```
/** Import an XLSX file.  **/

PROC IMPORT DATAFILE="/folders/myshortcuts/sas_university_edition/my folder/lesson2/Retail_Analysis.xlsx"
            OUT=WORK.Retail_Analysis
            DBMS=XLSX
            REPLACE;
RUN;

/** Print the results. **/

PROC PRINT DATA=WORK.Retail_Analysis; RUN;



/* Descriptive Statistics for the Retail_Analysis */

proc means data = Retail_Analysis N mean mode median max min p25 p50 p75;

             var Sales Profit Discount Shipping_Cost;
run;



/* Correlation Analysis for Retail_Analysis*/

ODS graphics on;
proc corr data = Retail_Analysis plot= matrix(Histogram) ;

            var Sales Profit Discount Shipping_Cost;
run;
ODS Graphics off;

/* linear correlation between Sales and Profit is strong */
/* Pearson cofficient value for sales and profit is 0.89 its nearby 1 means relationship between sales and profit is perfact */



/* Significance of independent variable */

proc reg data = Retail_Analysis;

       model Sales = Quantity Profit Discount;
  run;

 /* This model provide a strong relationship between Response and explanetry variable with high accuracy (r^2 = 86% and adj r^2 = 85% ) */
 /* The p-value for the profit and quantity is less than 0.05 than variable are found to be significant */



/* Creating a new variable Total_Sales using Sales of the Product and Quantity */

data Retail_Analysis;
       set Retail_Analysis;
        Total_Sales = Sales*Quantity;
run;



/* regression Analysis(Prediction) using Total_Sales */

proc reg data = Retail_Analysis;

       model Total_Sales = Quantity Profit;
  run;

/* Total_Sales give us better idea about r^2 value(88%) and adj r^2(87.8%) means good accuracy */



/* Creating a new dataset with log, exponential, square and cube of the variable */

data Retail_NewAnalysis;
            set Retail_Analysis;
            Sales_log = log10(Sales); Quantity_log = log10(Quantity); Profit_log = log10(Profit); Discount_log = log10(Discount); ShippingCost_log = log10(Shipping_Cost);TotalSales_log = log10(Total_Sales);
            Sales_exp = exp(Sales); Quantity_exp = exp(Quantity); Profit_exp = exp(Profit); Discount_exp = exp(Discount); ShippingCost_exp = exp(Shipping_Cost);TotalSales_exp = exp(Total_Sales);
            Sales_sqr = Sales**2; Quantity_sqr = Quantity**2; Profit_sqr = Profit**2; Discount_sqr = Discount**2; ShippingCost_sqr = Shipping_Cost**2;TotalSales_sqr = Total_Sales**2;
            Sales_cube = Sales**3; Quantity_cube = Quantity**3; Profit_cube = Profit**3; Discount_cube = Discount**3; ShippingCost_cube = Shipping_Cost**3;TotalSales_cube = Total_Sales**3;
run;



/*  Regression test of the new table */

/* Regression test for log variable */

proc reg data = Retail_NewAnalysis;

       model TotalSales_log = Quantity_log Profit_log Discount_log;
  run;

 /* The r^2 value is 84.8% which tells that the variation between TotalSales_log and and variable is pretty strong */
/* The P-value for the for the Profit is less than 5%  and therefore the variable Profit is significant at 95% confidence level*/



/* Regression test for sqr variable */

proc reg data = Retail_NewAnalysis;

       model TotalSales_sqr = Quantity_sqr Profit_sqr Discount_sqr;
  run;

/* The r^2 value(74%)  of sqr is less compare to regression analysis of log  */
/* for cube variable  gives the value of r^2 and adj r^2 is below 60% */
```