Seneca

Academic Year	**2023 – 2024**		
Semester	☒ Fall	☐ Winter	☐ Summer
Course Code - Name	BAN130		
Instructor	Zeynep Cevik, PhD		
Assessment	Projects		
Due Date	Tuesday, November 28, 2023		

Student ID	Student Name	Role
141305227	Mamat Jasseh	Member

Projects

You are required to choose a project from the list of the projects specified in this document and complete it within groups of max. 4. The project is expected to test your technical skills in SAS base programming and soft skills.

Since, this is a group project, it is required to be done in groups of maximum 4. In exceptional cases, there can be 1 member in a group. Each group should have a Group Lead who would be responsible for submitting the project on Blackboard (Please note that not all the members of the group are required to submit the project separately on Blackboard. One submission from the Group Lead would be sufficient). It is also the responsibility of the Group Lead to inform the instructor about the Group members and the project they have chosen by **November 6th, 2023 (please send instructor an email about the details)**. In case an email is not received, instructor would assign the groups, group leads and the project on **November 28th, 2023.**

The detailed requirements for each project are available in this document, so please go through the details and fulfil all the requirements to avoid missing any marks.

Finally, follow the below mentioned instructions carefully.

Instructions:

In order to obtain maximum marks in this assessment, please ensure the followings:

- Don't forget to write your name and ID on the first page of this document. The student IDs and names of all the students in the group should be mentioned along with the roles.
- Submit the project by writing your solution in this document under the Solution heading below. Do not use a separate document. Everything related to the project should be included in this document, e.g., code, screenshots and etc.
- This project has a weightage of **30%** marks of the course.
- This is a group project so **only 1 submission from the group lead is required.**
- Group Leads are required to submit the project on Blackboard as instructed. Submissions through emails will not be accepted.
- The project deadline is **November 28, 2023**. Submissions after the deadline will not be accepted.

Rubric:

Your assessment will be graded based on the following rubric:

	Excellent (7 - 10)	Average (4 – 6.9)	Poor (<4)
Project Completion and Code (10)	The project was completed without any errors and output is as expected. Fulfills all/most of the requirements for the project.	The project was completed with few errors. Fulfills some of the requirements for the project.	The project is incomplete. Does not fulfill all/most of the requirements.
Detailed Explanation (10)	The student has a good contribution to the project. Knows ins and outs of the project. The student has written his/her part of the project very well. Knows everything / most of his/her part.	The student has average contribution to the project. Does not know the whole project. The student has averagely written his/her part of the project. Knows few of the things about his/her part.	The student has no contribution to the project. Does not know anything / most about the project. The student has poorly wrote the project. Does not know much about the project.
Report (10)	Student has contributed well in preparing the project report and knows all the aspects of the report.	Student has contributed partially in preparing the project report and knows some aspects of the report.	Student has not contributed in preparing the report.

Project 1

Project Name: Adventure Works Product Sales Analysis
Max. no. of students in a group: 4 students

Dataset: AdventureWorks.xlsx (Available on Blackboard)

Requirements:
Below are bare minimum requirements for this project, however, you are free to add more features to your project:

1. Data Import
   - This phase requires you to import the data from the provided excel file into SAS using Proc Import.
     - Product sheet in excel file should be imported as Product dataset in SAS.
     - SalesOrderDetail sheet in excel file should be imported as SalesOrderDetail dataset in SAS.
2. Data Cleaning
   - This phase requires you to clean your data before data analysis phase.
     - Product_Clean:
       - Create a Product_Clean dataset from Product dataset by bringing in only ProductID, Name, ProductNumber, Color and ListPrice
       - All the missing values in Color column should be replaced by 'NA'
       - ListPrice column should be numeric (final column name should be ListPrice) and format should have a dollar sign with 2 decimal places
       - No un-necessary columns should be part of the Product_Clean dataset. Please see below expected output.
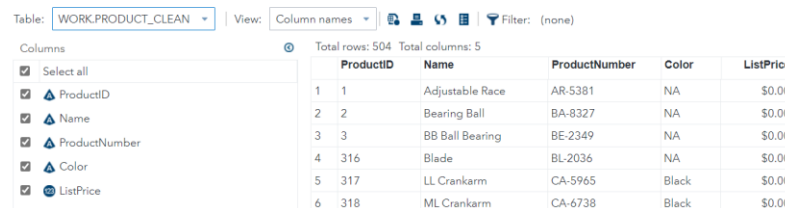

     - SalesOrderDetail_Clean:
       - Create SalesOrderDetail_Clean dataset from SalesOrderDetail dataset by bringing in only SalesOrderID SalesOrderDetailID OrderQty ProductID UnitPrice LineTotal and ModifiedDate
       - ModifiedDate should be numeric with column name ModifiedDate

- UnitPrice should be numeric with column name UnitPrice
- LineTotal should be numeric with column name LineTotal
- OrderQty should be numeric with column name OrderQty
- Include date for year 2013 and 2014 in ModifiedDate only
- ModifiedDate should be mmddyy10. Format
- UnitPrice and LineTotal should have a dollar with 2 decimal places
- No un-necessary columns should be part of the SalesOrderDetail_Clean dataset. Please see expected output below:

Table: WORK.SALESORDERDETAIL_CLEAN ▾ | View: Column names ▾ | 🗔 🖳 ↺ ▦ | ▼Filter: (none)

Columns ⊗ | Total rows: 93912 Total columns: 7

		SalesOrderID	SalesOrderDetailID	ProductID	ModifiedDate	UnitPrice	LineTotal	OrderQty
☑ Select all	1	49464	27765	707	01/28/2013	$20.19	$80.75	4
☑ △ SalesOrderID	2	49473	27871	707	01/28/2013	$20.19	$60.56	3
☑ △ SalesOrderDetailID	3	49479	27910	707	01/28/2013	$20.19	$60.56	3
☑ △ ProductID	4	49481	27996	707	01/28/2013	$20.19	$121.12	6
☑ 🗓 ModifiedDate	5	49484	28036	707	01/28/2013	$20.19	$40.37	2
☑ ◉ UnitPrice	6	49490	28160	707	01/28/2013	$20.19	$201.87	10
☑ ◉ LineTotal	7	49498	28289	707	01/28/2013	$20.19	$20.19	1
☑ ◉ OrderQty	8	49501	28369	707	01/28/2013	$20.19	$100.93	5

