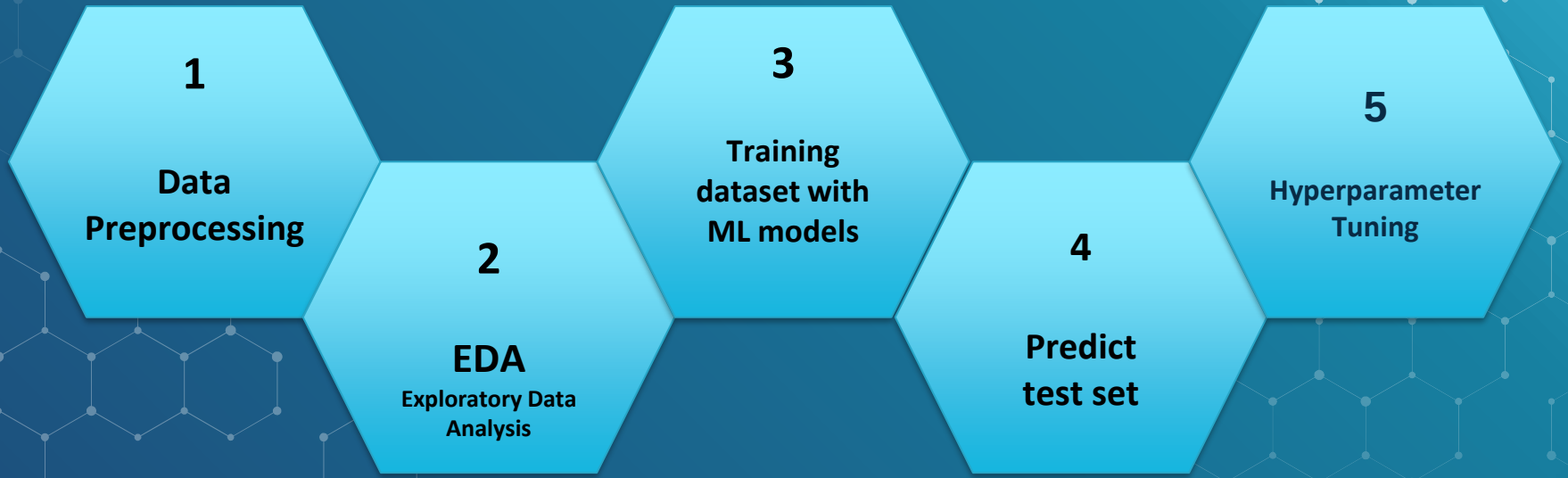
Store Sales Forecasting

Process of Project



Data Preprocessing

The following methods are used for having normal datasets.

- Combining data

- Data Types :
  - Numerical
  - Categorical

- Data Wrangling:
  - Splitting column
  - Label Encoding
  - Handle Missing Value

