##### Case Study 1

An ITES company wants to understand its data related to services requests for customers. These requests are related to particular products that the company manufactures. The requests can be divided into 3 priorities: - Low, Medium and High. Each of these priority levels have a different SLA (Service Level Agreement) for resolution. The data consists of the following fields :-

1 ServiceRequestNo:- Unique id for each Service request logged on the system

2 ServiceRequestStatus:- As of date status on the Service request

3 TypeOfEnagagement:- What type of work is the company doing for the client

4 Incident/ Problem:- Is the Service request a problem?

5 SR Priority:- What is the priority level of the Service Request

6 SR Open Date:- Date on which the Service Request was logged on system

7 SR Close Date:- Date on which the Service Request was closed on system

8 Product :- The name of the product for which the Service Request was raised

9 Geography:- The continent in which the client belongs

10 Country:- The country to which the client belongs

Clean the data and make it ready for analysis of Resolution time i.e., Time, in number of days ,taken to resolve the Service Request (to be calculated as difference between SR Close Date and SR Open Date.)

Let's create a Project Plan on the lines of DCOVA&I

1. Define the problem :- Clean the data and create a Project Datamart to do Analytics for Resolution Time i.e., Time, in number of days ,taken to resolve the Service Request (to be calculated as difference between SR Close Date and SR Open Date.)

1.1. Create the y variable

2. Colect the relevant data: - The data for the project comes from one file "CaseStudy1.csv"

3. Organise the data: - Manipulate data , create derived variables thru calculation, Understand missing values

4. Visualise the data :-

4.1. Univariate analysis of y

4.2. Multi-variate analysis :- Correlation

5. Create the final Project Data mart

5.1. Drop variables :-

5.1.1. Variables used to create y

5.1.2. Variables which are non-numeric

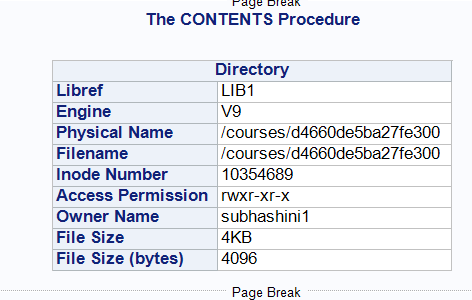
5.2. Numeric values :- ServiceRequestStatus, TypeOfEnagagement, Incident/ Problem, SR Priority, Product, Geography, Country

Now execute the above mentioned steps in SAS

1. Create Libname - In SAS on Demand for Academics , the library is pre-set for the course . You can view the same in the SAS Studio link. (Hint :- Google is your best friend to find out relevant documents and DIY - Do It Yourself processes. SAS has extensive online help. )

libname lib1 " /home/subhashini1/my\_content "; run;

PROC CONTENTS data = lib1.\_ALL\_ NODS;run;



2. Import data :- I loaded the csv file into the SAS on Demand cloud directory.

FILENAME REFFILE "/home/subhashini1/my\_content/CaseStudy1.csv" TERMSTR=CR;

PROC IMPORT DATAFILE=REFFILE

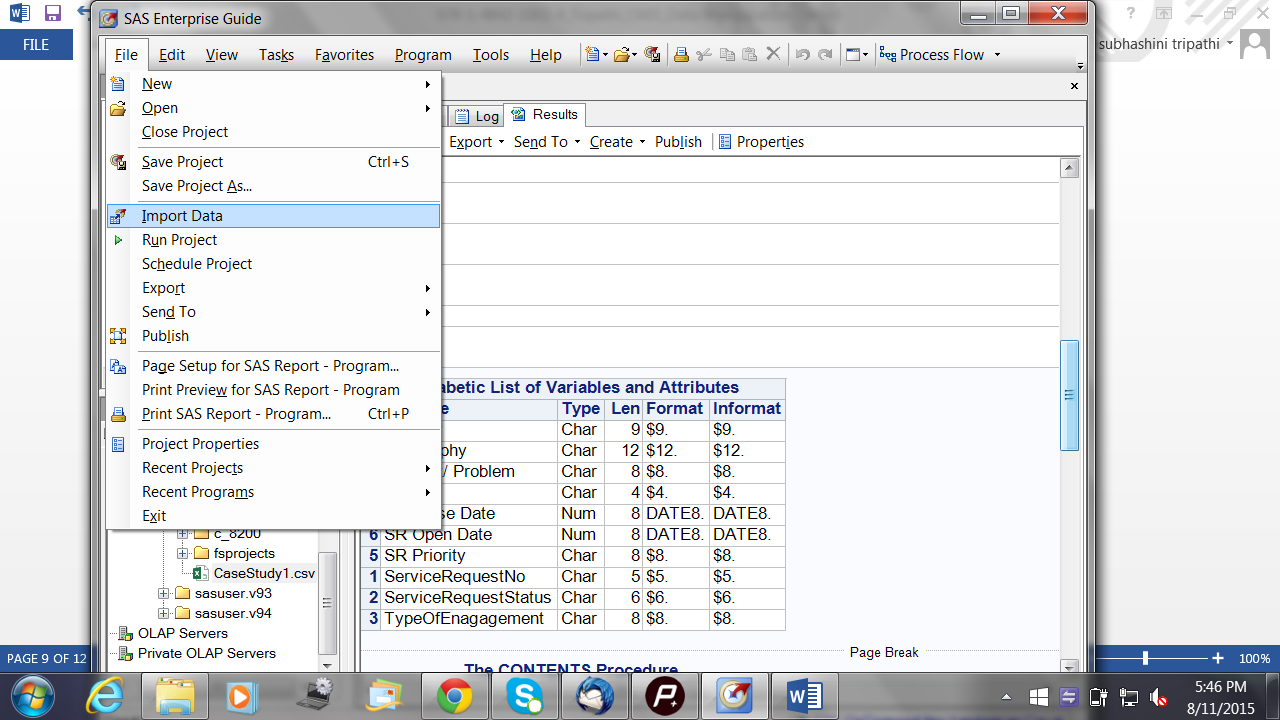
DBMS=CSV

OUT=WORK.IMPORT;