3. Joining and Merging
- This phase requires you to join/merge your datasets to create a dataset for analysis.
  - SalesDetails:
    - Create a SalesDetails dataset by joining SalesOrderDetail_Clean and Product_Clean datasets
    - Use ProductID column for joining the tables
    - SalesDetails table should contain all the observations from SalesOrderDetail_Clean table along with columns from Product_Clean
    - Drop SalesOrderID SalesOrderDetailID ProductNumber and ListPrice from the result dataset. Please see expected output below:

Table: WORK.SALESDETAILS ▾ | View: Column names ▾ | 🗔 🖳 ↺ ▦ | ▼Filter: (none)

Columns ⊗ | Total rows: 93912 Total columns: 7

		ProductID	ModifiedDate	UnitPrice	LineTotal	OrderQty	Name	Color
☑ Select all	1	707	01/28/2013	$20.19	$80.75	4	Sport-100 Helmet, Red	Red
☑ △ ProductID	2	707	01/28/2013	$20.19	$60.56	3	Sport-100 Helmet, Red	Red
☑ 🗓 ModifiedDate	3	707	01/28/2013	$20.19	$60.56	3	Sport-100 Helmet, Red	Red
☑ ◉ UnitPrice	4	707	01/28/2013	$20.19	$121.12	6	Sport-100 Helmet, Red	Red
☑ ◉ LineTotal	5	707	01/28/2013	$20.19	$40.37	2	Sport-100 Helmet, Red	Red
☑ ◉ OrderQty	6	707	01/28/2013	$20.19	$201.87	10	Sport-100 Helmet, Red	Red
☑ △ Name	7	707	01/28/2013	$20.19	$20.19	1	Sport-100 Helmet, Red	Red
☑ △ Color	8	707	01/28/2013	$20.19	$100.93	5	Sport-100 Helmet, Red	Red

  - SalesAnalysis:
    - Create a SalesAnalysis dataset from SalesDetails dataset that groups all the products by ProductID (hint: research on obtaining a total for each by group in SAS)
    - Create SubTotal and SubOrderQty columns in SalesAnalysis that provides an aggregate sum and quantity of each product by its ProductID.

- SubTotal column should have a dollar and 2 decimal places.
- Please see below expected output:

			ProductID	ModifiedDate	UnitPrice	LineTotal	OrderQty	Name	Color	SubTotal
		1	707	06/30/2014	$34.99	$34.99	1	Sport-100 Helmet, Red	Red	$126,263.88
		2	708	06/30/2014	$34.99	$34.99	1	Sport-100 Helmet, Black	Black	$126,940.27
		3	711	06/30/2014	$34.99	$34.99	1	Sport-100 Helmet, Blue	Blue	$128,596.20
		4	712	06/30/2014	$8.99	$8.99	1	AWC Logo Cap	Multi	$38,013.93
		5	713	06/30/2014	$49.99	$49.99	1	Long-Sleeve Logo Jersey, S	Multi	$21,445.71
		6	714	06/30/2014	$49.99	$49.99	1	Long-Sleeve Logo Jersey, M	Multi	$77,087.41
		7	715	06/30/2014	$49.99	$49.99	1	Long-Sleeve Logo Jersey, L	Multi	$123,614.75
		8	716	06/30/2014	$49.99	$49.99	1	Long-Sleeve Logo Jersey, XL	Multi	$58,127.87
		9	717	05/01/2014	$858.90	$1,717.80	2	HL Road Frame - Red, 62	Red	$219,253.75

4. Data Analysis
   - This phase requires you to analyze the SalesAnalysis for Adventure Works and answer the following 5 questions by generating reports using Proc Print for each of the 5 questions:
     - How many Red color Helmets are sold in 2013 and 2014?
     - How many items sold in 2013 and 2014 have a Multi color?
     - What is the combined Sales total for all the helmets sold in 2013 and 2014?
     - How many Yellow Color Touring-1000 where sold in 2013 and 2014?
     - What was the total sales in 2013 and 2014?
   - Create at least one chart in SAS for any analysis of your choice from SalesAnalysis dataset (this analysis can be of your choice and not necessarily from above 5 questions.)
5. Project Report
   - This phase requires you to create a report in MS Word with the following requirements (you can convert to pdf format if the file size is too big):
     - Explain each and every phase of the project (from Phase 1 to 4) along with the screenshots of the output and the related SAS code
     - Include answers to questions in Phase 4 in your report along with the chart you have chosen to create along with its justification
     - Make sure not to miss any phase and output of its screenshot

Project 2

Project Name: Adventure Works Territory Sales Analysis
Max. no. of students in a group: 4 students

Dataset: AdventureWorks.xlsx (Available on Blackboard)

Requirements:
Below are bare minimum requirements for this project, however, you are free to add more features to your project:

1. Data Import
   - This phase requires you to import the data from the provided excel file into SAS using Proc Import.
     - SalesTerritory sheet in excel file should be imported as SalesTerritory dataset in SAS.
     - SalesOrderHeader sheet in excel file should be imported as SalesOrderHeader dataset in SAS.

2. Data Cleaning
   - This phase requires you to clean your data before data analysis phase.
     - SalesOrderHeader_Clean:
       - Create a SalesOrderHeader_Clean dataset from SalesOrderHeader dataset by bringing in only SalesOrderID OrderDate OnlineOrderFlag TerritoryID TotalDue
       - TotalDue column should be numeric (final column name should be TotalDue) and format should have a dollar sign with 2 decimal places
       - OnlineOrderFlag column should be numeric (final column name should be OnlineOrderFlag)
       - OrderDate column should be numeric (final column name should be OrderDate) and format should be mmddyy10.
       - TerritoryID column should be numeric (final column name should be TerritoryID)
       - No un-necessary columns should be part of the SalesOrderHeader_Clean dataset. Please see below expected output.