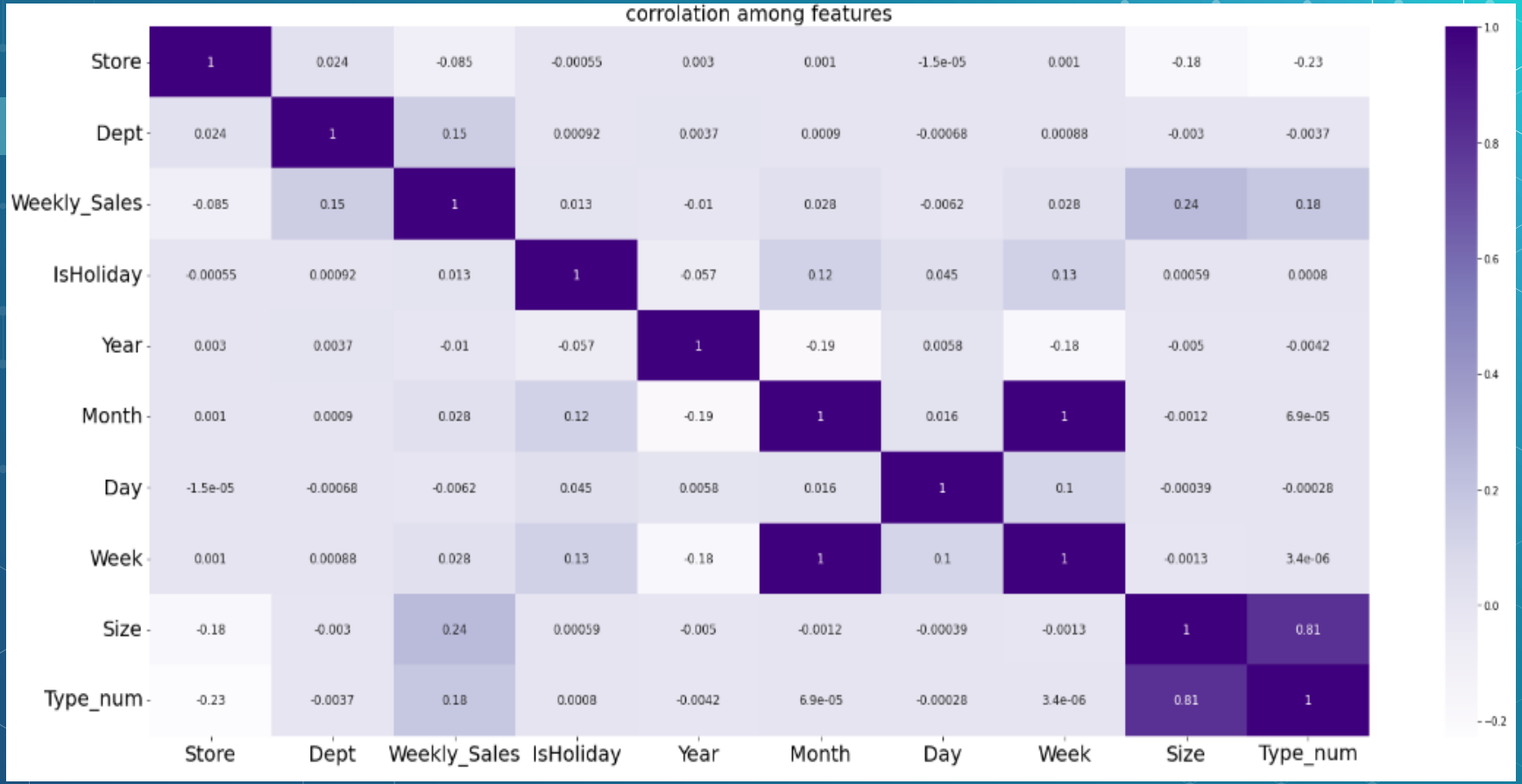
Results of Preprocessing

- Row Data

	Store	Dept	Date	Weekly_Sales	IsHoliday	Type	Size
0	1	1	2010-02-05	24924.50	False	A	151315
1	1	1	2010-02-12	46039.49	True	A	151315
2	1	1	2010-02-19	41595.55	False	A	151315
3	1	1	2010-02-26	19403.54	False	A	151315
4	1	1	2010-03-05	21827.90	False	A	151315

- Processed Data

	Store	Dept	Weekly_Sales	IsHoliday	Size	Year	Week	Type_num
0	1	1	24924.50	False	151315	2010	5.0	3
1	1	1	46039.49	True	151315	2010	6.0	3
2	1	1	41595.55	False	151315	2010	7.0	3
3	1	1	19403.54	False	151315	2010	8.0	3
4	1	1	21827.90	False	151315	2010	9.0	3


corrolation among features
Store 1 0.024 -0.085 -0.00055 0.003 0.001 -1.5e-05 0.001 -0.18 -0.23
Dept 0.024 1 0.15 0.00092 0.0037 0.0009 -0.00068 0.00088 -0.003 -0.0037
Weekly_Sales -0.085 0.15 1 0.013 -0.01 0.028 -0.0062 0.028 0.24 0.18
IsHoliday -0.00055 0.00092 0.013 1 -0.057 0.12 0.045 0.13 0.00059 0.0008
Year 0.003 0.0037 -0.01 -0.057 1 -0.19 0.0058 -0.18 -0.005 -0.0042
Month 0.001 0.0009 0.028 0.12 -0.19 1 0.016 1 -0.0012 6.9e-05
Day -1.5e-05 -0.00068 -0.0062 0.045 0.0058 0.016 1 0.1 -0.00039 -0.00028
Week 0.001 0.00088 0.028 0.13 -0.18 1 0.1 1 -0.0013 3.4e-06
Size -0.18 -0.003 0.24 0.00059 -0.005 -0.0012 -0.00039 -0.0013 1 0.81
Type_num -0.23 -0.0037 0.18 0.0008 -0.0042 6.9e-05 -0.00028 3.4e-06 0.81 1
Store Dept Weekly_Sales IsHoliday Year Month Day Week Size Type_num
1.0 0.8 0.6 0.4 0.2 0.0 −0.2