GETNAMES=YES;

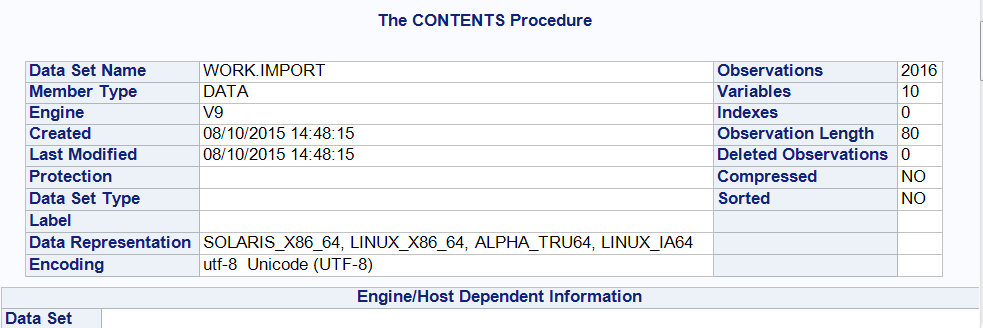
RUN;

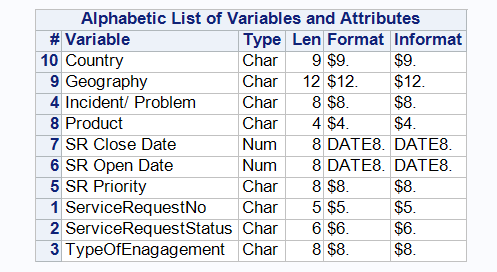
Note :- Easier way to import without touching the cloud based server is to use the button driven option - File > Import Data. This will open an Import Wizard thru which you can pull the data into the SAS system from your desktop !!



Did the file get loaded ? Is all the Data present? PROC CONTENTS will help us understand.

PROC CONTENTS DATA=WORK.IMPORT; RUN;



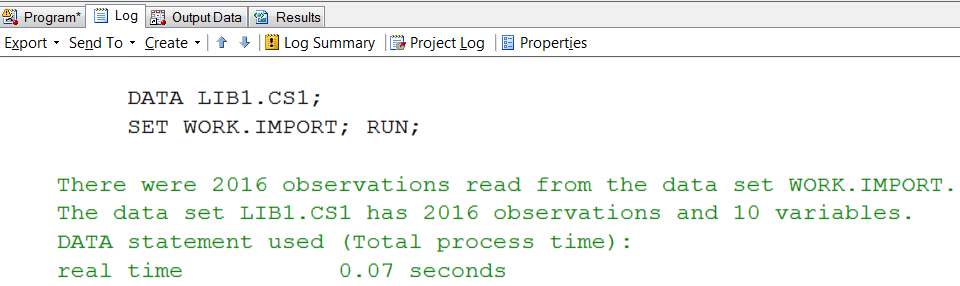


Note :- Look up the Little SAS Book for Enterprise Guide (purchase from the SAS website - <https://www.sas.com/store/prodBK_61861_en.html>) . For free help on SAS coding , look up <http://www.sascommunity.org/wiki/Sasopedia/Topics>.

3. Store the SAS dataset back in SAS on Demand library

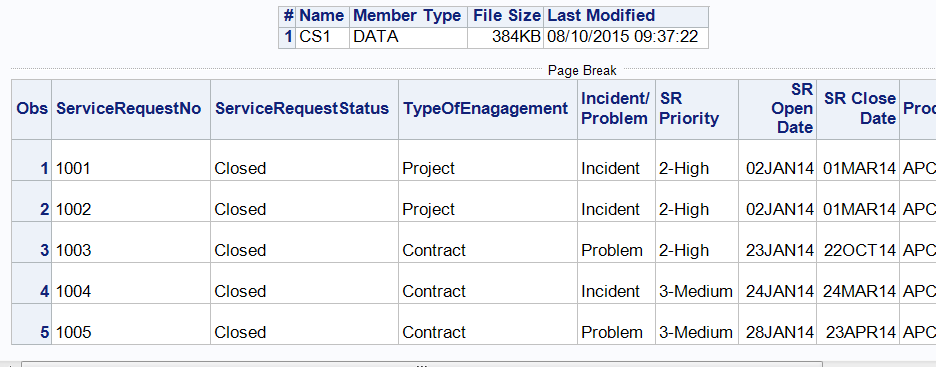
DATA LIB1.CS1;

SET WORK.IMPORT; RUN;



4. View 5 lines from SAS file CS1 (CaseStudy1)to see how the data looks.

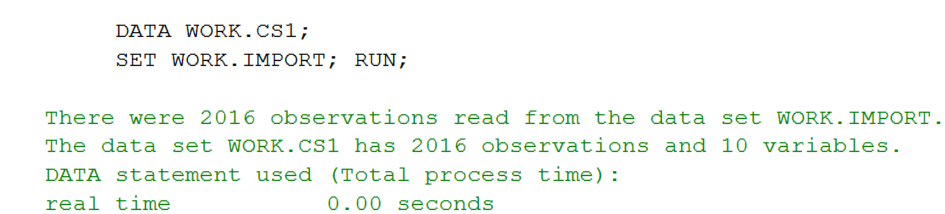
PROC PRINT DATA=LIB1.CS1(OBS=5);RUN;



5. Work on file WORK.IMPORT in the WORK directory . Then save the final, modified dataset in the permanent directory (LIB1). Create copy of file WORK.IMPORT as WORK.CS1.

