****

Programming for analytics

(BAN 130)

**PROJECT REPORT**

**on**

**Adventure Works Product Sales Analysis**

****

**GROUP 8**

|  |  |  |
| --- | --- | --- |
| Student Name | Student ID | Task |
| Aaron Gonsalves | 161288196 | Project Report |
| Goushalya Kothai Nachiar Ayyappan | 135058204 | Data Analysis |
| Parin Mahajan | 142504208 | Data Cleaning - 2 |
| Sambhav Sharma | 151963204 | Data Import & Data Cleaning - 1 |
| Sidek Hussain Abbas Hussain | 156384208 | Joining and Merging |

Contents

[Part 1: Data Import 2](#_Toc78647396)

[Part 2: Data Cleaning 7](#_Toc78647397)

[2.1 Product\_Clean 7](#_Toc78647398)

[2.1.1 Creating Product\_Clean dataset from Product Dataset 7](#_Toc78647399)

[2.1.2 Dealing with Missing Values 9](#_Toc78647400)

[2.1.3 Changing type of column ListPrice and formatting it. 10](#_Toc78647401)

[2.2 SalesOrderDetail\_Clean 11](#_Toc78647402)

[2.2.1 Creating SalesOrderDetail\_Clean from SalesOrderDetail 11](#_Toc78647403)

[2.2.2 Performing type conversions for various columns 13](#_Toc78647404)

[2.2.3 Considering dates for years 2013 and 2014 with the given format 15](#_Toc78647405)

[Part 3: Joining and Merging 16](#_Toc78647406)

[3.1 SalesDetails 16](#_Toc78647407)

[3.1.1 Creating SalesDetails dataset by joining datasets made in Part 2 16](#_Toc78647408)

[3.2 SalesAnalysis 17](#_Toc78647409)

[3.2.1 Creating SalesAnalysis dataset from SalesDetails dataset 17](#_Toc78647410)

[Part 4. Data Analysis 18](#_Toc78647411)

[4.1 How many Red color Helmets are sold in 2013 and 2014? 18](#_Toc78647412)

[4.2 How many items sold in 2013 and 2014 have a Multi color? 19](#_Toc78647413)

[4.3 What is the combined Sales total for all the helmets sold in 2013 and 2014? 20](#_Toc78647414)

[4.4 How many Yellow Color Touring-1000 where sold in 2013 and 2014? 21](#_Toc78647415)

[4.5 What was the total sales in 2013 and 2014? 22](#_Toc78647416)

[4.5.1 Approach 1 22](#_Toc78647417)

[4.5.2 Approach 2 23](#_Toc78647418)

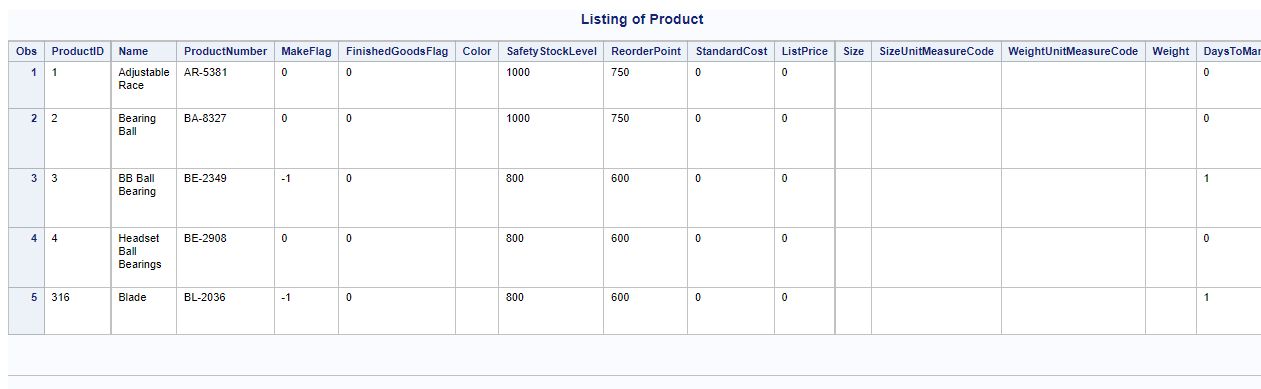
[Part 5. Creating a Chart 24](#_Toc78647419)

[Part 6. Program Code 25](#_Toc78647420)

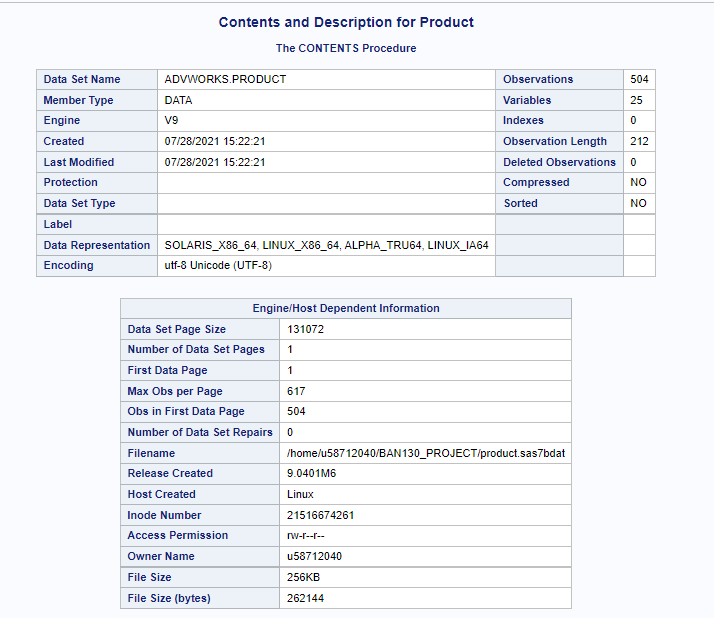
# Part 1: Data Import



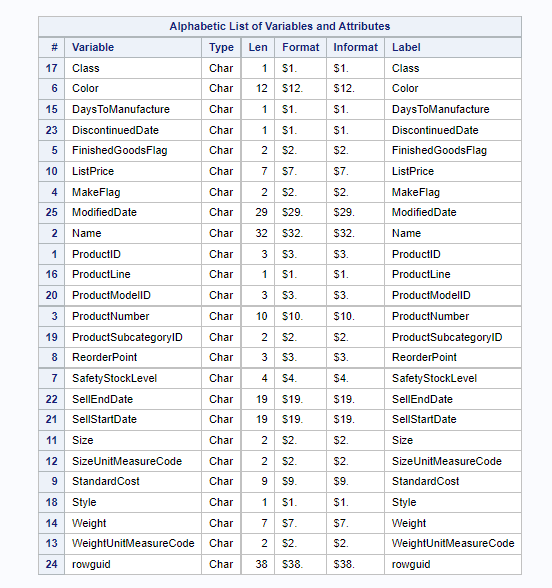
* Firstly, a library reference named ‘AdWorks’ was assigned by using the libname statement, which is used in further phases to store datasets permanently.
* In order to import the required data from the given excel file, ‘AdventureWorks.xlsx’, the procedure, ‘proc import’ was used.
* As there are 4 sheets in the ‘AdventureWorks’ excel workbook, and we require only 2 sheets namely: Product and SalesOrderDetail from them, ‘proc import’ was used twice to import these sheets.
* In the above picture, the sheet, ‘Product’ has been imported into the dataset, ‘AdvWorks.Product’.
* After importing, 5 observations from the dataset, ‘AdvWorks.Product’ were been printed using ‘proc print’.
* Also, a ‘proc contents’ was performed to obtain the description of ‘AdvWorks.Product’ dataset.
* The general details found from this were:
  + There are 504 observations.
  + There are 25 variables and all of them are character variables.