Train Set:

- Shape: 421570 rows, 8 columns
- Number of Stores: 45
- Number of Dept: 99
- Number of Holidays: 29661
- Number of Stores by Type :

  A: 215478, B: 163495, C: 42597
- Number of Weeks: 143
- Date:

  Start: 2010-02-05 , End: 2012-10-26

Test set:

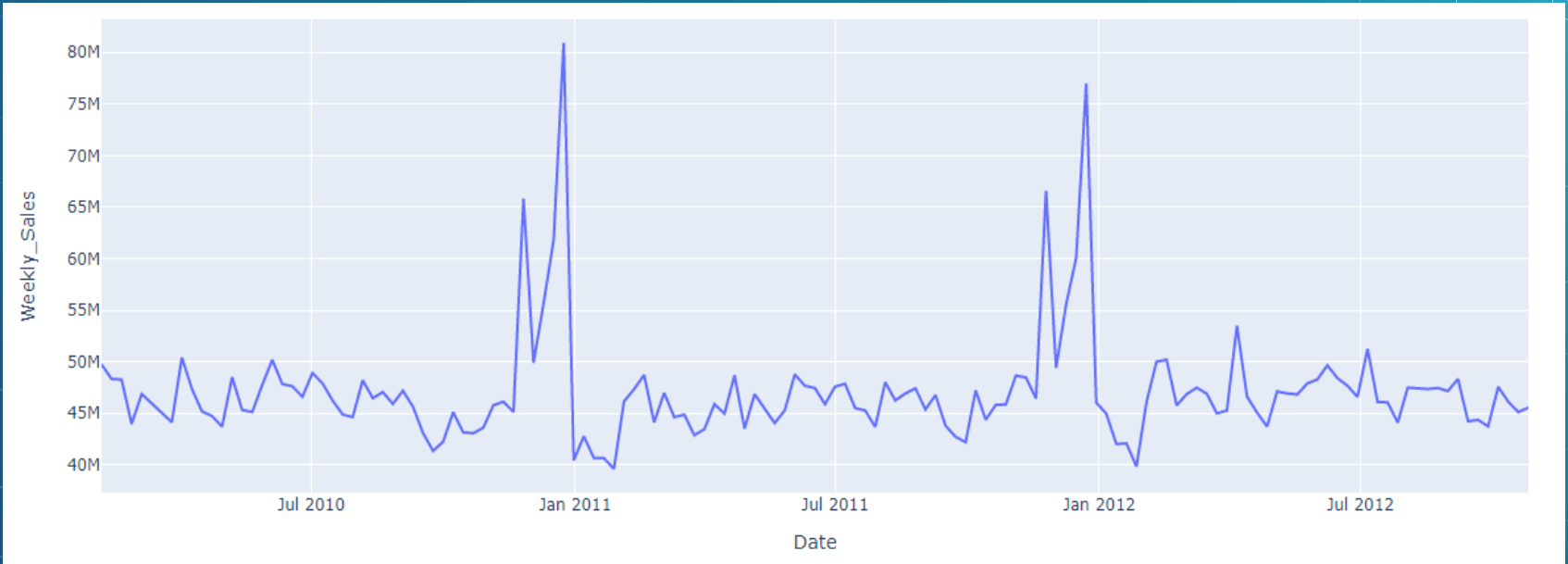
- Shape: 115064 rows, 7 columns
- Number of Stores: 45
- Number of Dept: 99
- Number of Holidays: 8928
- Number of Stores by Type :

  A: 58713, B: 44500, C: 11851
- Number of Weeks: 61
- Date:

  Start: 2012-11-02 , End: 2013-07-26

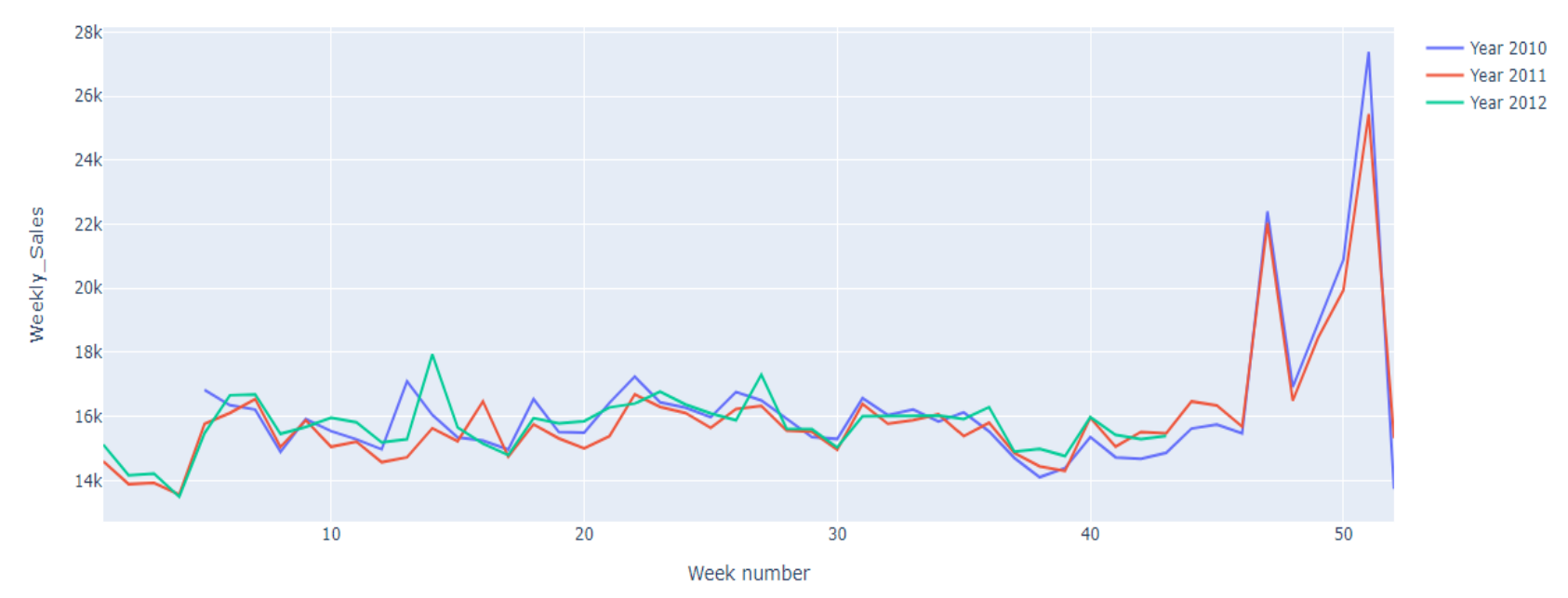
EDA (Exploratory Data Analysis)

