\*1. DATA IMPORT;  
proc import file="/home/u60700703/PFA PROJECT/AdventureWorks(2).xlsx"  
 out=Product  
 dbms=xlsx  
 replace;  
 sheet="Product";  
run;  
   
  
proc import file="/home/u60700703/PFA PROJECT/AdventureWorks(2).xlsx"  
 out=SalesOrderDetail  
 dbms=xlsx  
 replace;  
 sheet="SalesOrderDetail";  
run;  
  
\*2 DATA CLEANING;  
data Product\_clean;   
 set Product (keep = ProductID Name ProductNumber Color ListPrice);  
 if missing(color) then color = "NA";  
 ListPricenew = input(ListPrice, 7.);  
 Format ListPriceNew dollar8.2;  
 Drop ListPrice;  
 rename ListPriceNew = ListPrice;  
run;  
   
data SalesOrderDetail\_Clean;  
 set SalesOrderDetail (Keep= SalesOrderID SalesOrderDetailID OrderQty ProductID UnitPrice LineTotal ModifiedDate);   
 ModifiedDatenew = input(ModifiedDate ,YYMMDD19.);  
 UnitPricenew = input(UnitPrice,9. );  
 LineTotalnew = input(LineTotal, 12.);  
 OrderQtynew = input(OrderQty,2.);  
 year = year(ModifiedDatenew);  
 \*if year IN (2013,2014);  
 Format ModifiedDatenew mmddyy10. UnitPricenew dollar11.2 LineTotalnew dollar14.2;  
 Drop ModifiedDate UnitPrice LineTotal OrderQty; \*Year;   
 rename ModifiedDatenew = ModifiedDate  
 UnitPricenew = UnitPrice  
 LineTotalnew = Linetotal  
 OrderQtynew = OrderQty;   
run;  
  
data SalesOrderDetail\_Clean;  
 set SalesOrderDetail\_Clean;  
 where Year IN(2013,2014);  
 Drop year;  
run;  
  
  
\*3. Joining and Merging;  
proc sort data = SalesOrderDetail\_Clean out = SalesOrderDetail\_Cleansorted;  
 by ProductID;  
run;  
  
proc sort data = Product\_clean out = Product\_cleansorted;  
 by ProductID;  
run;  
  
data SalesDetails;  
 merge SalesOrderDetail\_Cleansorted (IN= in1)  
 Product\_cleansorted (IN= in2);  
 by ProductID;  
 if in1 = 1 and in2 = 1;  
 Drop SalesOrderID SalesOrderDetailID ProductNumber ListPrice;  
run;  
  
\*Grouping;  
Proc sort data = SalesDetails out = SalesDetails1;  
by ProductID;  
run;  
  
data SaleAnalysis1;  
 set SalesDetails1;  
 by ProductID;  
 if First.ProductID then Subtotal=0;  
 Subtotal + Linetotal;  
 if Last.ProductID;  
 format Subtotal dollar15.2;  
run;  
  
data SaleAnalysis2;  
 set SalesDetails1;  
 by ProductID;  
 if First.ProductID then SubOrderQty=0;  
 SubOrderQty + OrderQty;  
 if Last.ProductID;  
 format SubOrderQty comma10.2;  
run;  
  
proc sort data = SaleAnalysis1;  
by ProductID;  
run;  
  
proc sort data = SaleAnalysis2;  
by ProductID;  
run;  
  
data SalesAnalysis;  
merge SaleAnalysis1 SaleAnalysis2;  
by ProductID;  
run;  
  
\*Data Analysis;  
\*Question1;  
data Question1;  
 set SalesAnalysis;  
 if Name = 'Sport-100 Helmet, Red' and Color = 'Red' then do;  
 RED\_HELMETS = SubOrderQty;  
 end;  
 output;  
run;  
  
title 'How many Red color Helmets are sold in 2013 and 2014?';   
proc print data = Question1 (obs = 1) noobs;  
var RED\_HELMETS;  
format RED\_HELMETS comma8.2;  
run;  
  
  
\*Question 2;  
title 'Items sold in 2013 and 2014 have a Multi color';  
proc print data = SalesAnalysis noobs;  
var Name Color SubOrderQty;  
where color = 'Multi';  
sum SubOrderQty;  
run;  
  
\*Question 3;  
data Question3;  
 set SalesAnalysis;  
 ExtractName = scan(Name,1,',');  
run;  
title 'the combined Sales total for all the helmets sold in 2013 and 2014';  
proc print data = Question3;  
where ExtractName = 'Sport-100 Helmet';  
var ExtractName Subtotal;  
sum SubTotal;  
run;  
  
\*Question 4;  
data Question4;  
 set SalesAnalysis;  
 ExtractName = scan(Name,1,',');  
run;  
title 'Yellow Color Touring-1000 sold in 2013 and 2014';  
proc print data = Question4;  
where ExtractName = 'Touring-1000 Yellow' and color = 'Yellow';  
var ExtractName SubOrderQty;  
sum SubOrderQty;  
run;  
  
  
\*Question5;  
title 'total sales in 2013 and 2014';  
proc print data = SalesAnalysis;  
var Name Subtotal;  
sum Subtotal;  
run;  
  
proc sgplot data = SalesAnalysis;  
vbox SubOrderQty / category = color;  
run;  
  
  
  
  
  
proc sgplot data = SalesAnalysis;  
vbar color / datalabel datalabelattrs=(size=12) fillattrs=(color="RED");  
run;