Use case 5 :

Retails data analytics :

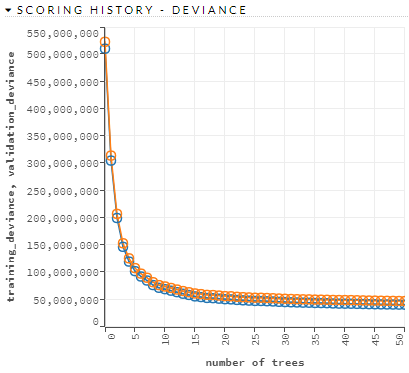
This dataset contains historical sales data, which covers to 2010-02-05 to 2012-11-01. We shall find the following fields:

* Store - the store number
* Dept - the department number
* Date - the week
* Weekly\_Sales - sales for the given department in the given store
* IsHoliday - whether the week is a special holiday week

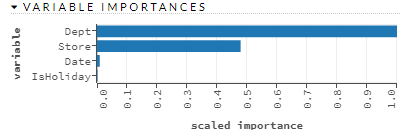
I have used the GBM algorithm to regress the dataset.

I have divided the dataset into 3 sections : training(50%), validation (25%) and test(25%)

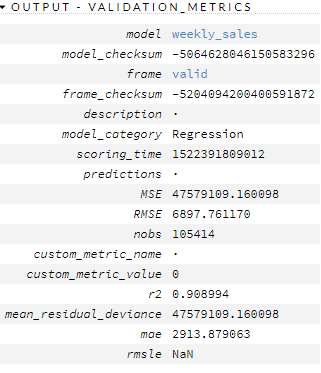
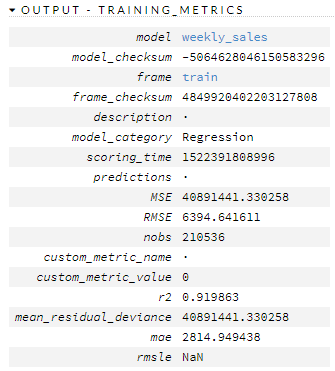
SCORING HISTORY :

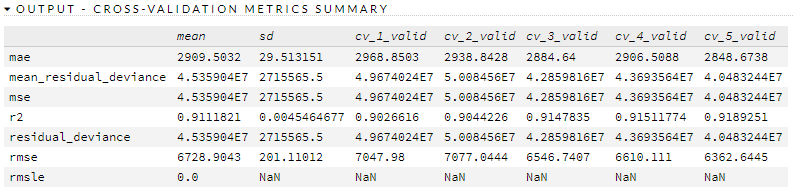


This shows that te deviance is low . Hence it has greater bond, which shows the model has a good reliability.



According to the variable importance the sales is not dependent upon date, rather it's highly dependent on Department and the store.

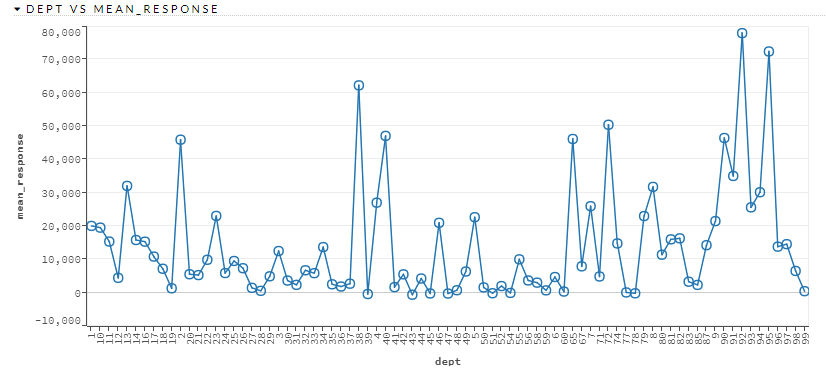




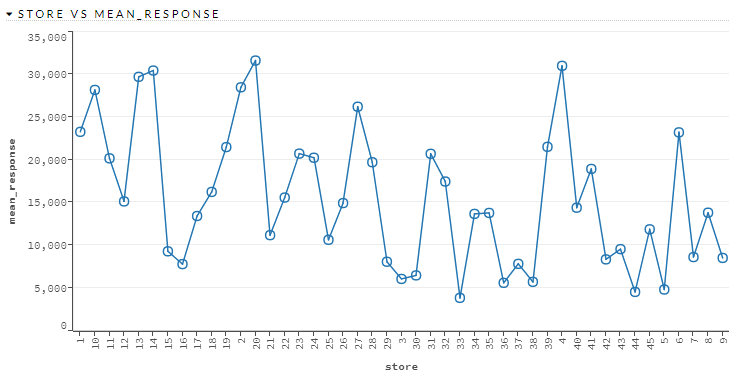
From the above metrices, we can observe the R2 of training, validation and cross validation models , that bits increasing gradually and the mean of R2 is 91% which tells us how fairly predictable the model is. Moreover, the all the models have their R2 in the same range and are not highly varying, which shows the model is not over fitted.

Now, the figures below demonstrate the variation in the weekly sales per department(dept.), store and date, through which we can find out which department or store is having maximum, minimum sales or the desired sales.

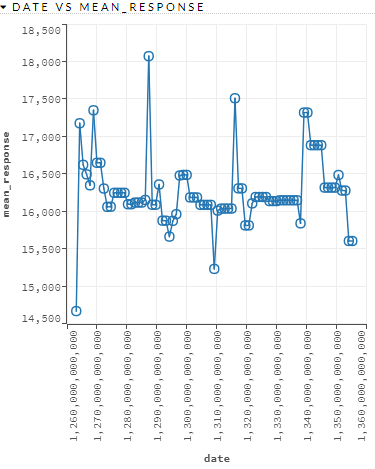
By Department :



By Store :



By Date :



We see the weekly sales status is highly driven department-wise and store-wise, which says that the sales is dependent on the people's choices and demand and region where the stores are located, for which the sales is driven department-wise and store-wise.