Overall Weekly Sales of Walmart



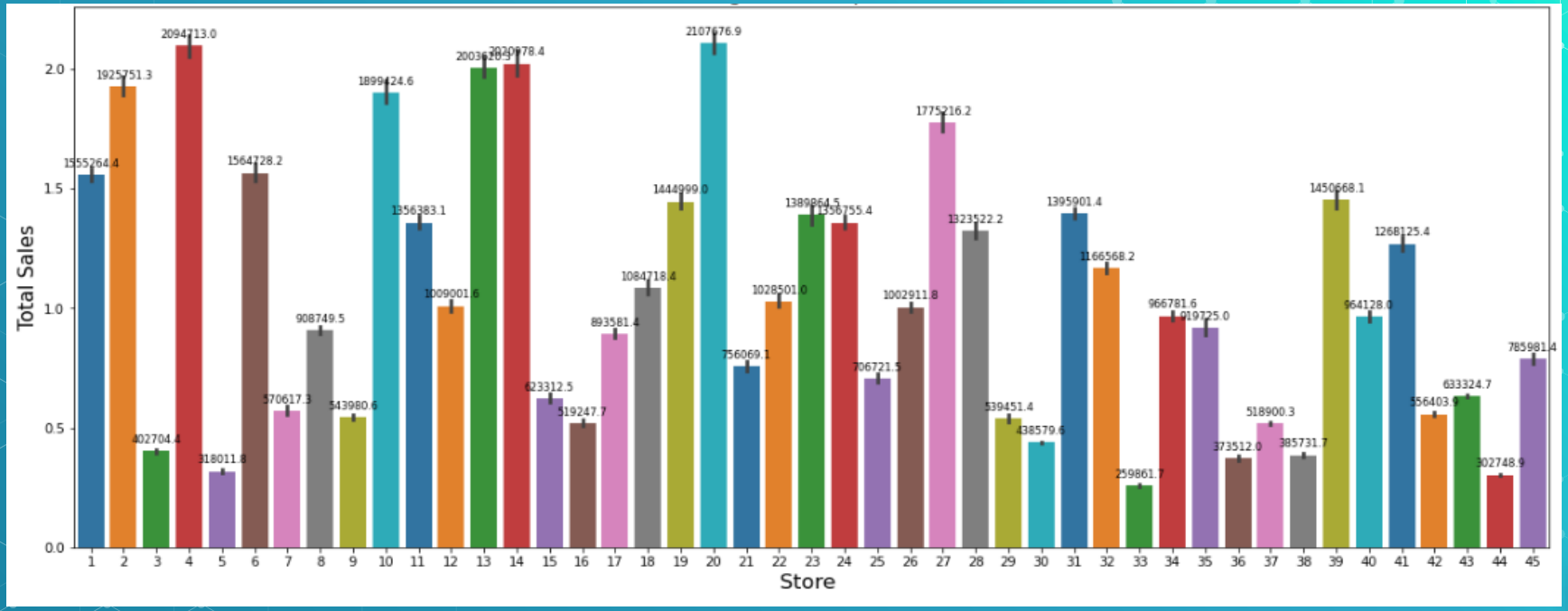
EDA (Exploratory Data Analysis)

Overall Weekly Sales of Walmart



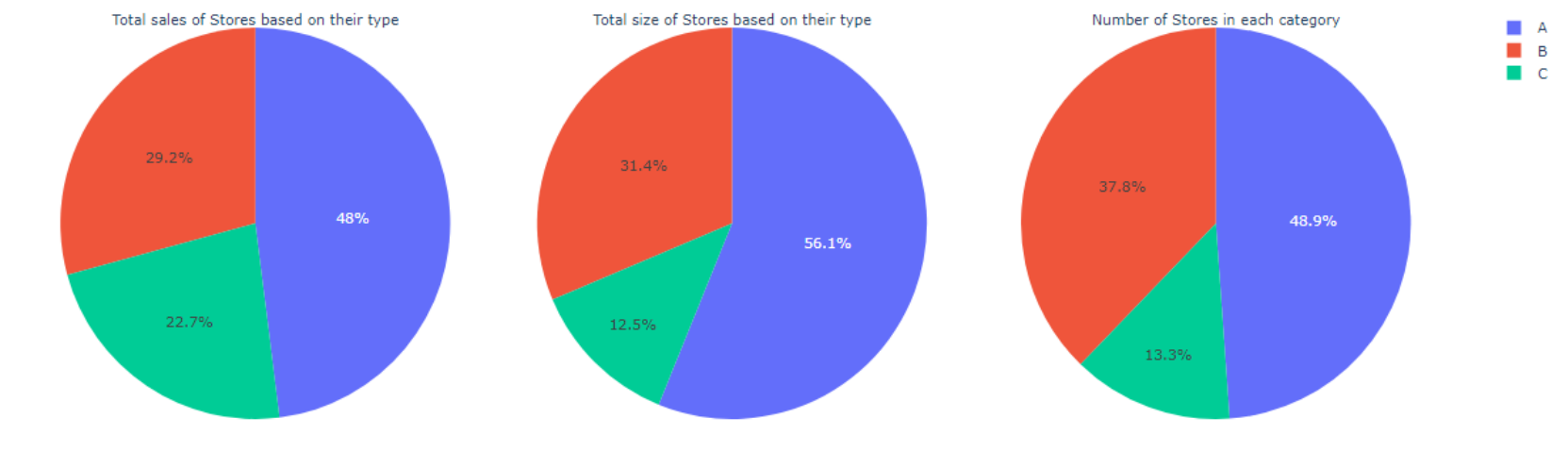
EDA (Exploratory Data Analysis)