Table: WORK.SALESORDERHEADER_CLEAN ▾ | View: Column names ▾ | 🗋 🖥 ↻ ▦ | ▼ Filter: (none)

Columns | ⊚ | Total rows: 31465 Total columns: 5

☑ Select all		SalesOrderID	TotalDue	OnlineOrderFlag	OrderDate	TerritoryID
☑ △ SalesOrderID	1	43664	$27,510.40	0	05/31/2011	1
☑ ⊕ TotalDue	2	43665	$16,158.60	0	05/31/2011	1
☑ ⊕ OnlineOrderFlag	3	43669	$807.26	0	05/31/2011	1
☑ 📅 OrderDate	4	43671	$9,153.60	0	05/31/2011	1
☑ ⊕ TerritoryID	5	43683	$48,204.00	0	05/31/2011	1
	6	43686	$3,899.67	0	05/31/2011	1

- ▪ Territory_Clean:
  - • Create Territory_Clean dataset from SalesTerritory dataset by bringing in only TerritoryID Name CountryRegionCode Group SalesYTD
  - • SalesYTD column should be numeric (final column name should be SalesYTD) and format should have a dollar sign with 2 decimal places
  - • TerritoryID column should be numeric (final column name should be TerritoryID)
  - • No un-necessary columns should be part of the Territory_Clean dataset. Please see below expected output.

Table: WORK.TERRITORY_CLEAN ▾ | View: Column names ▾ | 🗋 🖥 ↻ ▦ | ▼ Filter: (none)

Columns | ⊚ | Total rows: 10 Total columns: 5

☑ Select all		Name	CountryRegionCode	Group	SalesYTD	TerritoryID
☑ △ Name	1	Northwest	US	North America	$7,887,186.79	1
☑ △ CountryRegionCode	2	Northeast	US	North America	$2,402,176.85	2
☑ △ Group	3	Central	US	North America	$3,072,175.12	3
☑ ⊕ SalesYTD	4	Southwest	US	North America	$10,510,853.87	4
☑ ⊕ TerritoryID	5	Southeast	US	North America	$2,538,667.25	5
	6	Canada	CA	North America	$6,771,829.14	6

3. Joining and Merging
   - • This phase requires you to join / merge your datasets to create a dataset for analysis.
     - ▪ SalesDetails:
       - • Create a SalesDetails dataset by joining SalesOrderHeader_Clean and Territory_Clean datasets
       - • Use TerritoryID column for joining the tables
       - • SalesDetails table should contain all the observations from SalesOrderHeader_Clean table along with columns from Territory_Clean
       - • Please see expected output below:

Table: WORK.SALESDETAILS ▾ | View: Column names ▾ | 🗋 🖥 ↻ ▦ | ▼ Filter: (none)

Columns | ⊚ | Total rows: 31465 Total columns: 9

☑ Select all		SalesOrderID	TotalDue	OnlineOrderFlag	OrderDate	TerritoryID	Name	CountryRegionCode ▲	Group	SalesYTD
☑ △ SalesOrderID	1	73238	$1,276.80	-1	05/18/2014	9	Australia	AU	Pacific	$5,977,814.92
☑ ⊕ TotalDue	2	70823	$76.21	-1	04/17/2014	9	Australia	AU	Pacific	$5,977,814.92
☑ ⊕ OnlineOrderFlag	3	65877	$1,342.40	-1	02/06/2014	9	Australia	AU	Pacific	$5,977,814.92
☑ 📅 OrderDate	4	67004	$2,582.68	-1	02/24/2014	9	Australia	AU	Pacific	$5,977,814.92
☑ ⊕ TerritoryID	5	69571	$44.18	-1	03/31/2014	9	Australia	AU	Pacific	$5,977,814.92
☑ △ Name	6	64658	$1,972.38	-1	01/20/2014	9	Australia	AU	Pacific	$5,977,814.92
☑ △ CountryRegionCode	7	72101	$1,297.80	-1	05/03/2014	9	Australia	AU	Pacific	$5,977,814.92
☑ △ Group	8	62292	$54.11	-1	12/17/2013	9	Australia	AU	Pacific	$5,977,814.92
☑ ⊕ SalesYTD	9	74418	$26.28	-1	06/08/2014	9	Australia	AU	Pacific	$5,977,814.92
	10	60768	$30.14	-1	11/24/2013	9	Australia	AU	Pacific	$5,977,814.92
	11	67685	$2,588.92	-1	03/05/2014	9	Australia	AU	Pacific	$5,977,814.92

- SalesAnalysis:
  - Create a SalesAnalysis dataset from SalesDetails dataset that groups all the Territories by TerritoryID (hint: research on obtaining a total for each by group in SAS)
  - Create SubTotal and SubQty columns in SalesAnalysis that provides an aggregate sum and quantity of each Territory by its TerritoryID.
  - SubTotal column should have a dollar and 2 decimal places.
  - Drop TotalDue, TerritoryID SalesOrderID and OnlineOrderFlag
  - Please see below expected output:

	Name	CountryRegionCode	Group	SalesYTD	SubTotalDue
1	Northwest	US	North America	$7,887,186.79	$18,061,631.97
2	Northeast	US	North America	$2,402,176.85	$7,820,197.92
3	Central	US	North America	$3,072,175.12	$8,913,285.31
4	Southwest	US	North America	$10,510,853.87	$27,150,546.52
5	Southeast	US	North America	$2,538,667.25	$8,884,083.76
6	Canada	CA	North America	$6,771,829.14	$18,398,897.29
7	France	FR	Europe	$4,772,398.31	$8,119,733.27
8	Germany	DE	Europe	$3,805,202.35	$5,479,809.95
9	Australia	AU	Pacific	$5,977,814.92	$11,814,353.41
10	United Kingdom	GB	Europe	$5,012,905.37	$8,574,034.66

Table: WORK.SALESANALYSIS | View: Column names | Filter: (none)
Total rows: 10 Total columns: 5