1. Printing 5 records from AdvWorks.Product

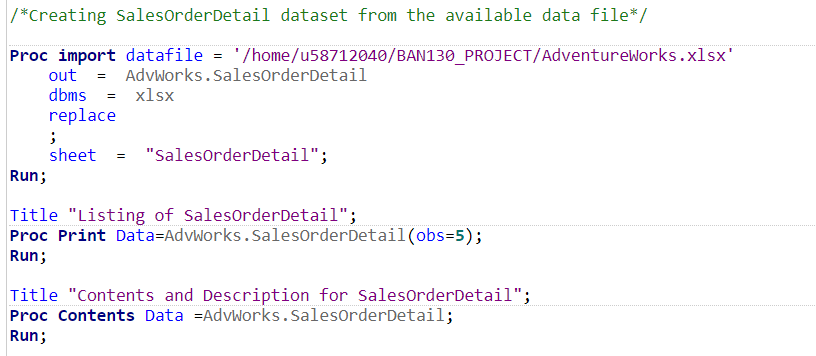


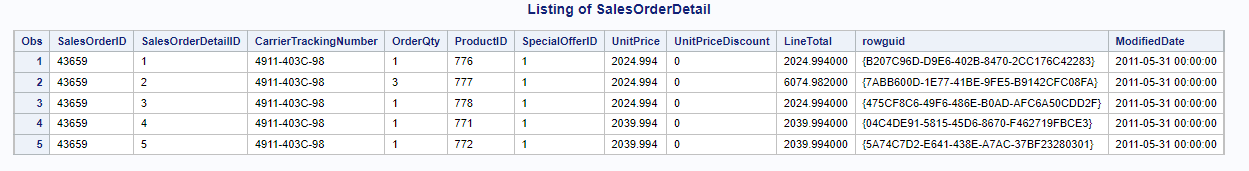
(b) Getting summary of AdvWorks.Product -1



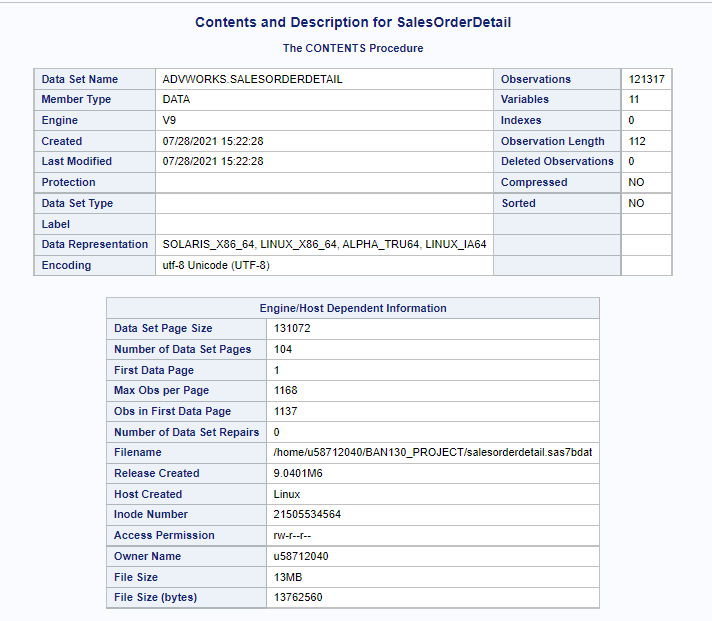
(c) Getting summary of AdvWorks.Product -2

* Similarly, as stated above, another ‘proc import’ was also applied to import the sheet, ‘SalesOrderDetail’ into AdvWorks.SalesOrderDetail.
* And accordingly, 5 observations from it were printed using ‘proc print’ and the general summary of the dataset was obtained using ‘proc contents’.
* The general details found from this were:
  + There are 121317 observations.
  + There are 11 variables and all of them are character variables.

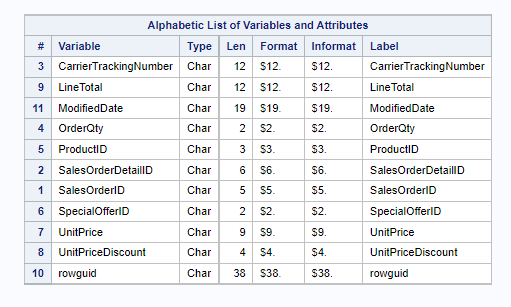




1. Printing 5 records from AdvWorks.SalesOrderDetail



(b) Getting summary of AdvWorks.SalesOrderDetail -1

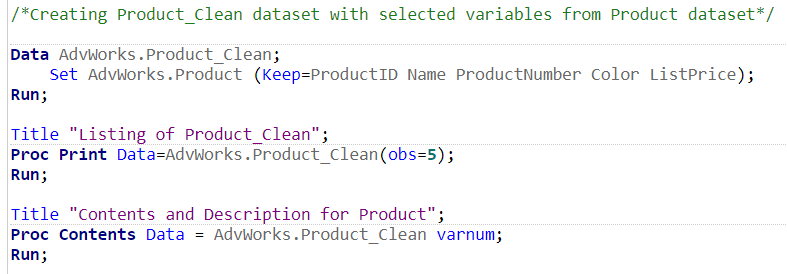


1. Getting summary of AdvWorks.SalesOrderDetail -1

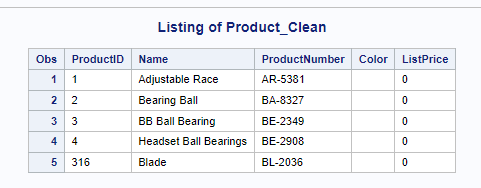
# Part 2: Data Cleaning

## Product\_Clean

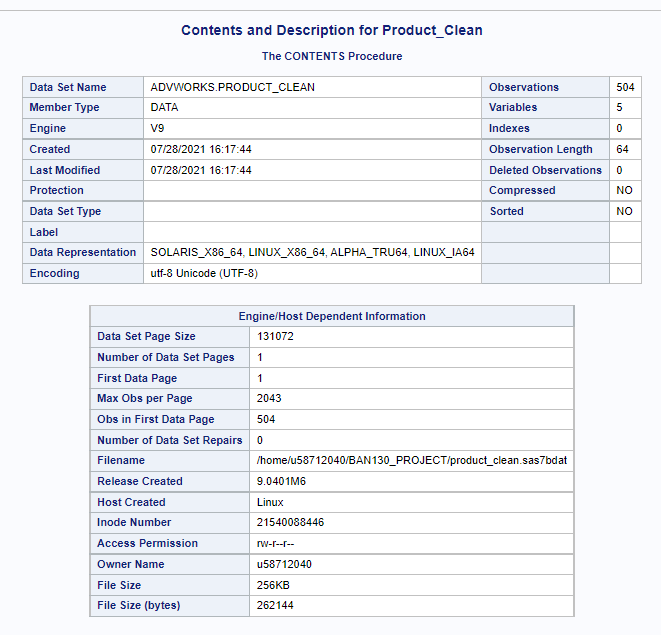
### 2.1.1 Creating Product\_Clean dataset from Product Dataset



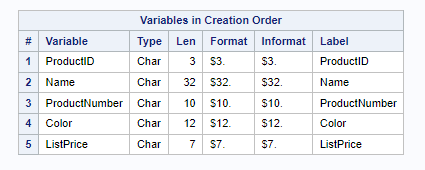
* + For the next step, a new dataset, ‘AdvWorks.Product\_Clean’ is created by reading observations from AdvWorks.Product with the use of ‘set’ statement.
  + In this step, only selected variables ProductID, Name, ProductNumber, Color and ListPrice are written to the ‘AdvWorks.Product\_Clean’ dataset with the help of ‘keep’ statement.
  + After this, 5 observations from this dataset are printed using ‘proc print’ along with the general summary of the dataset using ‘proc contents’.



