Seneca

Academic Year	**2023**		
Semester	☐ Fall	☒ Winter	☐ Summer
Course Code - Name	BAN110		
Instructor	Dr. Razi Iqbal		
Assessment	Assignment 2		

Student ID **176938215**

Student Name **Rongzhao Yi**

Assignment 1

The main purpose of this lab is to get students familiarize with Data Preparation using Character and Date comparisons.

Instructions:

- You are required to submit your answers in this document by pasting your SAS code under the solutions heading below.
- Please do not submit .sas files. Submit only this word document with your code inside it.
- Total Marks for this assignment are 5 marks.
- Students having exactly similar code will get a straight 0.
- You are required to complete these exercises using SAS.
- The deadline for submission of this assignment is Feb. 21 end of the day.

Question

You are provided with the dataset 'Superstore.xlsx' file. It is an excel file with three sheets, 'Orders', 'Returns' and 'People'. You are required to perform the following tasks using this dataset in SAS:

- Import the data from the excel sheet in SAS (You have already done something like this in Assignment 1).
- Create a dataset in SAS namely Orders which would get all the data from Orders sheet.
- Once the data is imported you need to create a new dataset called 'Orders_2016' which would bring only the following columns from the original Orders dataset: Order ID, Order Date, State, Region, Category, Quantity and Sales.
- Work in your data module to filter the data in Orders_2016 to contain data that has Order Date of November 2016. So, basically you are fetching all the records from the original Orders dataset into this dataset that have order date of November 2016. You might need to do some research on how to filter on Dates in SAS.
- Once you have your 'Orders_2016' dataset ready, create a report which only prints the records of **Furniture** Orders in **South** along with its **quantity**. So, basically, we are answering a question, "How many furniture items were ordered in November 2016 in South Region?"
- Below is the expected output of this program:

Order_ID	Order_Date	State	Region	Category	Sales	Quantity
CA-2016-152156	11/08/2016	Kentucky	South	Furniture	261.96	2
CA-2016-152156	11/08/2016	Kentucky	South	Furniture	731.94	3
CA-2016-168893	11/03/2016	Kentucky	South	Furniture	24.1	5
CA-2016-168893	11/03/2016	Kentucky	South	Furniture	842.94	3
CA-2016-163167	11/28/2016	Georgia	South	Furniture	182.67	3
US-2016-153815	11/06/2016	Florida	South	Furniture	207.984	2
US-2016-153815	11/06/2016	Florida	South	Furniture	35.568	2
CA-2016-113845	11/20/2016	Florida	South	Furniture	289.568	2
CA-2016-165827	11/04/2016	Florida	South	Furniture	50.496	6
CA-2016-104276	11/27/2016	Florida	South	Furniture	331.023	7
CA-2016-162236	11/04/2016	North Carolina	South	Furniture	876.3	10
CA-2016-130820	11/15/2016	North Carolina	South	Furniture	630.024	4
CA-2016-157280	11/05/2016	Virginia	South	Furniture	273.96	2
CA-2016-157280	11/05/2016	Virginia	South	Furniture	756.8	5
CA-2016-123533	11/24/2016	Florida	South	Furniture	339.92	5
						61

Make sure to include only the columns shown in the screenshot above.

Solution

```
options validvarname=v7;
PROC IMPORT OUT= WORK.Orders DATAFILE= "/home/u63055836/BAN110ZAA/Superstore.xlsx"
            DBMS=xlsx REPLACE;
     SHEET="Orders";
     GETNAMES=yes;
RUN;

data Orders_2016;
    set Orders(keep=Order_ID Order_Date State Region Category Quantity Sales);
    where Order_Date between '01NOV2016'd and '30NOV2016'd;
run;

proc print data=Orders_2016(where=(Category='Furniture' and Region='South')) noobs;
    sum Quantity;
run;
```