4. Data Analysis
   - This phase requires you to analyze the SalesAnalysis for Adventure Works and answer the following 5 questions by generating reports using Proc steps for each of the 5 questions:
     - What is the Total Due for all the North American Regions?
     - What is the total Sales YTD for U.S.?
     - How much is due from France and Germany?
     - What is the total Sales YTD for Europe?
     - How many total territories in U.S?
   - Create at least one chart in SAS for any analysis of your choice from SalesAnalysis dataset (this analysis can be of your choice and not necessarily from above 5 questions.)
5. Project Report
   - This phase requires you to create a report in MS Word with the following requirements (you can convert to pdf format if the file size is too big):
     - Explain each and every phase of the project (from Phase 1 to 4) along with the screenshots of the output and the related SAS code
     - Include answers to questions in Phase 4 in your report along with the chart you have chosen to create along with its justification
     - Make sure not to miss any phase and output of its screenshot

Solutions

Project Number: One

Project Name: Adventure Works Product Sales Analysis

For the purposes of this project, we will be working with the Adventure Works.xlsx dataset, with a particular emphasis on the Product and Sales Order Detail worksheets. Whereas the sales order detail worksheet has 121,317 observations and 11 variables, the product worksheet has 504 observations and 25 variables. To ascertain the quantity of observations and variables in every dataset, we perform a procedure of contents.

1. DATA IMPORT

Using the PROC IMPORT procedure, we imported the file. Then, we created the product dataset using the OUTPUT statement and read and wrote the Excel workbooks using the DBMS function. The sheet option allows us to choose which sheet to import. used the GETNAMES option at the end to tell SAS to use the first row of the imported data as variable names in the SAS dataset that was just created, as shown in the figure below.

CODE:

```
PROC IMPORT
DATAFILE='/home/u63568328/MYDATA/BAN130JACK/AdventureWorks(2).xlsx'
DBMS=XLSX
OUT=product
replace;
sheet='product';
GETNAMES=YES;
RUN;

PROC IMPORT
DATAFILE='/home/u63568328/MYDATA/BAN130JACK/AdventureWorks(2).xlsx'
DBMS=XLSX
OUT=SalesOrderDetail
replace;
sheet='SalesOrderDetail';
GETNAMES=YES;
RUN;
```

SCREENSHOT:

Product dataset

Salesorderdetail dataset

2. DATA CLEANING

To produce the Product_Clean dataset. ProductID, Name, ProductNumber, Color, and ListPrice were the only variables we focused on when reading a subset of variables from an existing dataset called Product using the SET function and KEEP. changing the original format of the ListPrice variable to a numeric format. The input function is used to accomplish this conversion. A variable called ListPrice1 holds the updated numerical value. It is then formatted to appear as a dollar amount with two decimal places. To simplify the data, the original ListPrice variable is eliminated from the dataset after ListPrice has been converted and formatted. For consistency and clarity, the new numeric ListPrice1 is then renamed back to ListPrice.

Lastly, the Color variable's missing values were handled using an IF function. It explicitly states that there is no data by substituting the string "NA" for any missing values.
The run; statement, which is carried out to generate the Product_Clean dataset, marks the end of the process. With a correctly formatted ListPrice and clearly marked missing values in the Color column, the dataset is now more concise and useful for upcoming data analysis tasks.

CODE

```
data Product_Clean; /* New dataset Product_Clean */
set Product(keep=ProductID Name ProductNumber Color ListPrice);
ListPrice1 = input(ListPrice, 15.); /*Converting ListPrice to
numeric*/
format ListPrice1 dollar12.2;
drop ListPrice;
rename ListPrice1 = ListPrice;
if missing(Color) then Color = 'NA'; /* Replacing missing  Color
column with 'NA' */
run;
proc print data = product_clean (obs=30);
run;
```

In order to produce a cleaned version called SalesOrderDetail_clean, a dataset named SalesOrderDetail will be manipulated and filtered.
We began the process by converting an existing dataset called SalesOrderDetail into a new dataset called SalesOrderDetail_clean. From the original dataset, only a subset of the variables (SalesOrderID, SalesOrderDetailID, OrderQty, ProductID, UnitPrice, LineTotal, and ModifiedDate) are retained.

Next, change the names of a few variables (ModifiedDate, UnitPrice, LineTotal, OrderQty) by adding _Char to their original names. The purpose of this renaming is to make the process of converting these variables from character to numeric formats easier. This conversion makes use of the input function. Using the yymmdd10. informat, which interprets the date in a YYYYMMDD format, ModifiedDate_Char is converted to a numeric format. Likewise, a 10. informat is used to convert UnitPrice_Char, LineTotal_Char, and OrderQty_Char to numeric formats.

Following conversion, the following variables are formatted for readability and context: OrderQty is formatted as a 10-width numeric format, UnitPrice and LineTotal are formatted as dollar12.2, and ModifiedDate is formatted as mmddyy10. These formats provide appropriate dollar formatting and date representations, which improves the readability of the data. Subsequent to conversion, the original character variables (ModifiedDate_Char, UnitPrice_Char, LineTotal_Char, OrderQty_Char) are removed from the dataset because they are inactive.

The SalesOrderDetail_clean dataset is further refined in the following DATA step by removing records based on ModifiedDate. Retained records are those whose ModifiedDate is between January 1, 2013, and December 31, 2014. A conditional if statement is used to perform this filtering, determining whether ModifiedDate falls inside the given date range.This dataset is now better suited for in-depth reporting or analysis within the given time frame.