1. Printing 5 records from AdvWorks.Product\_Clean

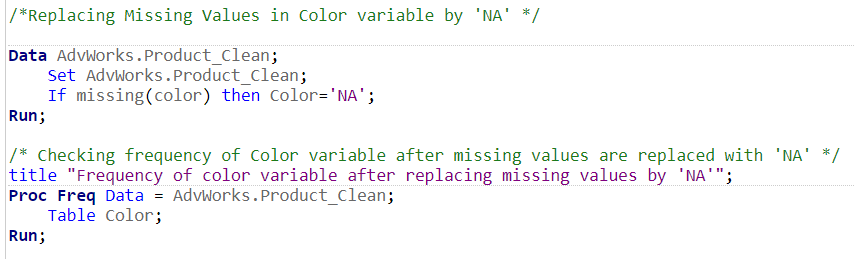


1. Getting summary of AdvWorks.Product\_Clean -1



1. Getting summary of AdvWorks.Product\_Clean -2

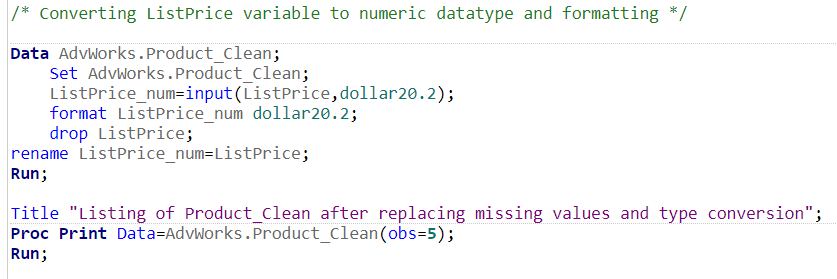
### Dealing with Missing Values



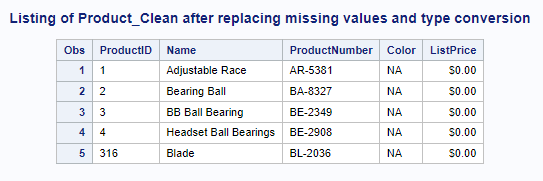
* + In the next step, any missing values in the ‘Color’ variable are replaced by ‘NA’ with the use of conditional statements.
  + A ‘proc freq’ step is performed after this, to check the change in frequency of ‘Color’ variable.



### Changing type of column ListPrice and formatting it.

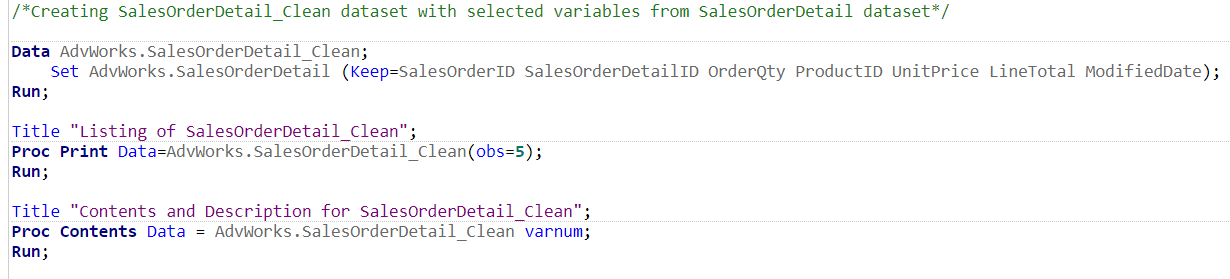


* + Here, a new variable ListPrice\_num is created by converting the existing variable ListPrice into numeric and storing its value in the new variable, which is done with the help of ‘input’ function.
  + Informat ‘dollar20.2’ is applied to this variable.
  + The old variable ‘ListPrice’ is dropped by using ‘drop’ statement.
  + After that the new variable ‘ListPrice\_num’ is renamed to ListPrice with the help of ‘rename’ statement.
  + And then 5 observations were printed, whose output is shown below.

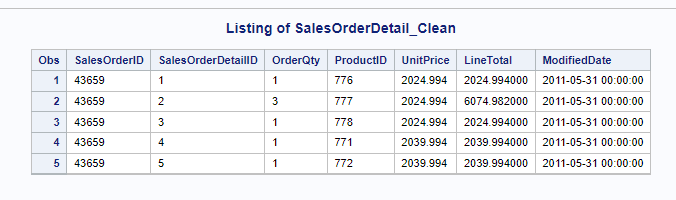


## SalesOrderDetail\_Clean

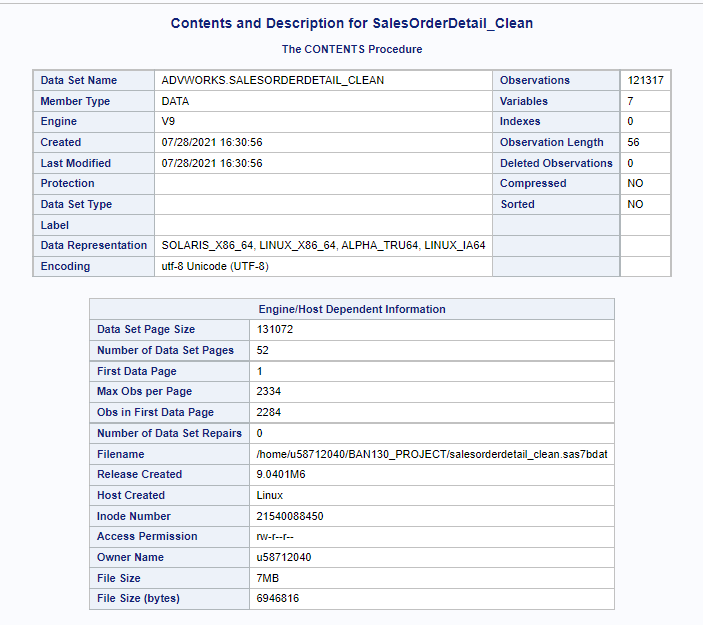
### Creating SalesOrderDetail\_Clean from SalesOrderDetail



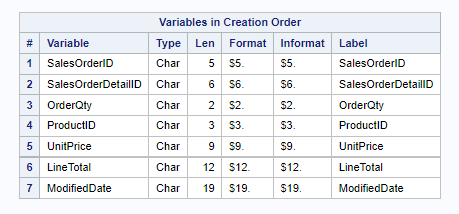
* + Again, a new dataset, ‘AdvWorks.SalesOrderDetail\_Clean’ is created by reading observations from AdvWorks.SalesOrderDetail with the use of ‘set’ statement.
  + In this step, only selected variables SalesOrderID, SalesOrderDetailID, OrderQty, ProductID, UnitPrice, LineTotal and ModifiedDate are written to ‘AdvWorks.SalesOrderDetail\_Clean’ with the help of a ‘keep’ statement.
  + After this, 5 observations from this dataset are printed using ‘proc print’ along with the general summary of the dataset using ‘proc contents’.