DATA WORK.CS1;

SET WORK.IMPORT; RUN;



`6. Create y variable of Resolution Time = difference between SR Close Date and SR Open Date.

VALIDNAME=ANY;

DATA WORK.CS1;

SET WORK.CS1;

RESOLUTIONTIME= 'SR Close Date'n-'SR Open Date'n; RUN;

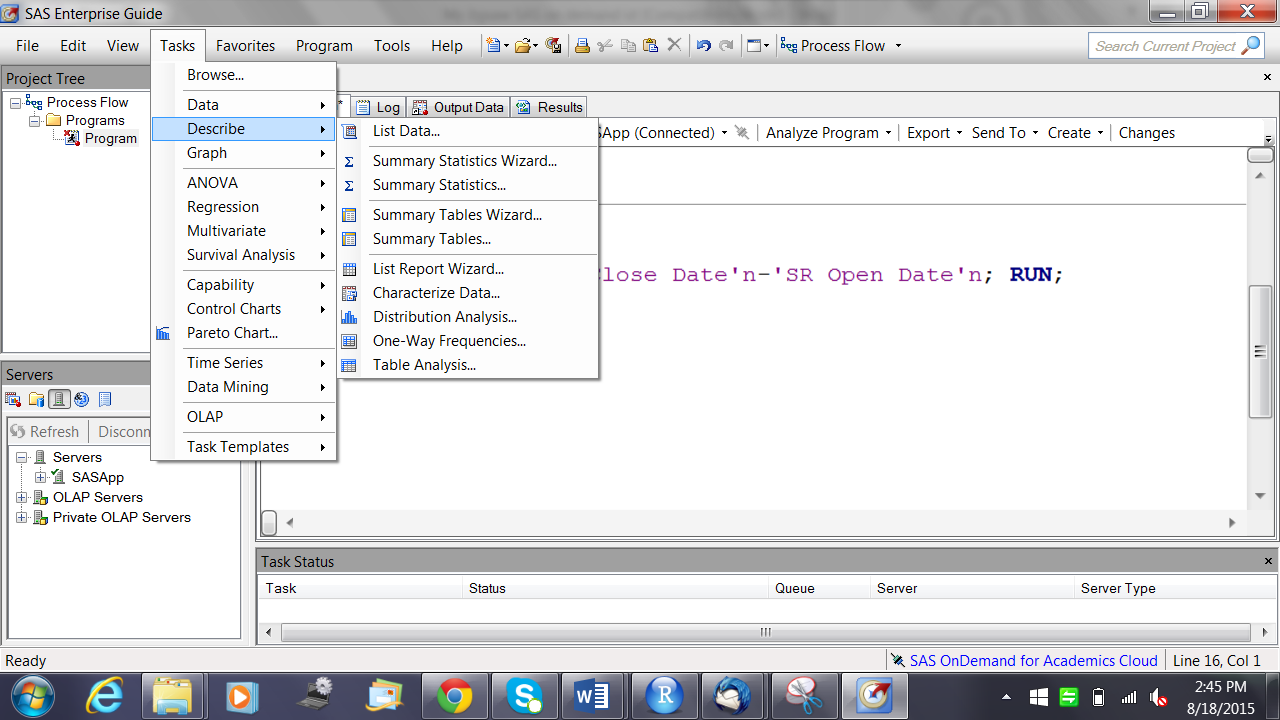
PROC PRINT DATA=WORK.CS1(OBS=5);RUN;

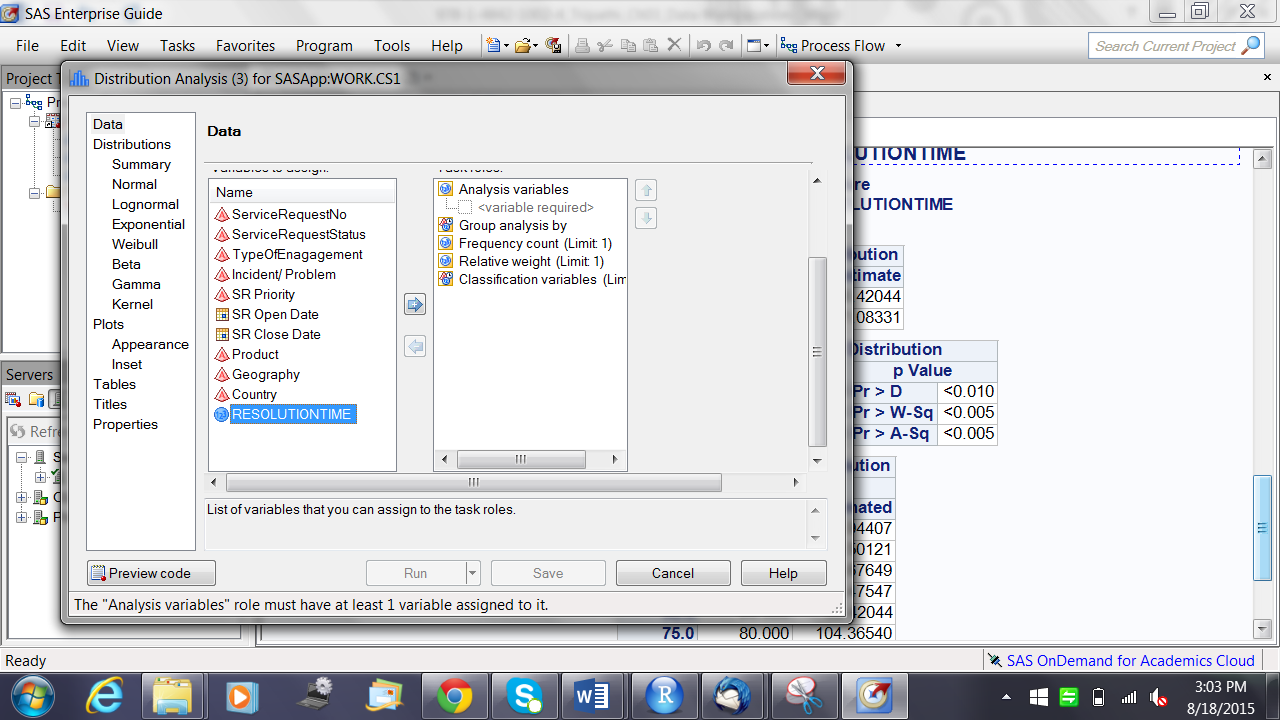
Faq :- Why do we use VALIDNAME=ANY; and n and with the variable names 'SR Close Date'n-'SR Open Date'n? Because there are spaces in the variable names which will not be acceptable to SAS . Ideally, during import, any spaces or special characters in variable names should be replaced by under-score (\_) in SAS.