CODE

```
data SalesOrderDetail_clean;
set SalesOrderDetail(keep=SalesOrderID SalesOrderDetailID OrderQty
ProductID UnitPrice
LineTotal ModifiedDate

rename = (ModifiedDate = ModifiedDate_Char
        UnitPrice = UnitPrice_Char
        LineTotal = LineTotal_Char
        OrderQty = OrderQty_Char));

ModifiedDate = input(ModifiedDate_Char, yymmdd10.);
UnitPrice = input(UnitPrice_Char, 10.);
LineTotal = input(LineTotal_Char, 10.);
OrderQty = input(OrderQty_Char, 10.);

format ModifiedDate mmddyy10.
       UnitPrice dollar12.2
       LineTotal dollar12.2
       OrderQty 10.;

drop ModifiedDate_Char
     UnitPrice_Char
     LineTotal_Char
     OrderQty_Char;
run;

data SalesOrderDetail_Clean;
set SalesOrderDetail_Clean;
if "01Jan2013"d <= ModifiedDate <= "31Dec2014"d;
run;
```

```
proc print data=SalesOrderDetail_clean (obs=30);
run;
```

SCREENSHOT

Product_clean dataset

Obs	ProductID	Name	ProductNumber	Color	ListPrice
1	1	Adjustable Race	AR-5381	NA	$0.00
2	2	Bearing Ball	BA-8327	NA	$0.00
3	3	BB Ball Bearing	BE-2349	NA	$0.00
4	4	Headset Ball Bearings	BE-2908	NA	$0.00
5	316	Blade	BL-2036	NA	$0.00
6	317	LL Crankarm	CA-5965	Black	$0.00
7	318	ML Crankarm	CA-6738	Black	$0.00
8	319	HL Crankarm	CA-7457	Black	$0.00
9	320	Chainring Bolts	CB-2903	Silver	$0.00
10	321	Chainring Nut	CN-6137	Silver	$0.00
11	322	Chainring	CR-7833	Black	$0.00
12	323	Crown Race	CR-9981	NA	$0.00
13	324	Chain Stays	CS-2812	NA	$0.00
14	325	Decal 1	DC-8732	NA	$0.00
15	326	Decal 2	DC-9824	NA	$0.00
16	327	Down Tube	DT-2377	NA	$0.00
17	328	Mountain End Caps	EC-M092	NA	$0.00
18	329	Road End Caps	EC-R098	NA	$0.00
19	330	Touring End Caps	EC-T209	NA	$0.00
20	331	Fork End	FE-3760	NA	$0.00
21	332	Freewheel	FH-2981	Silver	$0.00
22	341	Flat Washer 1	FW-1000	NA	$0.00
23	342	Flat Washer 6	FW-1200	NA	$0.00
24	343	Flat Washer 2	FW-1400	NA	$0.00
25	344	Flat Washer 9	FW-3400	NA	$0.00
26	345	Flat Washer 4	FW-3800	NA	$0.00
27	346	Flat Washer 3	FW-5160	NA	$0.00
28	347	Flat Washer 8	FW-5800	NA	$0.00
29	348	Flat Washer 5	FW-7160	NA	$0.00
30	349	Flat Washer 7	FW-9160	NA	$0.00

SalesOrderDetail_clean dataset

Obs	SalesOrderID	SalesOrderDetailID	ProductID	ModifiedDate	UnitPrice	Line Total	OrderQty
1	49181	27406	794	01/01/2013	$2,181.56	$2,181.56	1
2	49182	27407	790	01/01/2013	$2,443.35	$2,443.35	1
3	49183	27408	791	01/01/2013	$2,443.35	$2,443.35	1
4	49184	27409	784	01/01/2013	$2,049.10	$2,049.10	1
5	49185	27410	784	01/01/2013	$2,049.10	$2,049.10	1
6	49186	27411	791	01/01/2013	$2,443.35	$2,443.35	1
7	49187	27412	796	01/01/2013	$2,181.56	$2,181.56	1
8	49188	27413	782	01/01/2013	$2,049.10	$2,049.10	1
9	49189	27414	769	01/01/2013	$782.99	$782.99	1
10	49190	27415	798	01/01/2013	$1,000.44	$1,000.44	1
11	49191	27416	800	01/02/2013	$1,000.44	$1,000.44	1
12	49192	27417	800	01/02/2013	$1,000.44	$1,000.44	1
13	49193	27418	783	01/02/2013	$2,049.10	$2,049.10	1
14	49194	27419	766	01/02/2013	$782.99	$782.99	1
15	49195	27420	759	01/02/2013	$782.99	$782.99	1
16	49196	27421	780	01/02/2013	$2,071.42	$2,071.42	1
17	49197	27422	783	01/02/2013	$2,049.10	$2,049.10	1
18	49198	27423	770	01/02/2013	$782.99	$782.99	1
19	49199	27424	767	01/02/2013	$782.99	$782.99	1
20	49200	27425	789	01/03/2013	$2,443.35	$2,443.35	1
21	49201	27426	791	01/03/2013	$2,443.35	$2,443.35	1
22	49202	27427	794	01/03/2013	$2,181.56	$2,181.56	1
23	49203	27428	764	01/03/2013	$782.99	$782.99	1
24	49204	27429	779	01/03/2013	$2,071.42	$2,071.42	1
25	49205	27430	782	01/03/2013	$2,049.10	$2,049.10	1
26	49206	27431	794	01/03/2013	$2,181.56	$2,181.56	1
27	49207	27432	790	01/03/2013	$2,443.35	$2,443.35	1
28	49208	27433	794	01/03/2013	$2,181.56	$2,181.56	1
29	49209	27434	792	01/03/2013	$2,181.56	$2,181.56	1
30	49210	27435	781	01/03/2013	$2,071.42	$2,071.42	1

3. JOINING & MERGING

An important step before merging the datasets is to make sure that the data in both datasets is arranged in the same order based on the key variable. To do this, we used the proc sort function to order the data by the key variable ProductID.

Following the sorting of both datasets, a merging procedure is carried out. SalesDetails is a newly created dataset that was formed by combining product_clean and salesorderdetail_clean. The merge statement, which unites the datasets based on the shared variable ProductID, is used to complete this merging. Temporary binary variables are used in this process to indicate the presence of records from each original dataset: in=a for salesorderdetail_clean and in=b for product_clean.

Next, the DROP function was used to remove unnecessary columns from the combined SalesDetails dataset.

