/\* Retail Analysis \*/

/\* Step 1: Import the data \*/

/\* Generated Code (IMPORT) \*/

/\* Source File: Project Retail Analysis\_Dataset.xlsx \*/

/\* Source Path: /folders/myshortcuts/myFolders \*/

%web\_drop\_table(TRAINING.RetailAnalysis);

FILENAME REFFILE '/folders/myshortcuts/myFolders/Project 04\_Retail Analysis\_Dataset.xlsx';

PROC IMPORT DATAFILE=REFFILE

DBMS=XLSX

OUT=TRAINING.RetailAnalysis;

GETNAMES=YES;

RUN;

PROC CONTENTS DATA=TRAINING.RetailAnalysis; RUN;

/\* Step 2: Compute the Total Sales \*/

data Training.RetailAnalysis;

set Training.RetailAnalysis;

Total\_Sales = sales\*quantity;

run;

/\* Step 3: Model Total\_Sales (Dependent variable) against the other independent variables \*/

PROC REG DATA=training.retailanalysis;

MODEL Total\_Sales= Quantity Discount Profit Shipping\_Cost;

run;

/\* Since the Shipping cost variable is a multiple(0.1) of Profit variable, it can be dropped from the model \*/

PROC REG DATA=training.retailanalysis;

MODEL Total\_Sales= Quantity Discount Profit;

run;

/\* Discount variable is insignificant. Hence can be omitted from the model \*/

PROC REG DATA=training.retailanalysis;

MODEL Total\_Sales= Quantity Profit;

run;

/\* From the above model we can conclude that \*/

/\* 1) The independent variables Quantity & Profit have p-values < 0.05, hence both are significant \*/

/\* 2) The model is 87.81% accurate (adjusted R-square value). \*/

/\* 3) The linear equation is Total\_Sales = -333.042 + 166.88 \* Quantity + 4.1 \* Profit \*/

/\* From the above equation we can predict Total\_Sales values for given Quantity & Profit \*/

/\* Step 4: Checking the predicted values in the output dataset \*/

PROC REG DATA=training.retailanalysis;

MODEL Total\_Sales= Quantity Profit;

output out = PredictedSales p=Total\_Sales\_Predicted;

run;

/\* From the above output predicted value, we can notice that Product2 has -ve predictions \*/

/\* Running the model only for Product2 \*/

PROC REG DATA=training.retailanalysis;

MODEL Total\_Sales= Quantity Profit;

output out = PredictedSales p=Total\_Sales\_Predicted;

where Products = 'Product2';

run;

/\* We find that all the variables are insignificant and the accuracy is 56.65% \*/

/\* Running the model only for Product1 \*/

PROC REG DATA=training.retailanalysis;

MODEL Total\_Sales= Quantity Profit;

output out = PredictedSales p=Total\_Sales\_Predicted;

where Products = 'Product1';

run;

/\* We find that all the variables are insignificant and the accuracy is 42.46% \*/

/\* Running the model only for Product3 \*/

PROC REG DATA=training.retailanalysis;

MODEL Total\_Sales= Quantity Profit;

output out = PredictedSales p=Total\_Sales\_Predicted;

where Products = 'Product3';

run;

/\* We find that all the variables are significant and the accuracy is 100% \*/

/\* Running the model only for Product4 \*/

PROC REG DATA=training.retailanalysis;

MODEL Total\_Sales= Quantity Profit;

output out = PredictedSales p=Total\_Sales\_Predicted;

where Products = 'Product4';

run;

/\* We find that one of the variables is significant and the other is insignificant but the accuracy is 100% \*/

/\* Running the model only for Product5 \*/

PROC REG DATA=training.retailanalysis;

MODEL Total\_Sales= Quantity Profit;

output out = PredictedSales p=Total\_Sales\_Predicted;

where Products = 'Product5';

run;

/\* We find that we cannot say if the variables are significant or not but the accuracy is 100% \*/

/\* Running the model only for Product6 \*/

PROC REG DATA=training.retailanalysis;

MODEL Total\_Sales= Quantity Profit;

output out = PredictedSales p=Total\_Sales\_Predicted;

where Products = 'Product6';

run;

/\* We find that we cannot say if the variables are significant or not but the accuracy is 100% \*/

/\* Running the model only for Product7 \*/

PROC REG DATA=training.retailanalysis;

MODEL Total\_Sales= Quantity Profit;

output out = PredictedSales p=Total\_Sales\_Predicted;

where Products = 'Product7';

run;

/\* We find that we cannot say if the variables are significant or not but the accuracy is 100% \*/

/\* Running the model only for Product8 \*/

PROC REG DATA=training.retailanalysis;

MODEL Total\_Sales= Quantity Profit;

output out = PredictedSales p=Total\_Sales\_Predicted;

where Products = 'Product8';

run;

/\* We find that we cannot say if the variables are significant or not but the accuracy is 100% \*/

/\* Running the model only for Product9 \*/

PROC REG DATA=training.retailanalysis;

MODEL Total\_Sales= Quantity Profit;

output out = PredictedSales p=Total\_Sales\_Predicted;

where Products = 'Product9';

run;

/\* We find that we cannot say if the variables are significant or not but the accuracy is 100% \*/