1. Printing 5 records from SalesOrderDetail\_Clean

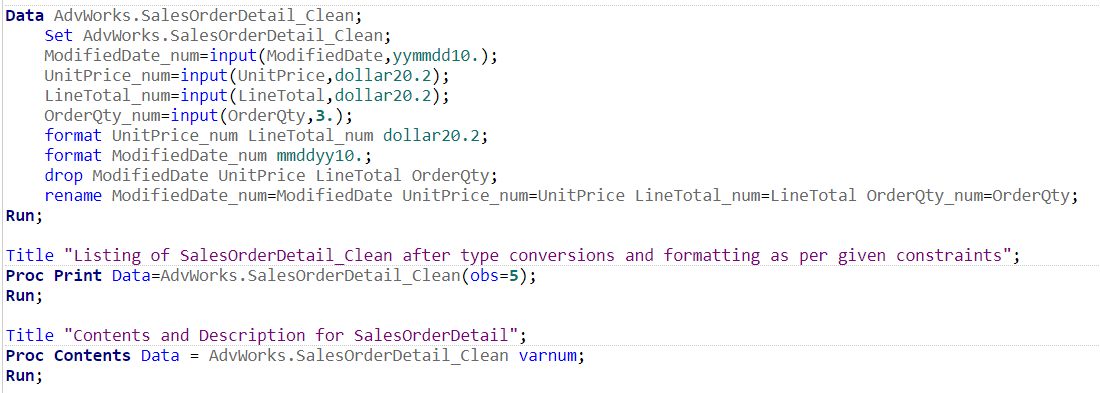


1. Getting summary of AdvWorks.SalesOrderDetail\_Clean -1

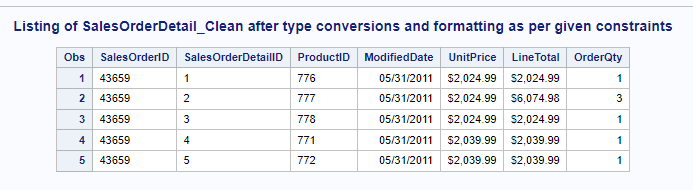


1. Getting summary of AdvWorks.SalesOrderDetail\_Clean -2

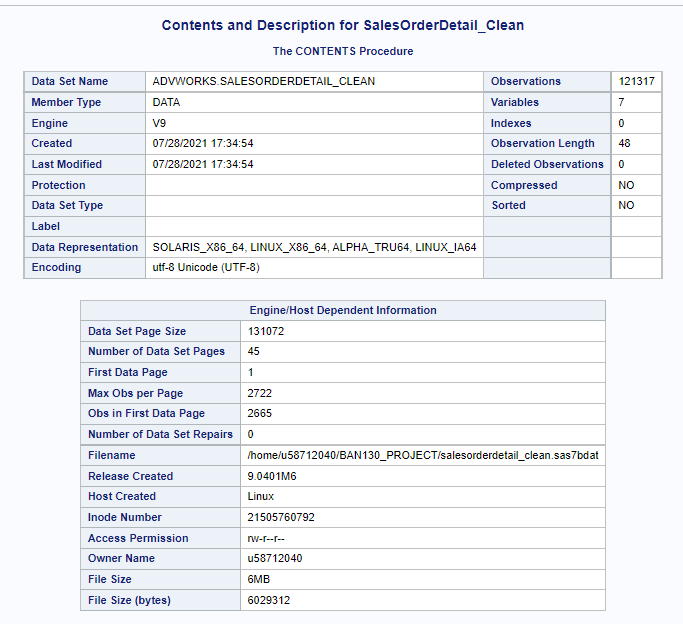
### Performing type conversions for various columns



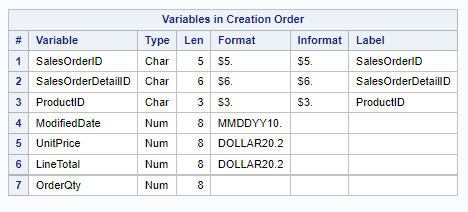
* + For the next step, various columns like ModifiedDate\_num, UnitPrice\_Num, LineTotal\_num and OrderQty\_num were created by converting the existing variables into numeric and storing their values in their respective new variables, which was done using ‘input’ function.
  + Informat ‘dollar20.2’ was used for the variables UnitPrice\_Num and LineTotal\_num.
  + The old variables ModifiedDate, UnitPrice, LineTotal and OrderQty were dropped by using the ‘drop’ statement.
  + After this, the new variables ModifiedDate\_num, UnitPrice\_Num, LineTotal\_num and OrderQty\_num were renamed to ModifiedDate, UnitPrice, LineTotal and OrderQty by using the ‘rename’ statement.
  + And then 5 observations were printed, along with ‘proc contents’ to show the reflected changes in type of variables.



1. Printing 5 records from SalesOrderDetail\_Clean

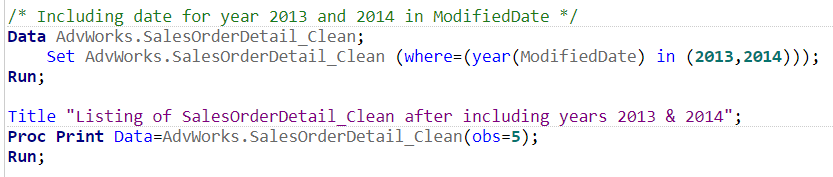


1. Showing general details of SalesOrderDetail\_Clean - 1

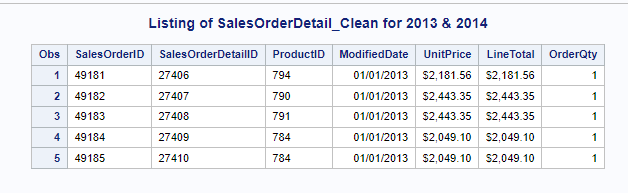


1. Showing general details of SalesOrderDetail\_Clean – 2

### Considering dates for years 2013 and 2014 with the given format



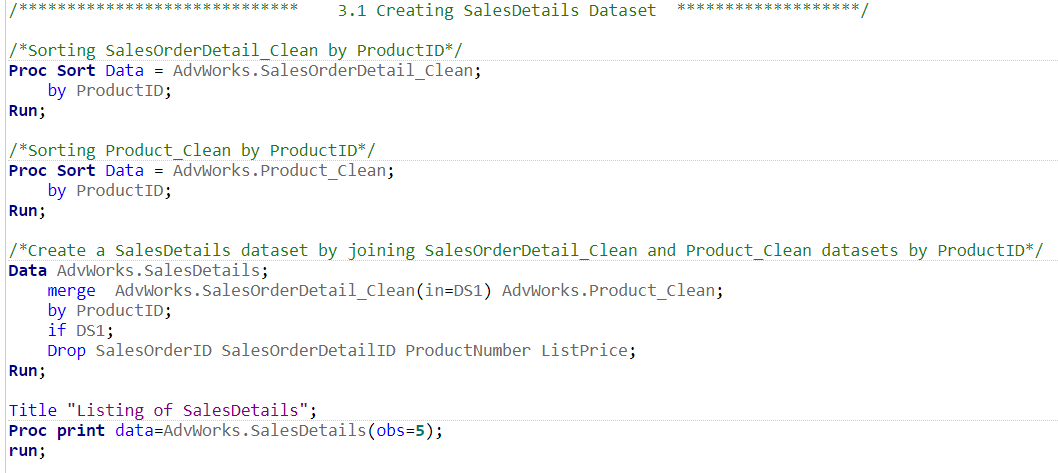
* + A new dataset is created which is also called ‘AdvWorks.SalesOrderDetail\_Clean’ whose values are read in by ‘AdvWorks.SalesOrderDetail\_Clean’ with the use of ‘set’ statement.
  + Here, using ‘where’ clause, and ‘year’ function on ModifiedDate, the values where year is either 2013 or 2014 are taken.
  + The ‘mmddyy10.’ format was already applied in the previous step.
  + After this, 5 observations were printed.



