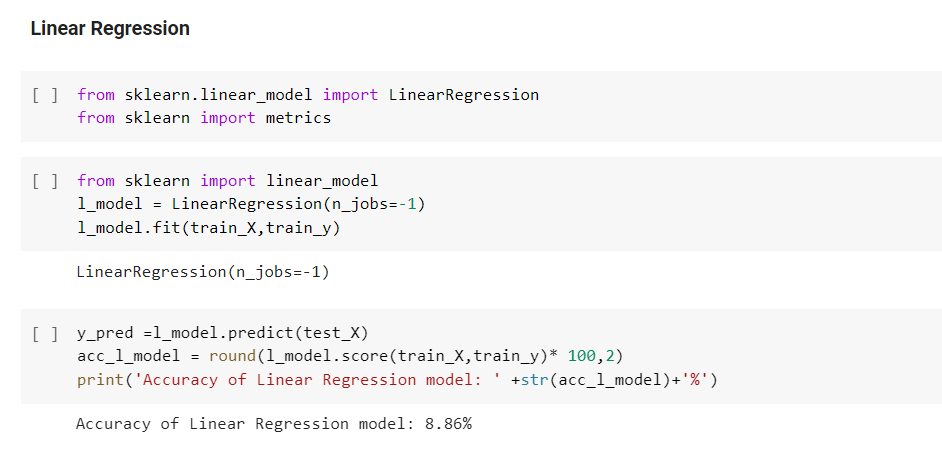
****

**Converting all float into integers:**

****

* **After these data building process we get into the classification of the dataset.**
* **Here the data ready for the prediction using linear regression,random forest regression,**

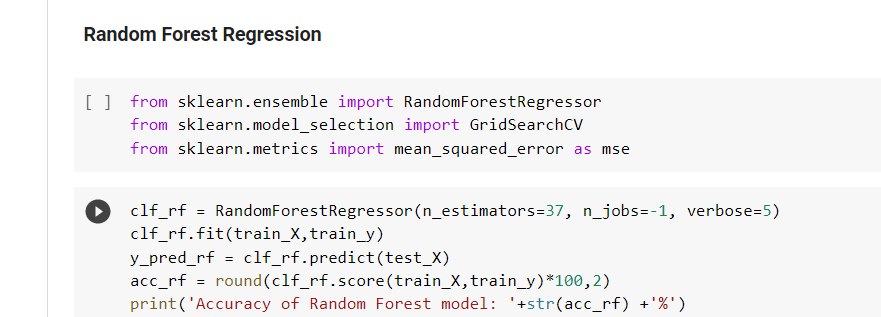
**Linear Regression:**

****

* From Sklearn library we predicted the Linear Regression for the dataset.
* The Accuracy for Linear Regression model is 8.86% .

**Random Forest Regression :**

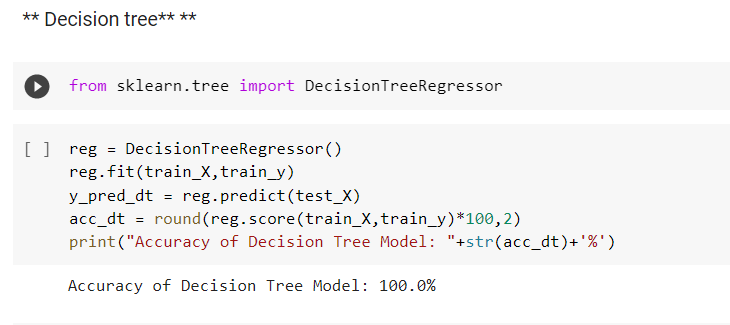
* Then we used Random forest regression to predict the dataset for more accuracy better than the other.

****

**The Accuracy for the Random forest regression is 99.64 %**

**Decision tree :**

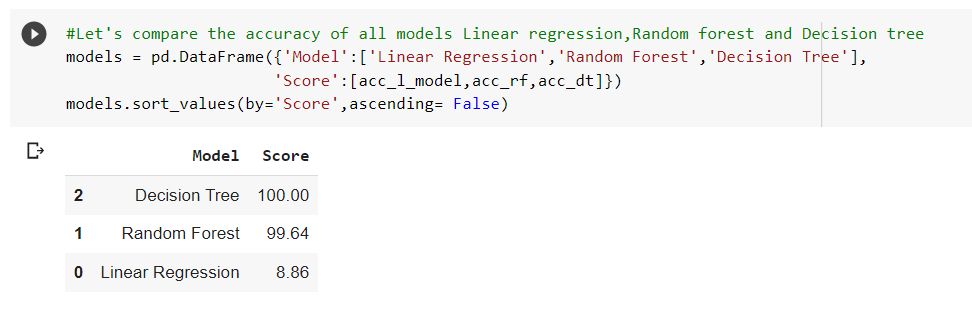
* We cannot conclude with two predictions so we used Decision tree model to predict the dataset accuracy.



* The decision tree provides the model with 100% accuracy score.

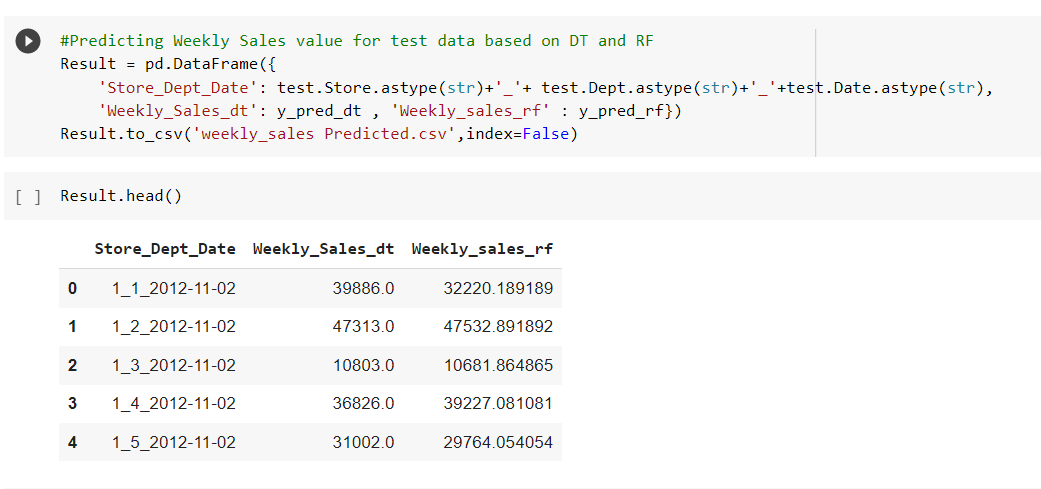
**Comparision the Accuracy score:**

* Lets compare the accuracy score for the for better result .



**Report:**

* Here we predicted the weekly sales for the model based on the DT and RF tools.
* The final predicted result shown in figure below.



THANK YOU