Distribution of total sales per store



EDA (Exploratory Data Analysis)

Analyzing each store based on Type

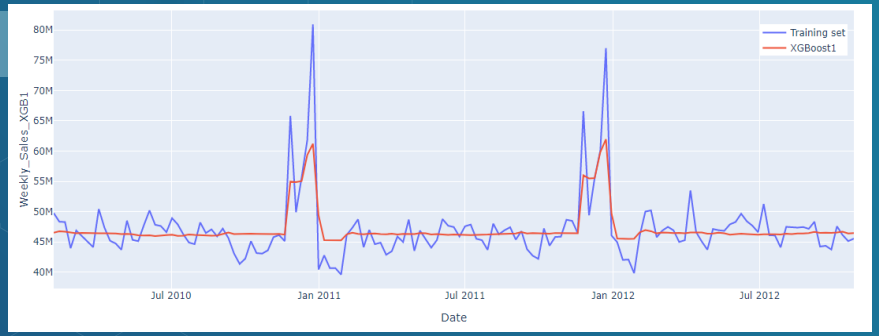


Training dataset with ML models

- Data Normalization
- XGBoost
- Random Forrest

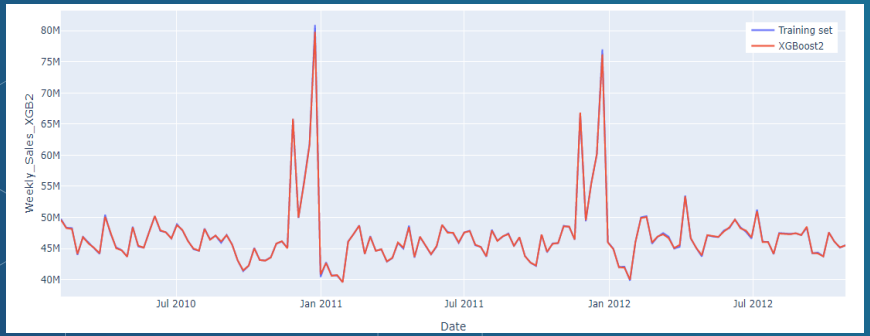
XGBoost Train Forecasting

Initial Parameters



wMAE= 7069.97

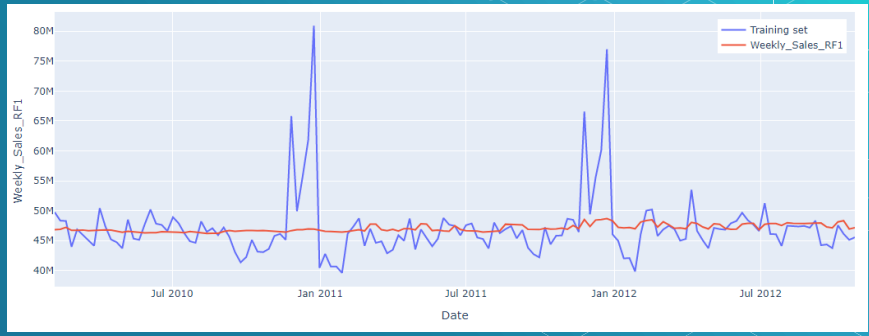
HyperParameters



wMAE= 0.03

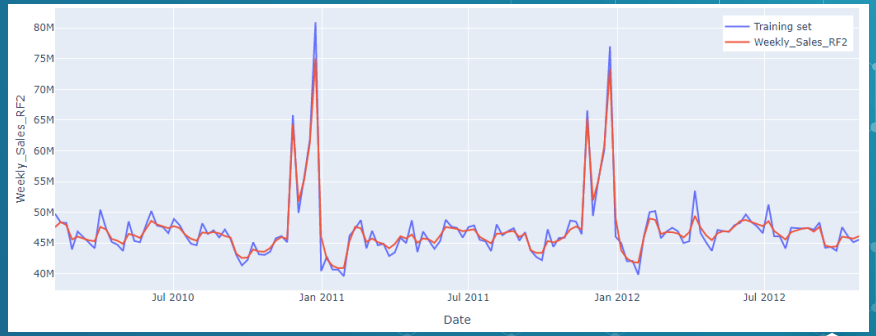
RandomForrest Train Forecasting

Initial Parameters



wMAE= 13046.2

HyperParameters



wMAE= 750.63

Experimental result:

Table: Forecasting results of models

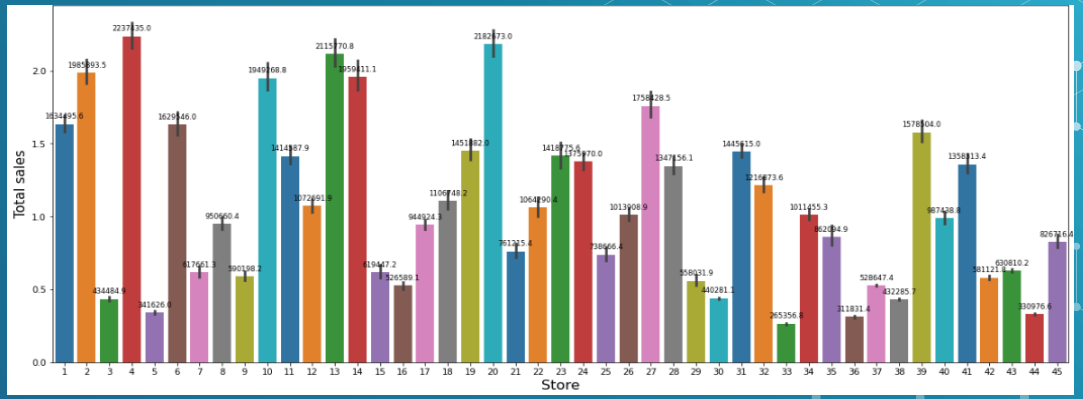
	Store	Dept	Date	IsHoliday	Weekly_Sales_XGB1	Weekly_Sales_XGB2	Weekly_Sales_RF1	Weekly_Sales_RF2
0	1	1	2012-11-02	False	25300.494141	35023.472656	20793.374716	29759.633429
1	1	1	2012-11-09	False	25300.494141	16426.974609	20793.374716	19786.616138
2	1	1	2012-11-16	False	25300.494141	16820.726562	20793.374716	19568.582196
3	1	1	2012-11-23	True	31682.855469	20637.279297	20793.374716	20376.302944
4	1	1	2012-11-30	False	31682.855469	24474.115234	20793.374716	23590.892365