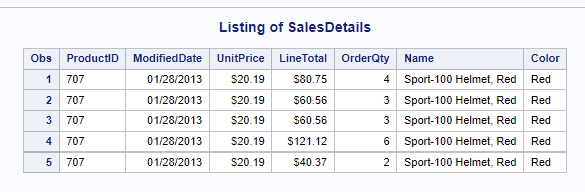
# Part 3: Joining and Merging

## SalesDetails

## Creating SalesDetails dataset by joining datasets made in Part 2

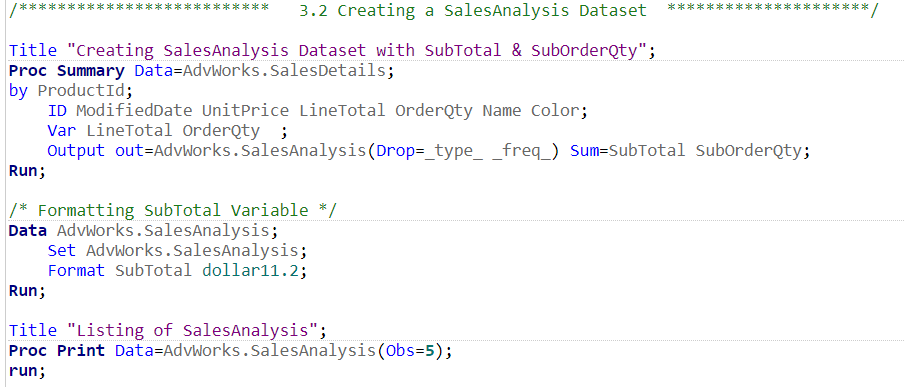


* Before joining datasets, both datasets were sorted as it is a pre-requirement.
* A new dataset, ‘SalesDetails’ was created by joining the datasets, ‘Product\_Clean’ and ‘SalesOrderDetail\_Clean’ with the use of ‘merge’ statement, on ProductID which exists in both datasets.
* The columns for SalesOrderID, SalesOrderDetailID, ProductNumber and ListPrice were dropped with the help of ‘drop’ statement.
* 5 observations of the new created dataset were printed.

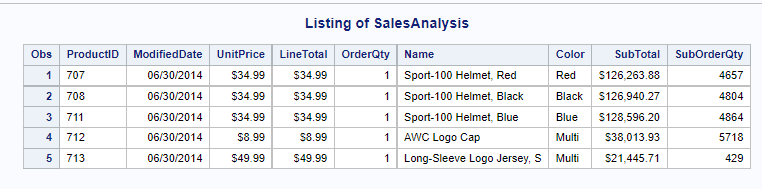


## SalesAnalysis

## Creating SalesAnalysis dataset from SalesDetails dataset

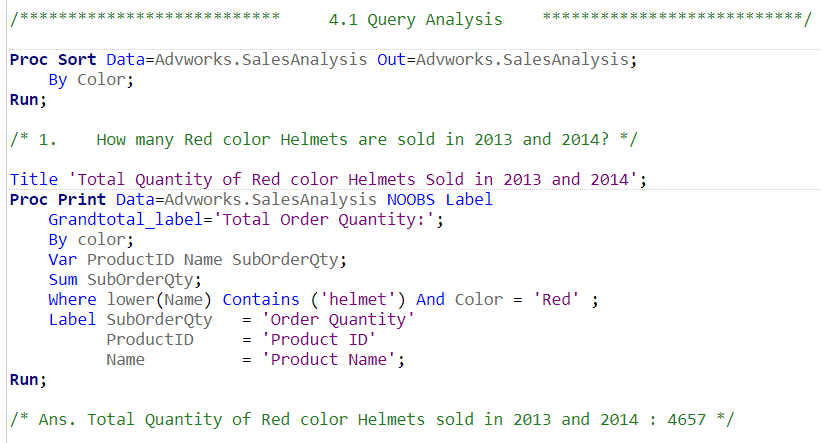


* ‘proc summary’ is used to calculate descriptive statistics for LineTotal and OrderQty
* Also, other variables are given in ID as identification variables for more data and better visuals.
* The columns \_type\_ and \_freq\_ are dropped as they come as extras from the output of ‘proc summary’.
* Using the results of ‘proc summary’ another dataset was created which was named the same and given a format of ‘dollar11.2’ for variable Subtotal.
* 5 observations were printed from the new dataset.



# Part 4. Data Analysis

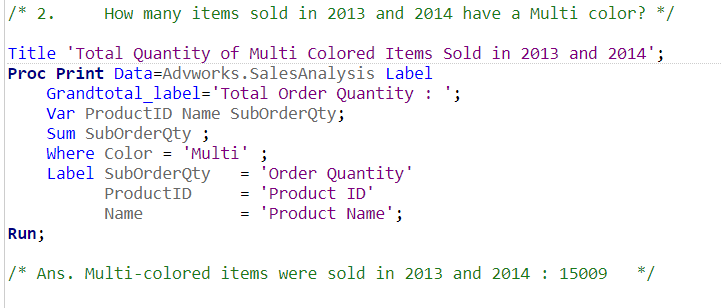
## 4.1 How many Red color Helmets are sold in 2013 and 2014?



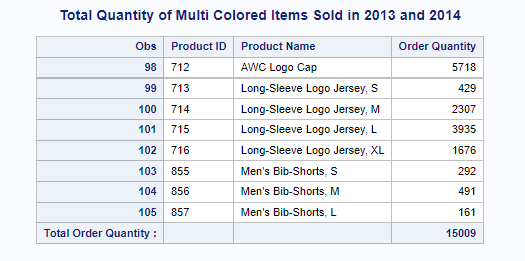
* Firstly, the dataset, ‘SalesAnalysis’ is sorted by the variable ‘color’.
* After that, using ‘proc print’, the required data is obtained showing the no. of ‘Red’ color Helmets that are sold in 2013 and 2014.
* For this, the output data is arranged by ‘color’.
* Using ‘var’ statement, the columns are arranged as shown in output.
* ‘sum’ function is used on SubOrderQty to get the summation of it.
* And a ‘where’ clause is applied in order to find the answer to the given question, which in this case is to look for red color helmets. ‘contains’ and ‘and’ operators are also used here, to check for the given string and condition.
* The ‘Grandtotal\_label’ is also given a label which is displayed in the last line of output.
* The printed variables were also given appropriate labels.
  + **4657 red helmets were sold in 2013 and 2014.**



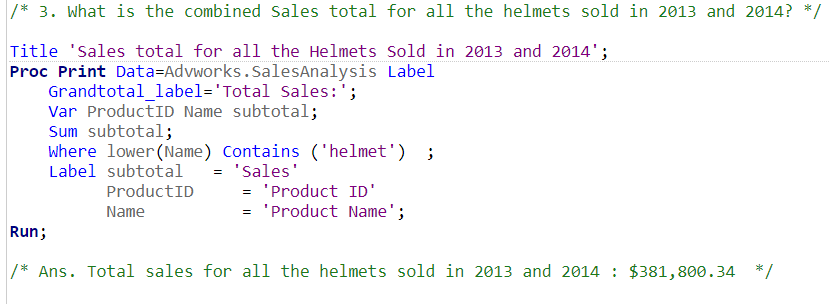
## 4.2 How many items sold in 2013 and 2014 have a Multi color?