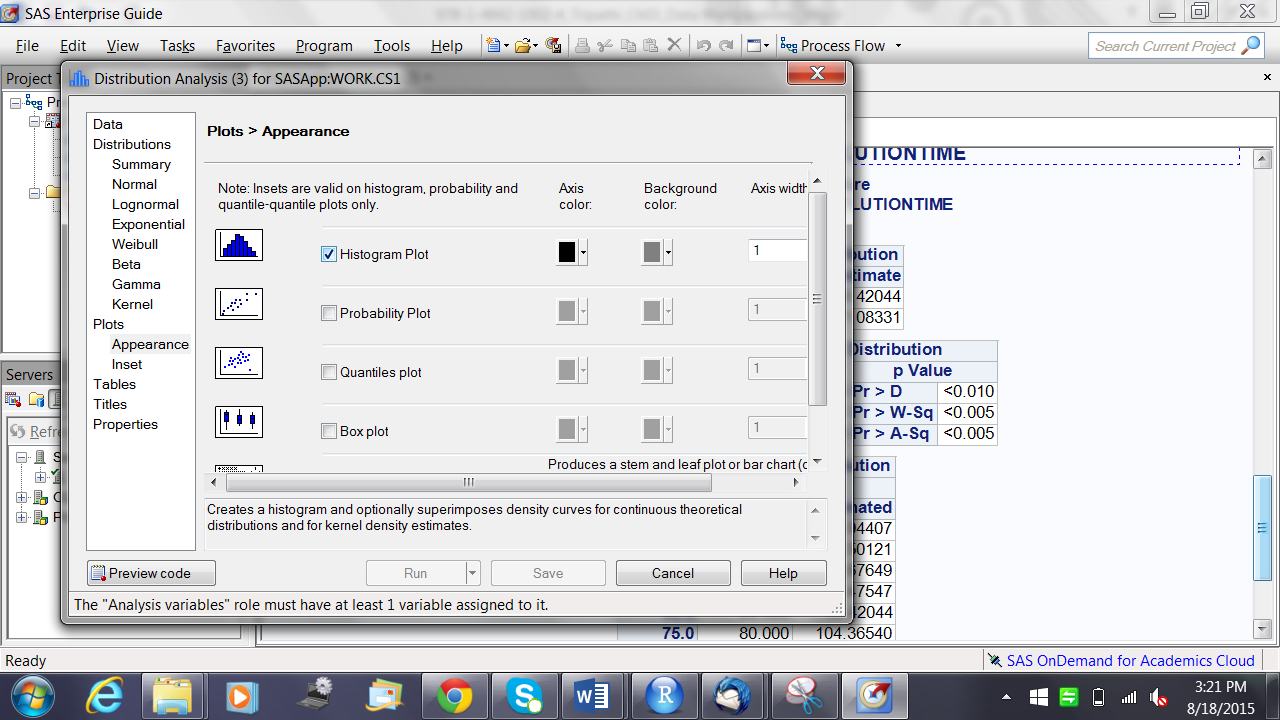
Faq:- How do we know the variable name in the SAS dataset ? We can see the names and type of variable, whether character / date / numeric, in the PROC CONTENTS output.

7. Understand the distribution of the y variable

Choose the Task Tab > Describe data tab and then choose Distribution Analysis.