CODE

```
proc sort data=Product_Clean;
by ProductID;
run;
proc sort data=SalesOrderDetail_Clean;
by ProductID;
run;
/*Joining datasets by productID*/
data SalesDetails;
merge salesorderdetail_clean (in=a)
      product_clean (in=b);
if a;
by productID;
run;
/* dropping unwanted variables*/
data salesdetails;
set salesdetails;
drop SalesOrderID SalesOrderDetailID ProductNumber ListPrice;
run;
Title 'listing of the joined datasets';
proc print data=SalesDetails (obs=30);
run;
```

SCREENSHOT

listing of the joined datasets

Obs	ProductID	ModifiedDate	UnitPrice	Line Total	OrderQty	Name	Color
1	707	01/28/2013	$20.19	$80.75	4	Sport-100 Helmet, Red	Red
2	707	01/28/2013	$20.19	$60.56	3	Sport-100 Helmet, Red	Red
3	707	01/28/2013	$20.19	$60.56	3	Sport-100 Helmet, Red	Red
4	707	01/28/2013	$20.19	$121.12	6	Sport-100 Helmet, Red	Red
5	707	01/28/2013	$20.19	$40.37	2	Sport-100 Helmet, Red	Red
6	707	01/28/2013	$20.19	$201.87	10	Sport-100 Helmet, Red	Red
7	707	01/28/2013	$20.19	$20.19	1	Sport-100 Helmet, Red	Red
8	707	01/28/2013	$20.19	$100.93	5	Sport-100 Helmet, Red	Red
9	707	01/28/2013	$20.19	$20.19	1	Sport-100 Helmet, Red	Red
10	707	01/28/2013	$20.19	$100.93	5	Sport-100 Helmet, Red	Red
11	707	01/28/2013	$20.19	$40.37	2	Sport-100 Helmet, Red	Red
12	707	01/28/2013	$20.19	$60.56	3	Sport-100 Helmet, Red	Red
13	707	02/28/2013	$20.19	$100.93	5	Sport-100 Helmet, Red	Red
14	707	02/28/2013	$20.19	$60.56	3	Sport-100 Helmet, Red	Red
15	707	02/28/2013	$20.19	$80.75	4	Sport-100 Helmet, Red	Red
16	707	02/28/2013	$20.19	$60.56	3	Sport-100 Helmet, Red	Red
17	707	02/28/2013	$19.51	$267.73	14	Sport-100 Helmet, Red	Red
18	707	02/28/2013	$20.19	$40.37	2	Sport-100 Helmet, Red	Red
19	707	02/28/2013	$20.19	$60.56	3	Sport-100 Helmet, Red	Red
20	707	02/28/2013	$20.19	$60.56	3	Sport-100 Helmet, Red	Red
21	707	02/28/2013	$20.19	$20.19	1	Sport-100 Helmet, Red	Red
22	707	02/28/2013	$20.19	$60.56	3	Sport-100 Helmet, Red	Red
23	707	02/28/2013	$20.19	$100.93	5	Sport-100 Helmet, Red	Red
24	707	02/28/2013	$20.19	$141.31	7	Sport-100 Helmet, Red	Red
25	707	02/28/2013	$20.19	$121.12	6	Sport-100 Helmet, Red	Red
26	707	02/28/2013	$20.19	$20.19	1	Sport-100 Helmet, Red	Red
27	707	02/28/2013	$20.19	$100.93	5	Sport-100 Helmet, Red	Red
28	707	02/28/2013	$20.19	$60.56	3	Sport-100 Helmet, Red	Red
29	707	02/28/2013	$20.19	$60.56	3	Sport-100 Helmet, Red	Red
30	707	02/28/2013	$20.19	$100.93	5	Sport-100 Helmet, Red	Red

SALES ANALYSIS

We performed another sorting on a dataset named SalesDetails, with the intention of creating a new dataset called SalesAnalysis that includes a subtotal of LineTotal for each ProductID.

We used PROC SORT procedure to sort the SalesDetails dataset by ProductID. The sorted dataset is then output as SalesAnalysis.

In order to further process the SalesAnalysis dataset, we then carried out a data step. The data is still being grouped by ProductID. The calculations are carried out in this section of the steps. Two variables, SubTotal and SubOrderQty, are initialized to zero for each new ProductID that is encountered. This is crucial in order to compute totals for every product individually. In order to effectively total the sales and order quantities for each ProductID, we further accumulate the values of LineTotal into SubTotal and orderqty into SubOrderQty. As indicated by the if Last.ProductID; statement, the totals are only kept for each product's final record. With this method, only the final accumulated totals for each product are included in the dataset. Additionally, we formatted SubTotal's values with a currency to improve readability.

CODE

```
proc sort data=SalesDetails out=SalesAnalysis;
by ProductID;
run;

data SalesAnalysis;
set SalesAnalysis;
by ProductID;
if First.ProductID then SubTotal=0;
if First.ProductID then SubOrderQty=0;
SubTotal + LineTotal;
SubOrderQty+orderqty;
if Last.ProductID;
format SubTotal dollar11.2;
run;
proc print data =SalesAnalysis (obs=20);
run;
```