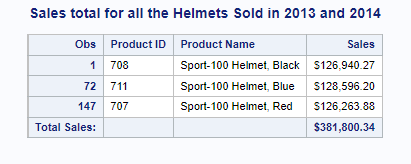
* Using ‘proc print’ on Advworks.SalesAnalysis, the required results for the question were obtained.
* Using ‘var’ statement, the columns are arranged as shown in output.
* ‘sum’ function is used on SubOrderQty to get the summation of it.
* And a ‘where’ clause is applied in order to find the answer to the given question. Which in this case is to look for items where the color has value of ‘Multi’.
* The ‘Grandtotal\_label’ is also given a label which is displayed in the last line of output.
* The printed variables were also given appropriate labels.
* **15009 items were sold in 2013 and 2014 where color had value ‘Multi’.**



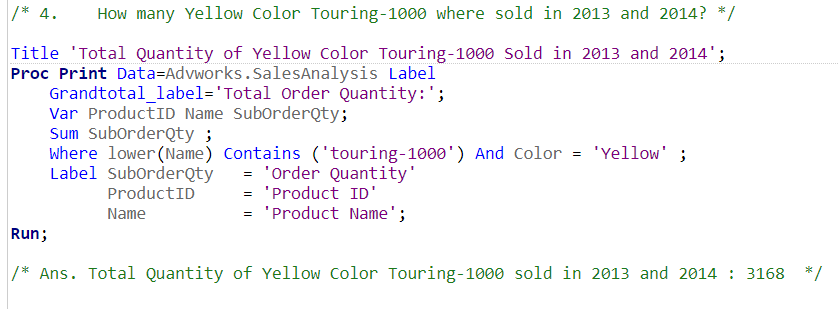
## 4.3 What is the combined Sales total for all the helmets sold in 2013 and 2014?



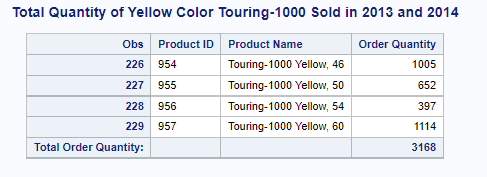
* Using ‘proc print’ on Advworks.SalesAnalysis, the answer to the question asked was found.
* Using ‘var’ statement, the columns are arranged as shown in output.
* ‘sum’ function is used on Subtotal to get the summation of it.
* And a ‘where’ clause is applied in order to find the answer to the given question, which in this case is to look for all product names which contain the word ‘helmet’ in them. For this, ‘contains’ and ‘and’ operators are used here, to check for the given string and condition.
* The ‘Grandtotal\_label’ is also given a label which is displayed in the last line of output.
* The printed variables were also given appropriate labels.
* **The combined sales total for all helmets sold in 2013 and 2014 are $381,800.34**



## 4.4 How many Yellow Color Touring-1000 where sold in 2013 and 2014?



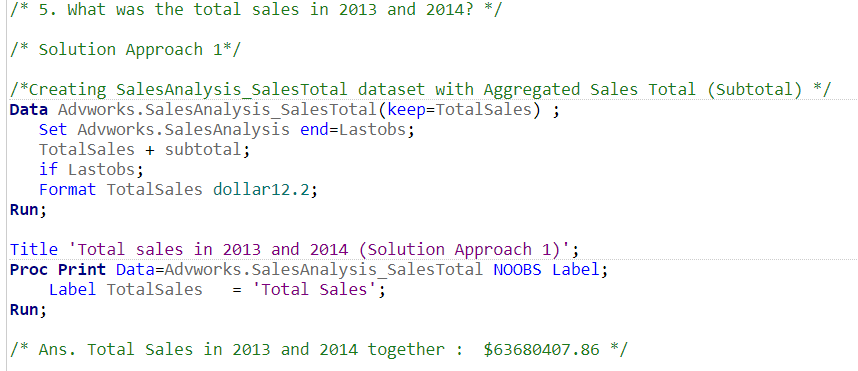
* Using ‘proc print’ on ‘Advworks.SalesAnalysis’ we get the answer to the question asked.
* Using ‘var’ statement, the columns are arranged as shown in output.
* ‘sum’ function is used on SubOrderQty to get the summation of it.
* And a ‘where’ clause is applied in order to find the answer to the given question, which in this case is to look for all product names which contain the word ‘touring-1000’ and whose color is ‘yellow’ in them. For this, ‘contains’ and ‘and’ operators are used here, to check for the given string and condition.
* The ‘Grandtotal\_label’ is also given a label which is displayed in the last line of output.
* The printed variables were also given appropriate labels.
* **The number of Yellow Color Touring-1000 sold in 2013 and 2014 are 3168**



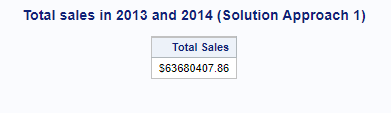
## 4.5 What was the total sales in 2013 and 2014?

* To answer this question, two approaches were considered.

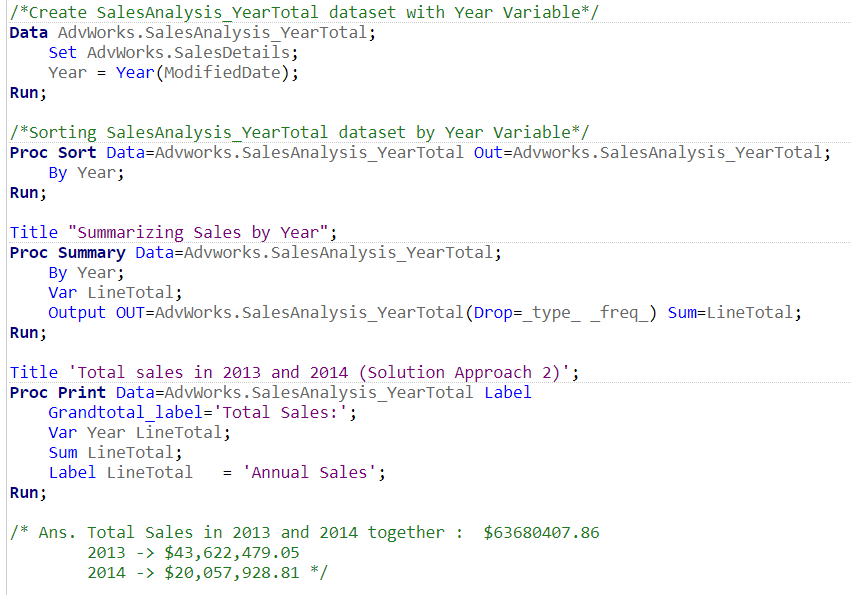
### 4.5.1 Approach 1



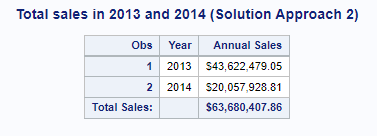
* The snippet shown above displays the first approach.
* Here, a new dataset named ‘Advworks.SalesAnalysis\_SalesTotal‘ is created by keeping variable ‘TotalSales’ from the Advworks.SalesAnalysis dataset.
* Also, using conditional statement ‘if’ for LastObs i.e. for the last observation value, a sum statement has been calculated which returns the Total Sales that are made in years 2013 and 2014.
* Proper label is applied to TotalSales in the ‘proc print’ step which gives the below output.
* **The total sales made in 2013 and 2014 was $63680407.86**