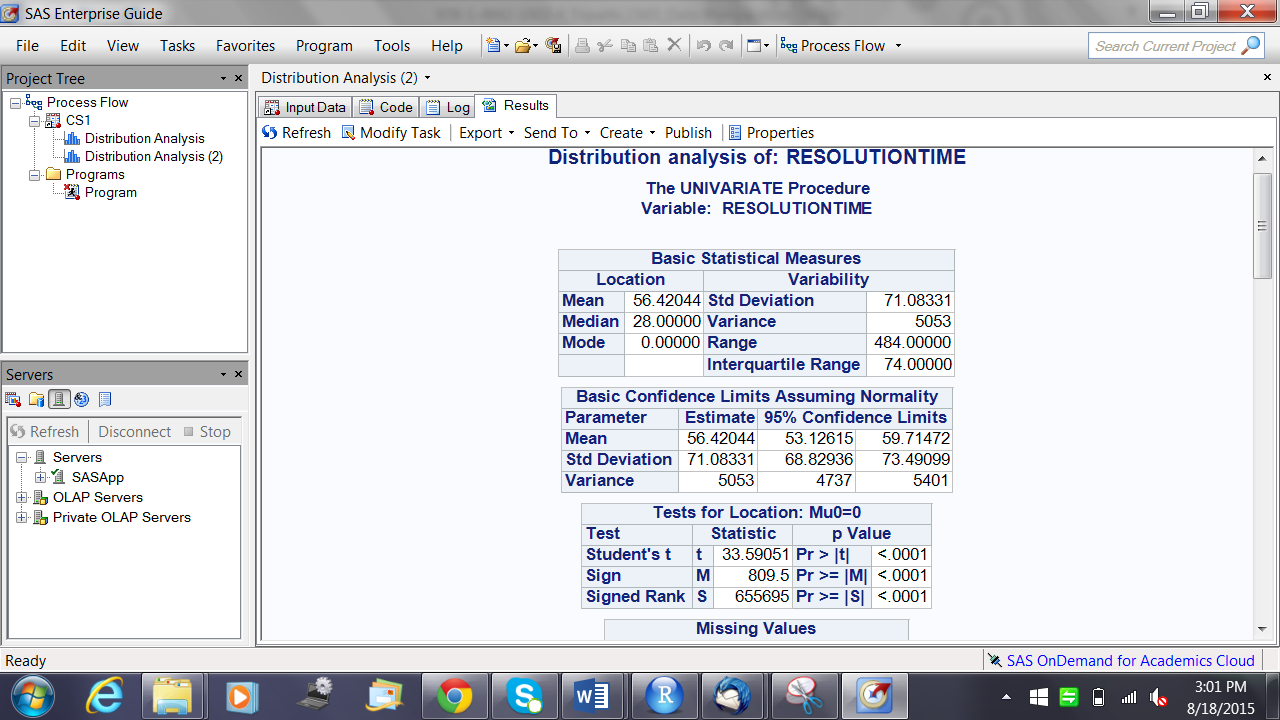


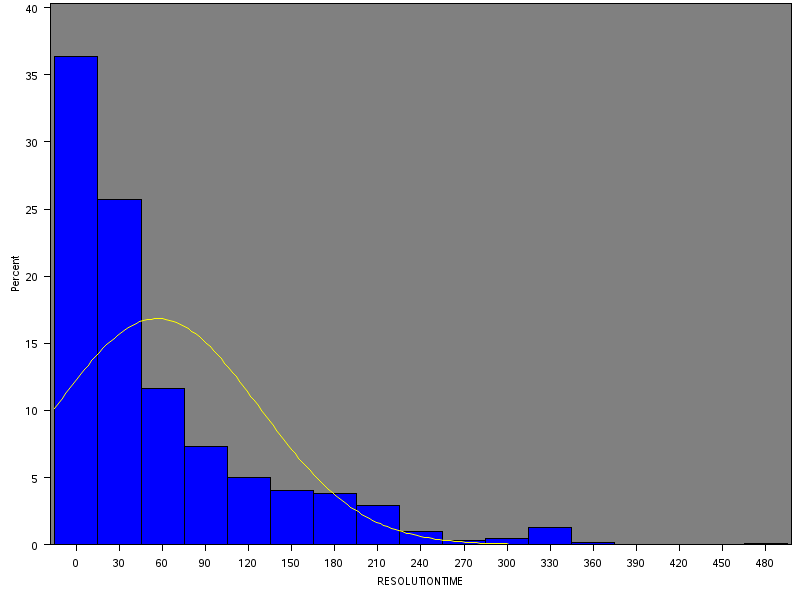


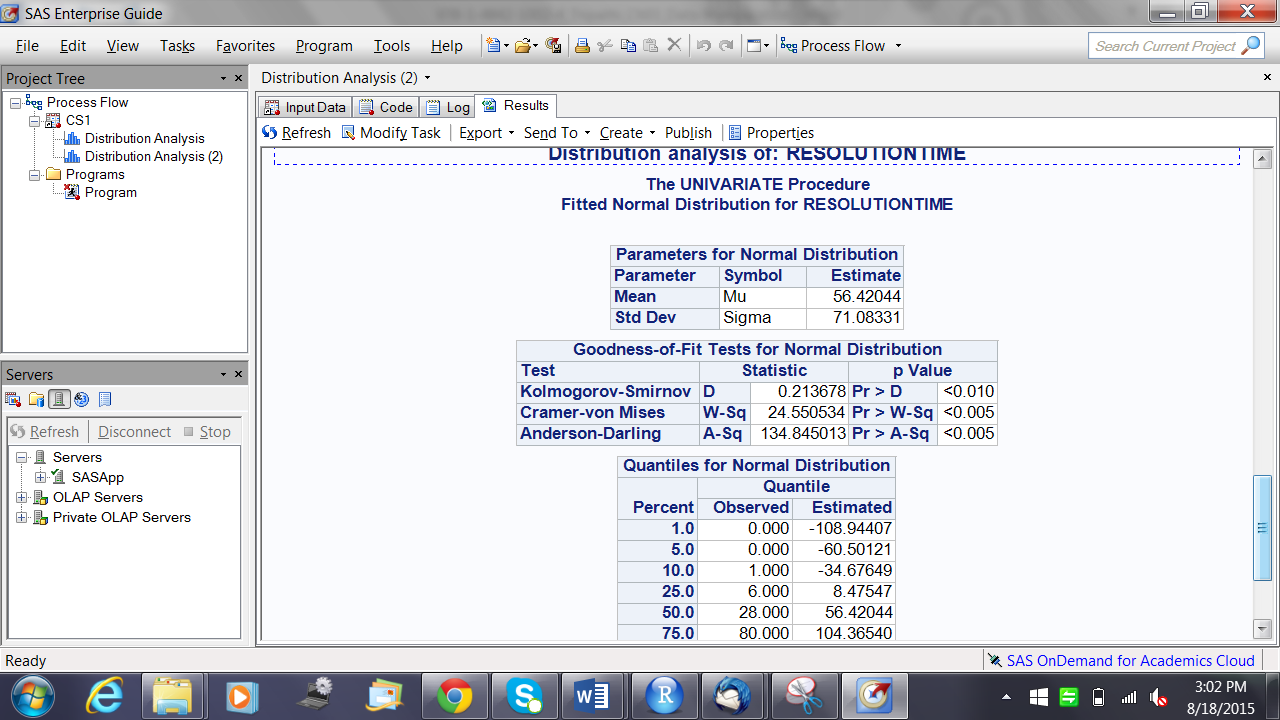


Choose the Normal Distribution and a Histogram Plot

Tip :- Click on Preview Code to see the SAS code for the options you have choosed thru the button menu !!







8. Univariate Analysis of Y variable . What does the descriptive stats say ?