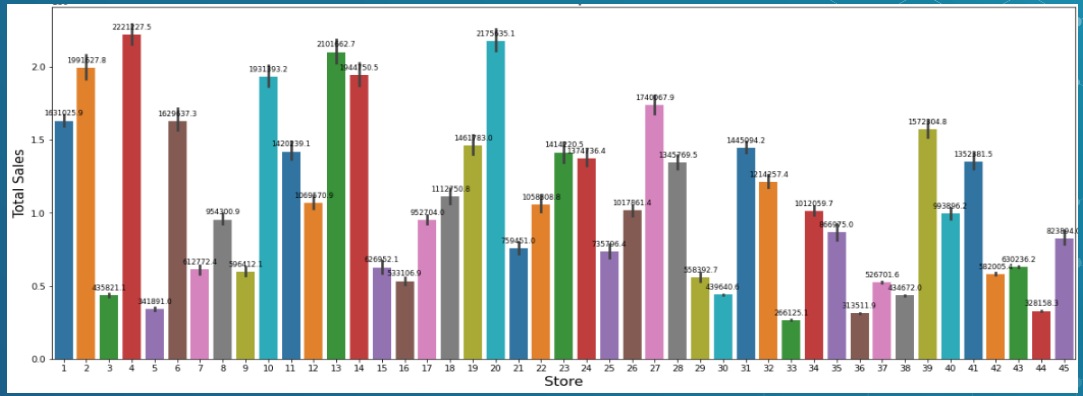
Experimental results

XGBoost results

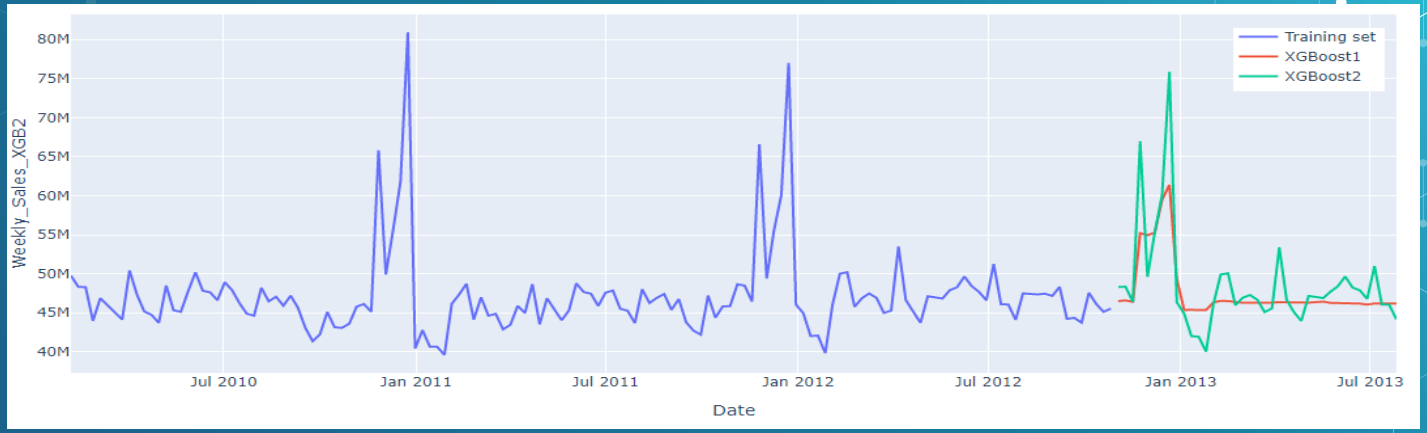
RandomForest results

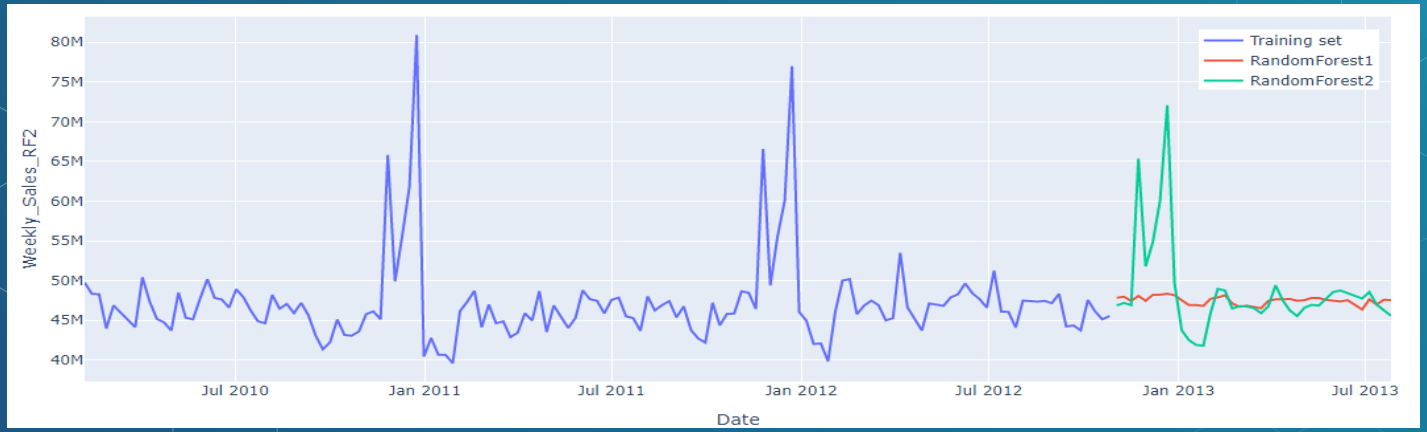
Distribution of total sales per store





Forecasting results of models





Experimental result:

Overall Weekly Sales of Walmart

Kaggle Scores

	Private Score	Public score
XGBoost	18900.44731	18497.52428
RandomForest	20810.73280	20460.24320
HyperParameter Tunning		
XGBoost	**2882.65763**	**2790.63733**
RandomForest	**2959.36412**	**2852.76808**