SCREENSHOT

Obs	ProductID	ModifiedDate	UnitPrice	LineTotal	OrderQty	Name	Color	SubTotal	SubOrderQty
1	707	06/30/2014	$34.99	$34.99	1	Sport-100 Helmet, Red	Red	$126,263.88	4657
2	708	06/30/2014	$34.99	$34.99	1	Sport-100 Helmet, Black	Black	$126,940.27	4804
3	711	06/30/2014	$34.99	$34.99	1	Sport-100 Helmet, Blue	Blue	$128,596.20	4864
4	712	06/30/2014	$8.99	$8.99	1	AWC Logo Cap	Multi	$38,013.93	5718
5	713	06/30/2014	$49.99	$49.99	1	Long-Sleeve Logo Jersey, S	Multi	$21,445.71	429
6	714	06/30/2014	$49.99	$49.99	1	Long-Sleeve Logo Jersey, M	Multi	$77,087.41	2307
7	715	06/30/2014	$49.99	$49.99	1	Long-Sleeve Logo Jersey, L	Multi	$123,614.75	3935
8	716	06/30/2014	$49.99	$49.99	1	Long-Sleeve Logo Jersey, XL	Multi	$58,127.87	1676
9	717	05/01/2014	$858.90	$1,717.80	2	HL Road Frame - Red, 62	Red	$219,253.75	260
10	718	05/01/2014	$858.90	$858.90	1	HL Road Frame - Red, 44	Red	$220,112.65	261
11	719	07/31/2013	$858.90	$2,576.70	3	HL Road Frame - Red, 48	Red	$52,392.90	61
12	722	05/01/2014	$202.33	$202.33	1	LL Road Frame - Black, 58	Black	$86,506.13	439
13	723	07/31/2013	$202.33	$607.00	3	LL Road Frame - Black, 60	Black	$12,342.25	61
14	725	04/30/2013	$202.33	$202.33	1	LL Road Frame - Red, 44	Red	$37,229.09	184
15	726	02/28/2014	$249.54	$249.54	1	LL Road Frame - Red, 48	Red	$28,016.24	138
16	727	04/30/2013	$202.33	$607.00	3	LL Road Frame - Red, 52	Red	$1,820.99	9
17	729	06/30/2013	$249.54	$499.09	2	LL Road Frame - Red, 60	Red	$37,525.84	185
18	730	03/30/2014	$249.54	$249.54	1	LL Road Frame - Red, 62	Red	$31,743.33	157
19	736	05/01/2014	$202.33	$202.33	1	LL Road Frame - Black, 44	Black	$22,311.70	113
20	738	05/01/2014	$202.33	$607.00	3	LL Road Frame - Black, 52	Black	$152,887.25	774

4. DATA ANALYSIS

We carry out various forms of analysis to identify specific trends or discover answers to all the questions.

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

CODE
```
title "Total Red Helmets Sold in 2013 and 2014";
Proc print data=SalesAnalysis;
where color='Red'and name like'%Helmet%';
sum SubOrderQty;
run; proc print data=RedHelmetsSales noobs;
run;
```

SCREENSHOT

Total Red Helmets Sold in 2013 and 2014

Obs	ProductID	ModifiedDate	UnitPrice	LineTotal	OrderQty	Name	Color	SubTotal	SubOrderQty
1	707	06/30/2014	$34.99	$34.99	1	Sport-100 Helmet, Red	Red	$126,263.88	4657
									4657

A total of **4,657** red helmets were sold in the years 2013 and 2014.

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

CODE

```
title "Total Multi color items sold in 2013 and 2014?";
Proc print data=SalesAnalysis;
where color='Multi';
sum subOrderQty;
run;
```

SCREENSHOT

Total Multi color items sold in 2013 and 2014?

Obs	ProductID	ModifiedDate	UnitPrice	LineTotal	OrderQty	Name	Color	SubTotal	SubOrderQty
4	712	06/30/2014	$8.99	$8.99	1	AWC Logo Cap	Multi	$38,013.93	5718
5	713	06/30/2014	$49.99	$49.99	1	Long-Sleeve Logo Jersey, S	Multi	$21,445.71	429
6	714	06/30/2014	$49.99	$49.99	1	Long-Sleeve Logo Jersey, M	Multi	$77,087.41	2307
7	715	06/30/2014	$49.99	$49.99	1	Long-Sleeve Logo Jersey, L	Multi	$123,614.75	3935
8	716	06/30/2014	$49.99	$49.99	1	Long-Sleeve Logo Jersey, XL	Multi	$58,127.87	1676
97	855	02/28/2014	$58.49	$116.99	2	Men's Bib-Shorts, S	Multi	$15,784.25	292
98	856	04/30/2014	$58.49	$233.97	4	Men's Bib-Shorts, M	Multi	$26,208.29	491
99	857	04/30/2013	$53.99	$161.98	3	Men's Bib-Shorts, L	Multi	$8,693.03	161
									15009

A total of **15,009** multi-color items were sold in the years 2013 and 2014.

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

CODE

```
title 'Combined Sales total for all helmets sold in 2013 and 2014?';
Proc print data=SalesAnalysis;
where name like'%Helmet%';
sum SubTotal;
run;
```

SCHREENSHOT

Combined Sales total for all helmets sold in 2013 and 2014?

Obs	ProductID	ModifiedDate	UnitPrice	LineTotal	OrderQty	Name	Color	SubTotal	SubOrderQty
1	707	06/30/2014	$34.99	$34.99	1	Sport-100 Helmet, Red	Red	$126,263.88	4657
2	708	06/30/2014	$34.99	$34.99	1	Sport-100 Helmet, Black	Black	$126,940.27	9461
3	711	06/30/2014	$34.99	$34.99	1	Sport-100 Helmet, Blue	Blue	$128,596.20	14325
								$381,800.34	

A total of **381,800.34** combined helmets were sold in the years 2013 and 2014.

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

CODE

```
title 'total Yellow Color Touring-1000 sold in 2013 and 2014?';
Proc print data=SalesAnalysis;
where color='Yellow'and name like'%Touring-1000%';
sum SubOrderQty;
run;
```

SCREENSHOT

			total Yellow Color Touring-1000 sold in 2013 and 2014?						
Obs	ProductID	ModifiedDate	UnitPrice	LineTotal	OrderQty	Name	Color	SubTotal	SubOrderQty
194	954	05/30/2014	$2,384.07	$2,384.07	1	Touring-1000 Yellow, 46	Yellow	$1426372.87	1005
195	955	05/25/2014	$2,384.07	$2,384.07	1	Touring-1000 Yellow, 50	Yellow	$981,187.85	652
196	956	05/29/2014	$2,384.07	$2,384.07	1	Touring-1000 Yellow, 54	Yellow	$667,158.15	397
197	957	05/25/2014	$2,384.07	$2,384.07	1	Touring-1000 Yellow, 60	Yellow	$1518133.10	1114
									3168

A total of **3,168** color touring-100 were sold in the years 2013 and 2014.