Note :- Use the Proc Unvariate Proceedure to run a similar output in code



Tip :- Copy paste the outputs from the Results Tab into an Excel sheet . You will find that the format fits perfectly in excel

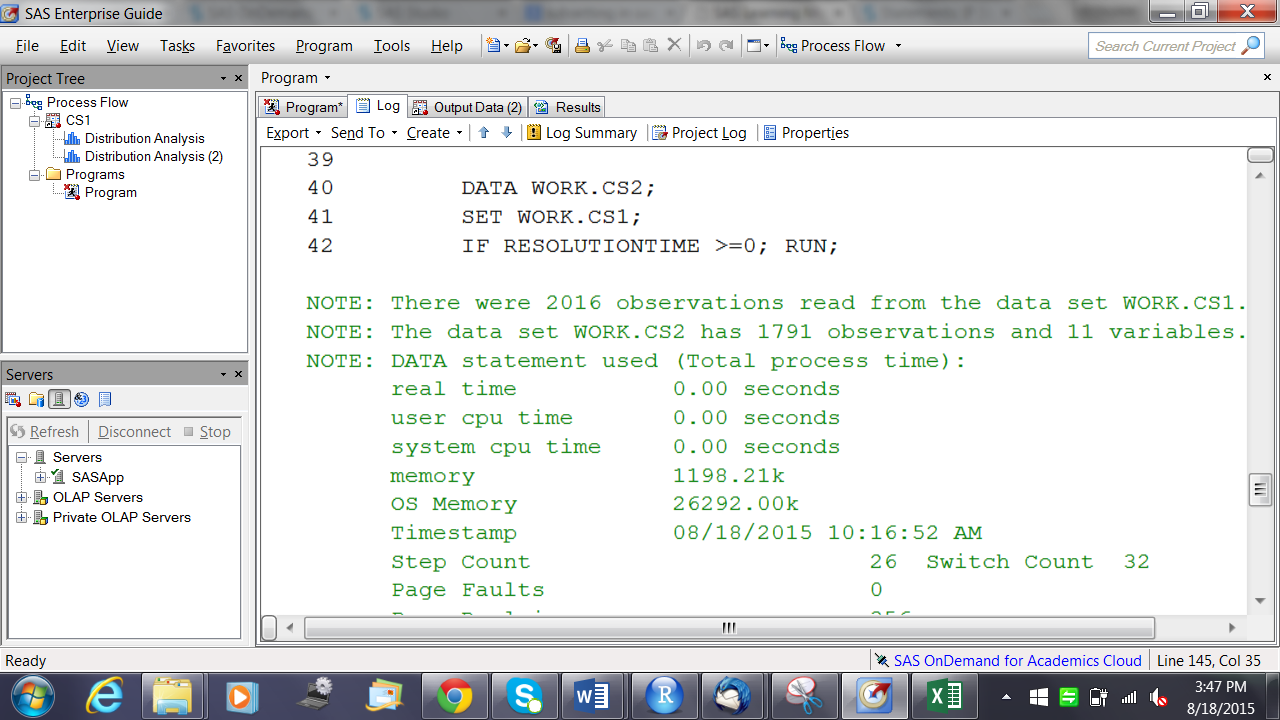
Note :- Codes in SAS are not case sensitive . You can write them in Upper or Lower case and it will run with equal effectiveness.

9. Drop the observations with Y missing .

DATA WORK.CS2;