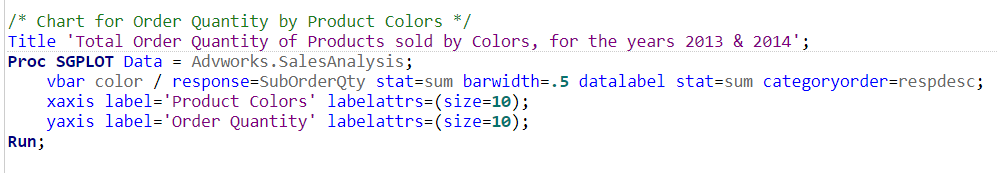
### 4.5.2 Approach 2



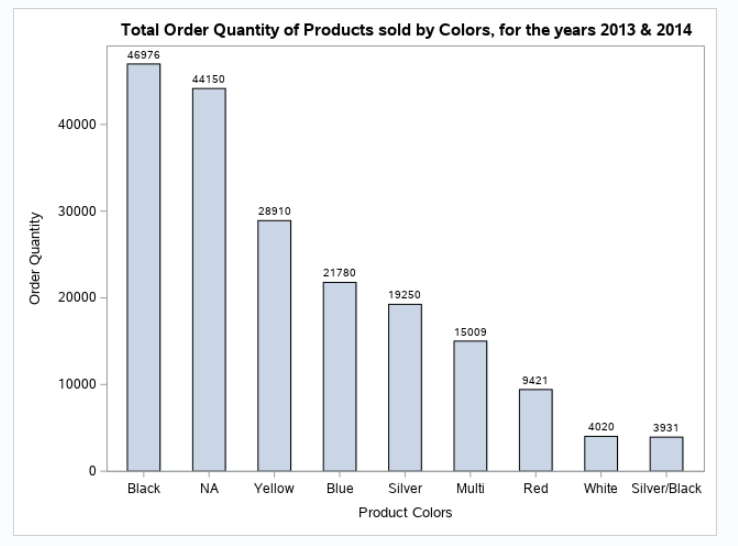
* Here, a new dataset is created called ‘AdvWorks.SalesAnalysis\_YearTotal’ whose data is read from ‘AdvWorks.SalesDetails’.
* Also, a new variable ‘Year’ was created by applying the function ‘Year’ on ModifiedDate.
* After this, ‘AdvWorks.SalesAnalysis\_YearTotal’ is sorted by ‘year’ variable.
* ‘proc summary’ is used here to get the sum of ‘LineTotal’ and is outputted in ‘AdvWorks.SalesAnalysis\_YearTotal’.
* Appropriate label is also given to LineTotal.
* The total sales made in 2013 were: $43,622,479.05
* The total sales made in 2014 were: $20,057,928.81
* **The sum total of sales made in 2013 and 2014 is: $63,680,407.86**



# Part 5. Creating a Chart



* Using ‘proc sgplot’ on AdvWorks.SalesAnalysis, a bar graph was plotted.
* The x-axis represents the ‘Product Colors’.
* The y-axis represents the ‘Order Quantity’.



* Here, the bar chart displays the quantity of orders that were sold for color of each product.
* It is arranged in descending order for the viewer to get an idea of which colored products have the highest no. of sales and which have the lowest.
* The exact value for no. of products sold for each color is shown on the top of each bar.
* The no. of ‘Multi’ colored products sold in 2013 and 2014 were 15009 which was already answered in Part 4, section 4.2.
* Using this chart, any viewer can easily tell how much quantity of products were sold in the years 2013 and 2014 based on their color.

# Part 6. Program Code

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*NCC BAN130 Project: Group 8

Project Name: Adventure Works Product Sales Analysis

Team Members: Aaron, Goushalya,Parin, Sambhav, Sidek

\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* STEP 1 - Data Import \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

libname AdvWorks '/home/u58712040/BAN130\_PROJECT';

/\*Creating Product dataset from the available data file\*/

Proc Import datafile = '/home/u58712040/BAN130\_PROJECT/AdventureWorks.xlsx'

out = AdvWorks.Product

dbms = xlsx

replace

;

sheet = "Product";

Run;

Title "Listing of Product";

Proc Print Data=AdvWorks.Product(obs=5);

Run;

Title "Contents and Description for Product";

Proc Contents Data = AdvWorks.Product;

Run;

/\*Creating SalesOrderDetail dataset from the available data file\*/

Proc import datafile = '/home/u58712040/BAN130\_PROJECT/AdventureWorks.xlsx'

out = AdvWorks.SalesOrderDetail

dbms = xlsx

replace

;

sheet = "SalesOrderDetail";

Run;

Title "Listing of SalesOrderDetail";

Proc Print Data=AdvWorks.SalesOrderDetail(obs=5);

Run;

Title "Contents and Description for SalesOrderDetail";

Proc Contents Data =AdvWorks.SalesOrderDetail;

Run;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* STEP 2 - Data Cleaning \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* 2.1 Product\_Clean \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*Creating Product\_Clean dataset with selected variables from Product dataset\*/

Data AdvWorks.Product\_Clean;

Set AdvWorks.Product (Keep=ProductID Name ProductNumber Color ListPrice);

Run;

Title "Listing of Product\_Clean";

Proc Print Data=AdvWorks.Product\_Clean(obs=5);

Run;

Title "Contents and Description for Product\_Clean";

Proc Contents Data = AdvWorks.Product\_Clean varnum;

Run;

/\*Replacing Missing Values in Color variable by 'NA' \*/

Data AdvWorks.Product\_Clean;

Set AdvWorks.Product\_Clean;

If missing(color) then Color='NA';

Run;

/\* Checking frequency of Color variable after missing values are replaced with 'NA' \*/

title "Frequency of color variable after replacing missing values by 'NA'";

Proc Freq Data = AdvWorks.Product\_Clean;

Table Color;

Run;

/\* Inference : 248 missing values in Color variable are replaced by 'NA' \*/

/\* Converting ListPrice variable to numeric datatype and formatting \*/

Data AdvWorks.Product\_Clean;

Set AdvWorks.Product\_Clean;

ListPrice\_num=input(ListPrice,dollar20.2);

format ListPrice\_num dollar20.2;

drop ListPrice;

rename ListPrice\_num=ListPrice;

Run;

Title "Listing of Product\_Clean after replacing missing values and type conversion";

Proc Print Data=AdvWorks.Product\_Clean(obs=5);

Run;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* 2.2 SalesOrderDetail\_Clean \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*Creating SalesOrderDetail\_Clean dataset with selected variables from SalesOrderDetail dataset\*/

Data AdvWorks.SalesOrderDetail\_Clean;

Set AdvWorks.SalesOrderDetail (Keep=SalesOrderID SalesOrderDetailID OrderQty ProductID UnitPrice LineTotal ModifiedDate);

Run;

Title "Listing of SalesOrderDetail\_Clean";

Proc Print Data=AdvWorks.SalesOrderDetail\_Clean(obs=5);

Run;

Title "Contents and Description for SalesOrderDetail\_Clean";

Proc Contents Data = AdvWorks.SalesOrderDetail\_Clean varnum;

Run;

Data AdvWorks.SalesOrderDetail\_Clean;

Set AdvWorks.SalesOrderDetail\_Clean;

ModifiedDate\_num=input(ModifiedDate,yymmdd10.);

UnitPrice\_num=input(UnitPrice,dollar20.2);

LineTotal\_num=input(LineTotal,dollar20.2);

OrderQty\_num=input(OrderQty,3.);

format UnitPrice\_num LineTotal\_num dollar20.2;

format ModifiedDate\_num mmddyy10.;

drop ModifiedDate UnitPrice LineTotal OrderQty;

rename ModifiedDate\_num=ModifiedDate UnitPrice\_num=UnitPrice LineTotal\_num=LineTotal OrderQty\_num=OrderQty;

Run;

Title "Listing of SalesOrderDetail\_Clean after type conversions and formatting as per given constraints";

Proc Print Data=AdvWorks.SalesOrderDetail\_Clean(obs=5);

Run;

Title "Contents and Description for SalesOrderDetail\_Clean";

Proc Contents Data = AdvWorks.SalesOrderDetail\_Clean varnum;

Run;

/\* Including date for year 2013 and 2014 in ModifiedDate \*/

Data AdvWorks.SalesOrderDetail\_Clean;

Set AdvWorks.SalesOrderDetail\_Clean (where=(year(ModifiedDate) in (2013,2014)));

