**WALMART STORE PREDICTION – FORECASTING**

Data: train merged and test merged

Data shape: 421570 x 16

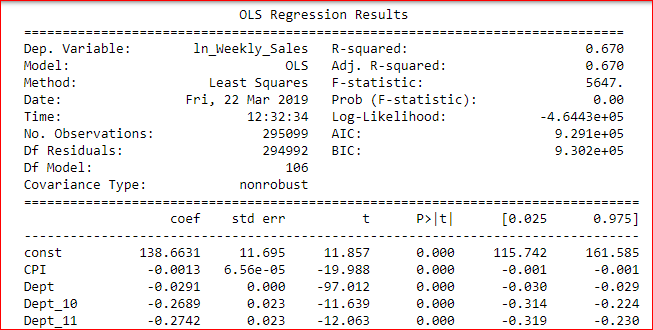
Objective:

To predict the store sales using the historic markdown data.

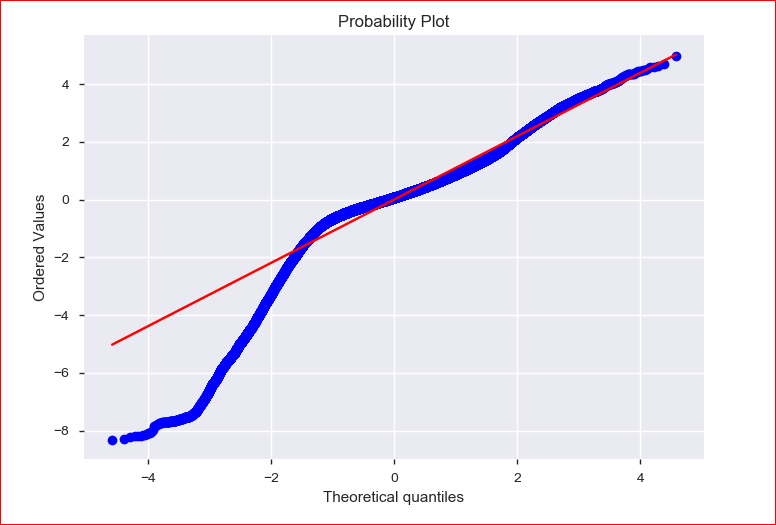
Steps to be followed:

* We are provided with the historic sales data for 45 Wal-Mart stores located in different regions. Each store has number of departments. You are tasked to predict the department – wise sales for each store.
* The problem is solved using the approach to proceed with Linear Regression.
* After importing the data, begin with the creation of derived variables that are provided in hints sheet.
* Start with the basic Data Pre-processing steps. Apply the Linear Regression model.
* To forecast the sales, we can also apply SARIMAX model.
* Import the clean data with only X and Y variables. Perform some basic pre-processing steps.
* Consider some random values for p, d, q. Then find the optimum pair of (p, d, q) and build the best model.

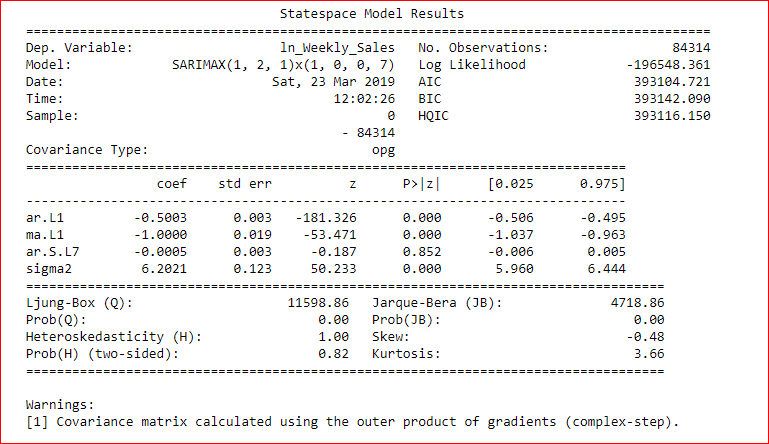
The below snip-it shows the OLS regression results in Linear Regression model. The R2 & Adj. R2 value turned out to be same – 67.



The dispersion of data around the regression line is shown in the Q – Q plot. The data is very much dispersed away from the line which shows the data is not linear.



The snip-it shows the result of the SARIMAX model when applied with the random values of (p, d, q),(P, D, Q, S).



The below snip-it shows the best SARIMAX model with the best values of (p,d,q).