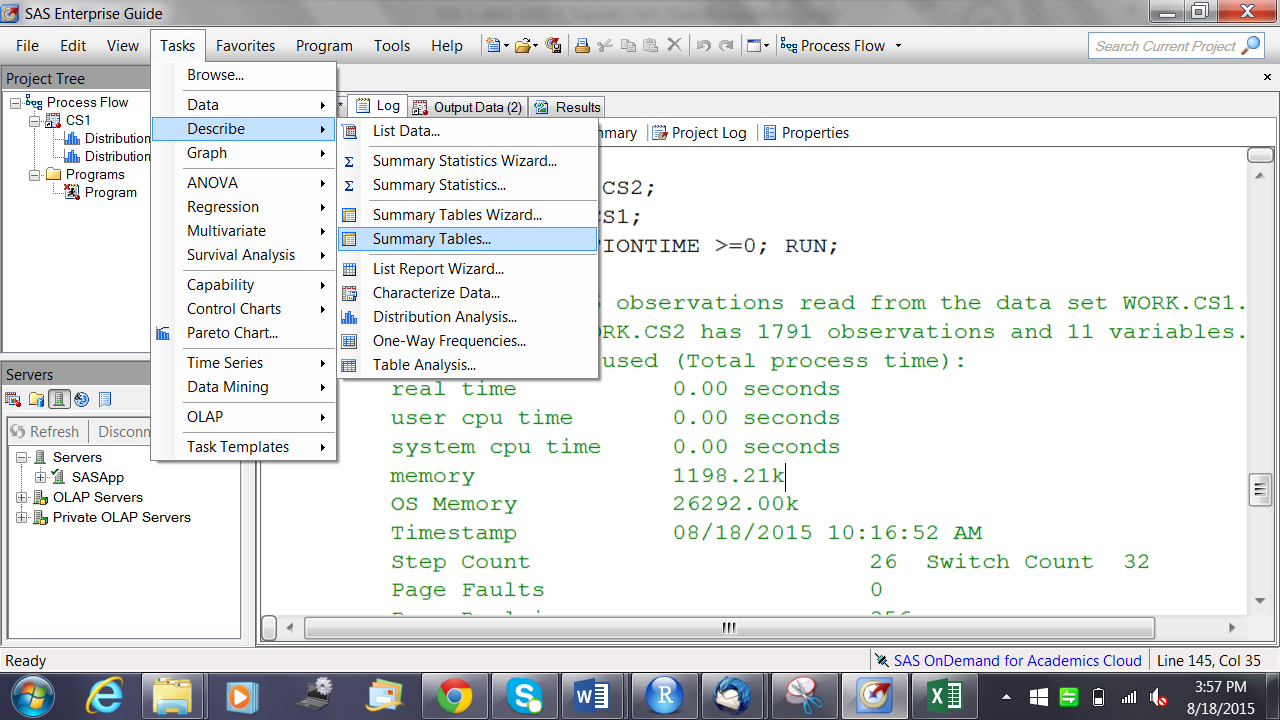
SET WORK.CS1;

IF RESOLUTIONTIME >=0; RUN;

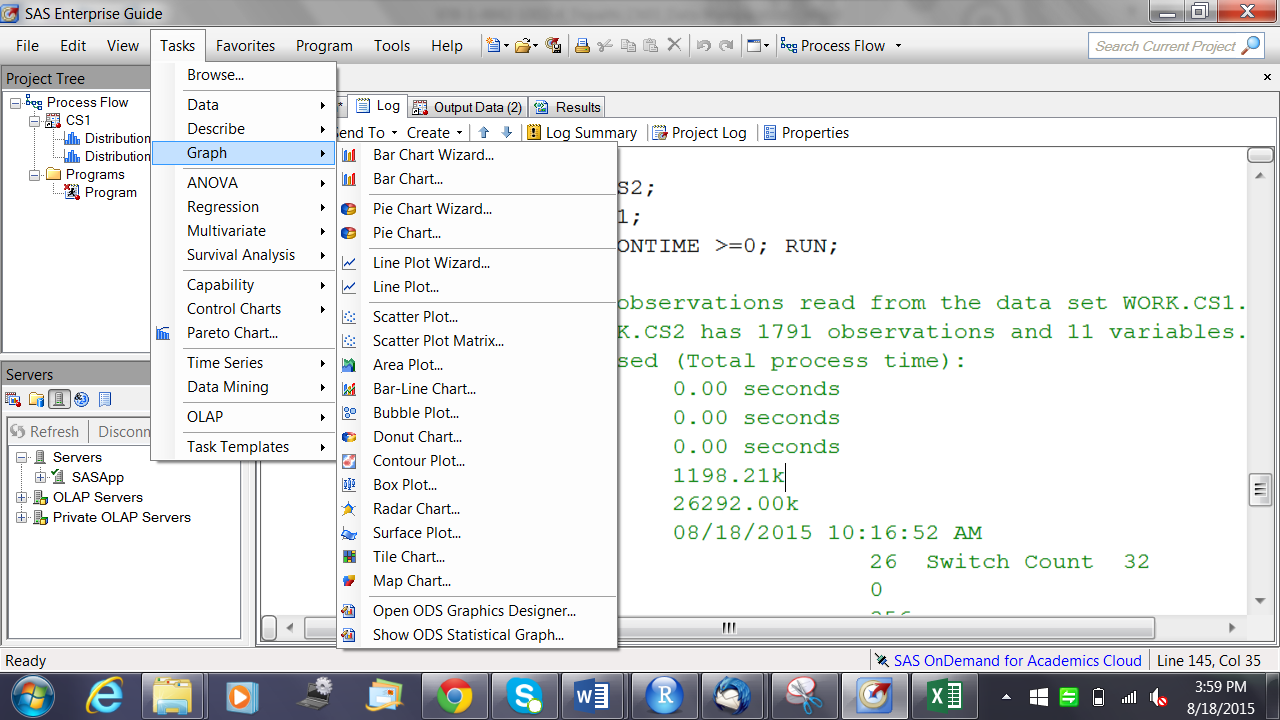


10. Do some visualisations to understand the data

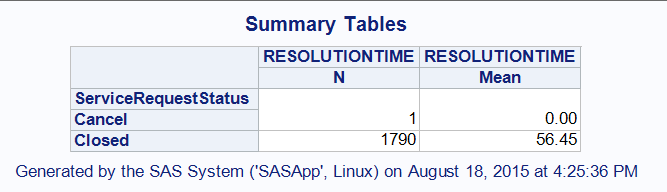
Run Tables from Tasks > Describe > Summary Tables