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

CODE

```
Proc print data=SalesAnalysis;
sum Subtotal;
run;
```

SCREENSHOT

Obs	ProductID	ModifiedDate	UnitPrice	LineTotal	OrderQty	Name	Color	SubTotal	SubOrderQty
1	707	06/30/2014	$34.99	$34.99	1	Sport-100 Helmet, Red	Red	$126,263.88	4657
2	708	06/30/2014	$34.99	$34.99	1	Sport-100 Helmet, Black	Black	$126,940.27	4804
3	711	06/30/2014	$34.99	$34.99	1	Sport-100 Helmet, Blue	Blue	$128,596.20	4864
4	712	06/30/2014	$8.99	$8.99	1	AWC Logo Cap	Multi	$38,013.93	5718
5	713	06/30/2014	$49.99	$49.99	1	Long-Sleeve Logo Jersey, S	Multi	$21,445.71	429
6	714	06/30/2014	$49.99	$49.99	1	Long-Sleeve Logo Jersey, M	Multi	$77,087.41	2307
7	715	06/30/2014	$49.99	$49.99	1	Long-Sleeve Logo Jersey, L	Multi	$123,614.75	3935
8	716	06/30/2014	$49.99	$49.99	1	Long-Sleeve Logo Jersey, XL	Multi	$58,127.87	1676
9	717	05/01/2014	$858.90	$1,717.80	2	HL Road Frame - Red, 62	Red	$219,253.75	260
10	718	05/01/2014	$858.90	$858.90	1	HL Road Frame - Red, 44	Red	$220,112.65	261
11	719	07/31/2013	$858.90	$2,576.70	3	HL Road Frame - Red, 48	Red	$52,392.90	61
12	722	05/01/2014	$202.33	$202.33	1	LL Road Frame - Black, 58	Black	$86,506.13	439
13	723	07/31/2013	$202.33	$607.00	3	LL Road Frame - Black, 60	Black	$12,342.25	61
14	725	04/30/2013	$202.33	$202.33	1	LL Road Frame - Red, 44	Red	$37,229.09	184
15	726	02/28/2014	$249.54	$249.54	1	LL Road Frame - Red, 48	Red	$28,016.24	138
16	727	04/30/2013	$202.33	$607.00	3	LL Road Frame - Red, 52	Red	$1,820.99	9
17	729	06/30/2013	$249.54	$499.09	2	LL Road Frame - Red, 60	Red	$37,525.84	185
18	730	03/30/2014	$249.54	$249.54	1	LL Road Frame - Red, 62	Red	$31,743.33	157
19	736	05/01/2014	$202.33	$202.33	1	LL Road Frame - Black, 44	Black	$22,311.70	113
20	738	05/01/2014	$202.33	$607.00	3	LL Road Frame - Black, 52	Black	$152,887.25	774
21	739	05/01/2014	$818.70	$4,093.50	5	HL Mountain Frame - Silver, 42	Silver	$146,845.01	181
22	742	05/01/2014	$818.70	$3,274.80	4	HL Mountain Frame - Silver, 46	Silver	$280,516.39	351
23	743	05/01/2014	$809.76	$809.76	1	HL Mountain Frame - Black, 42	Black	$515,964.36	652
216	976	05/27/2014	$1,700.99	$1,700.99	1	Road-350-W Yellow, 48	Yellow	$1774883.56	1622
217	977	05/30/2014	$539.99	$539.99	1	Road-750 Black, 58	Black	$361,145.31	892
218	978	05/30/2014	$742.35	$742.35	1	Touring-3000 Blue, 44	Blue	$133,365.03	292
219	979	05/25/2014	$742.35	$742.35	1	Touring-3000 Blue, 50	Blue	$348,581.50	841
220	980	05/29/2014	$769.49	$769.49	1	Mountain-400-W Silver, 38	Silver	$241,773.76	425
221	981	05/26/2014	$769.49	$769.49	1	Mountain-400-W Silver, 40	Silver	$323,703.82	617
222	982	05/29/2014	$769.49	$769.49	1	Mountain-400-W Silver, 42	Silver	$217,457.87	385
223	983	05/26/2014	$769.49	$769.49	1	Mountain-400-W Silver, 46	Silver	$227,347.67	401
224	984	05/28/2014	$564.99	$564.99	1	Mountain-500 Silver, 40	Silver	$145,089.43	450
225	985	05/29/2014	$564.99	$564.99	1	Mountain-500 Silver, 42	Silver	$141,360.50	455
226	986	05/27/2014	$564.99	$564.99	1	Mountain-500 Silver, 44	Silver	$122,512.43	381
227	987	05/29/2014	$564.99	$564.99	1	Mountain-500 Silver, 48	Silver	$142,897.27	457
228	988	05/27/2014	$564.99	$564.99	1	Mountain-500 Silver, 52	Silver	$161,293.35	515
229	989	05/15/2014	$539.99	$539.99	1	Mountain-500 Black, 40	Black	$101,734.12	282
230	990	05/27/2014	$539.99	$539.99	1	Mountain-500 Black, 42	Black	$136,293.48	388
231	991	05/27/2014	$539.99	$539.99	1	Mountain-500 Black, 44	Black	$125,925.67	350
232	992	05/28/2014	$539.99	$539.99	1	Mountain-500 Black, 48	Black	$157,569.08	449
233	993	05/26/2014	$539.99	$539.99	1	Mountain-500 Black, 52	Black	$96,982.20	272
234	994	05/01/2014	$32.39	$97.18	3	LL Bottom Bracket	NA	$12,244.93	378
235	996	05/01/2014	$72.89	$218.68	3	HL Bottom Bracket	NA	$39,581.44	543
236	997	05/30/2014	$539.99	$539.99	1	Road-750 Black, 44	Black	$290,298.62	656
237	998	05/30/2014	$539.99	$539.99	1	Road-750 Black, 48	Black	$578,174.31	1556
238	999	05/30/2014	$539.99	$539.99	1	Road-750 Black, 52	Black	$515,666.91	1338
								63680407.86	

The total sales in 2013 and 2014 is **63,680,407.86.**

6. A bar graph that depicts the total number of sales by color

CODE
```
proc sgplot data=salesanalysis;
    vbar color;
    xaxis label='color';
    yaxis label='sales';
```

```
    title 'Sales by color';
run;
```

SCREENSHOT

.