Run;

Title "Listing of SalesOrderDetail\_Clean for 2013 & 2014";

Proc Print Data=AdvWorks.SalesOrderDetail\_Clean(obs=5);

Run;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* STEP 3 - Joining and Merging \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* 3.1 Creating SalesDetails Dataset \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*Sorting SalesOrderDetail\_Clean by ProductID\*/

Proc Sort Data = AdvWorks.SalesOrderDetail\_Clean;

by ProductID;

Run;

/\*Sorting Product\_Clean by ProductID\*/

Proc Sort Data = AdvWorks.Product\_Clean;

by ProductID;

Run;

/\*Create a SalesDetails dataset by joining SalesOrderDetail\_Clean and Product\_Clean datasets by ProductID\*/

Data AdvWorks.SalesDetails;

merge AdvWorks.SalesOrderDetail\_Clean(in=DS1) AdvWorks.Product\_Clean;

by ProductID;

if DS1;

Drop SalesOrderID SalesOrderDetailID ProductNumber ListPrice;

Run;

/\* salesorderdetail is ds1 \*/

Title "Listing of SalesDetails";

Proc print data=AdvWorks.SalesDetails(obs=5);

run;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* 3.2 Creating a SalesAnalysis Dataset \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

Title "Creating SalesAnalysis Dataset with SubTotal & SubOrderQty";

Proc Summary Data=AdvWorks.SalesDetails;

by ProductId;

ID ModifiedDate UnitPrice LineTotal OrderQty Name Color;

Var LineTotal OrderQty ;

Output out=AdvWorks.SalesAnalysis(Drop=\_type\_ \_freq\_) Sum=SubTotal SubOrderQty;

Run;

/\* Formatting SubTotal Variable \*/

Data AdvWorks.SalesAnalysis;

Set AdvWorks.SalesAnalysis;

Format SubTotal dollar11.2;

Run;

Title "Listing of SalesAnalysis";

Proc Print Data=AdvWorks.SalesAnalysis(Obs=5);

run;

/\* Alternate Approach

Data AdvWorks.SalesAnalysis;

Set Advworks.SalesDetails;

by ProductId;

if first.ProductId then SubTotal=0 ;

Subtotal+ LineTotal ;

if first.ProductId then SubOrderQty=0;

SubOrderQty+ OrderQty;

if last.ProductId;

if last.ProductId;

format subtotal dollar11.2;

run;

\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* STEP 4 - Data Analysis \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* 4.1 Query Analysis \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

Proc Sort Data=Advworks.SalesAnalysis Out=Advworks.SalesAnalysis;

By Color;

Run;

/\* 1. How many Red color Helmets are sold in 2013 and 2014? \*/

Title 'Total Quantity of Red color Helmets Sold in 2013 and 2014';

Proc Print Data=Advworks.SalesAnalysis NOOBS Label

Grandtotal\_label='Total Order Quantity:';

By color;

Var ProductID Name SubOrderQty;

Sum SubOrderQty;

Where lower(Name) Contains ('helmet') And Color = 'Red' ;

Label SubOrderQty = 'Order Quantity'

ProductID = 'Product ID'

Name = 'Product Name';

Run;

/\* Ans. Total Quantity of Red color Helmets sold in 2013 and 2014 : 4657 \*/

/\* 2. How many items sold in 2013 and 2014 have a Multi color? \*/

Title 'Total Quantity of Multi Colored Items Sold in 2013 and 2014';

Proc Print Data=Advworks.SalesAnalysis Label

Grandtotal\_label='Total Order Quantity : ';

Var ProductID Name SubOrderQty;

Sum SubOrderQty ;

Where Color = 'Multi' ;

Label SubOrderQty = 'Order Quantity'

ProductID = 'Product ID'

Name = 'Product Name';

Run;

/\* Ans. Multi-colored items were sold in 2013 and 2014 : 15009 \*/

/\* 3. What is the combined Sales total for all the helmets sold in 2013 and 2014? \*/

Title 'Sales total for all the Helmets Sold in 2013 and 2014';

Proc Print Data=Advworks.SalesAnalysis Label

Grandtotal\_label='Total Sales:';

Var ProductID Name subtotal;

Sum subtotal;

Where lower(Name) Contains ('helmet') ;

Label subtotal = 'Sales'

ProductID = 'Product ID'

Name = 'Product Name';

Run;

/\* Ans. Total sales for all the helmets sold in 2013 and 2014 : $381,800.34 \*/

/\* 4. How many Yellow Color Touring-1000 where sold in 2013 and 2014? \*/

Title 'Total Quantity of Yellow Color Touring-1000 Sold in 2013 and 2014';

Proc Print Data=Advworks.SalesAnalysis Label

Grandtotal\_label='Total Order Quantity:';

Var ProductID Name SubOrderQty;

Sum SubOrderQty ;

Where lower(Name) Contains ('touring-1000') And Color = 'Yellow' ;

Label SubOrderQty = 'Order Quantity'

ProductID = 'Product ID'

Name = 'Product Name';

Run;

/\* Ans. Total Quantity of Yellow Color Touring-1000 sold in 2013 and 2014 : 3168 \*/

/\* 5. What was the total sales in 2013 and 2014? \*/

/\* Solution Approach 1\*/

/\*Creating SalesAnalysis\_SalesTotal dataset with Aggregated Sales Total (Subtotal) \*/

Data Advworks.SalesAnalysis\_SalesTotal(keep=TotalSales) ;

Set Advworks.SalesAnalysis end=Lastobs;

TotalSales + subtotal;

if Lastobs;

Format TotalSales dollar12.2;

Run;

Title 'Total sales in 2013 and 2014 (Solution Approach 1)';

Proc Print Data=Advworks.SalesAnalysis\_SalesTotal NOOBS Label;

Label TotalSales = 'Total Sales';

Run;

/\* Ans. Total Sales in 2013 and 2014 together : $63680407.86 \*/

/\* Solution Approach 2 \*/

/\*Create SalesAnalysis\_YearTotal dataset with Year Variable\*/

Data AdvWorks.SalesAnalysis\_YearTotal;

Set AdvWorks.SalesDetails;

Year = Year(ModifiedDate);

Run;

/\*Sorting SalesAnalysis\_YearTotal dataset by Year Variable\*/

Proc Sort Data=Advworks.SalesAnalysis\_YearTotal Out=Advworks.SalesAnalysis\_YearTotal;

By Year;

Run;

Title "Summarizing Sales by Year";

Proc Summary Data=Advworks.SalesAnalysis\_YearTotal;

By Year;

Var LineTotal;

Output OUT=AdvWorks.SalesAnalysis\_YearTotal(Drop=\_type\_ \_freq\_) Sum=LineTotal;

Run;

Title 'Total sales in 2013 and 2014 (Solution Approach 2)';

Proc Print Data=AdvWorks.SalesAnalysis\_YearTotal Label

Grandtotal\_label='Total Sales:';

Var Year LineTotal;

Sum LineTotal;

Label LineTotal = 'Annual Sales';

Run;

/\* Ans. Total Sales in 2013 and 2014 together : $63680407.86

2013 -> $43,622,479.05

2014 -> $20,057,928.81 \*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* 5. Chart Analysis \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* Chart for Order Quantity by Product Colors \*/

Title 'Total Order Quantity of Products sold by Colors, for the years 2013 & 2014';

Proc SGPLOT Data = Advworks.SalesAnalysis;

vbar color / response=SubOrderQty stat=sum barwidth=.5 datalabel stat=sum categoryorder=respdesc;

xaxis label='Product Colors' labelattrs=(size=10);

yaxis label='Order Quantity' labelattrs=(size=10);

Run;