Run Graphs from Tasks > Graphs



Plot Average Resolution time and count of observations against Service Request Status

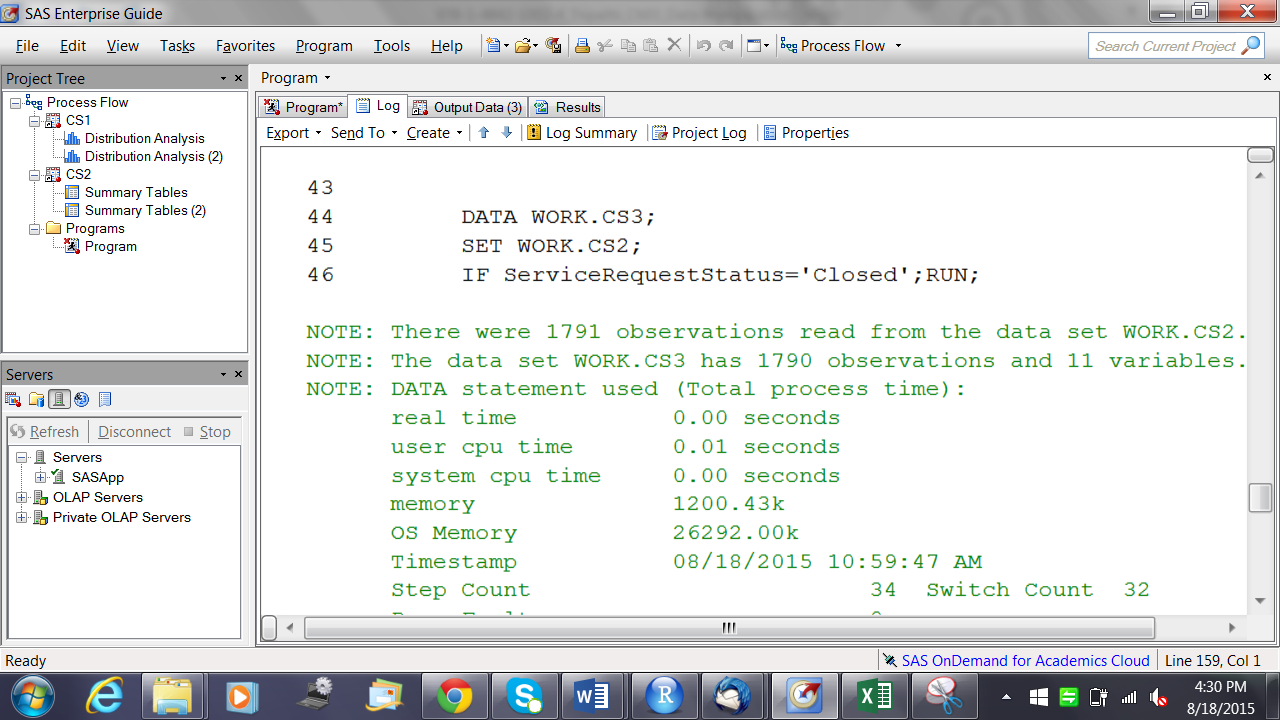


Remove the /observations where Status is Closed as it’s an outlier and should not be included in the analysis

DATA WORK.CS3;

SET WORK.CS2;

IF ServiceRequestStatus='Closed';RUN;



Lets check the spread of Resolution Time over Geography

Tip :- PROC TABULATE is used to create the Summary Tables that we can create from the automated button system .

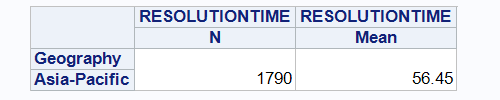
PROC TABULATE DATA=WORK.CS3;

VAR RESOLUTIONTIME;

CLASS Geography;

TABLE Geography, RESOLUTIONTIME\*N RESOLUTIONTIME\*MEAN;

RUN;



Since all the cases belong to the same Geography , we can drop this Variable.

Resolution Time by Country :-

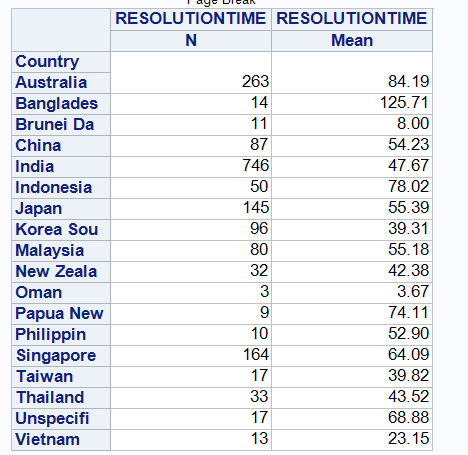
PROC TABULATE DATA=WORK.CS3;

VAR RESOLUTIONTIME;

CLASS Country;

TABLE Country, RESOLUTIONTIME\*N RESOLUTIONTIME\*MEAN;

RUN;



Resolution Time by Product :-

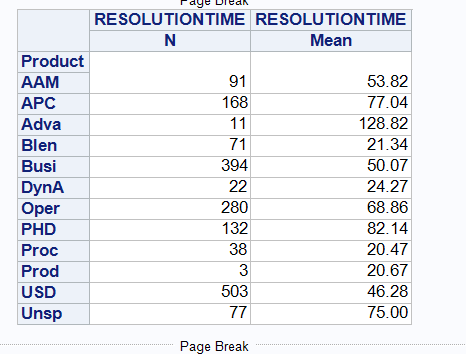
PROC TABULATE DATA=WORK.CS3;

VAR RESOLUTIONTIME;

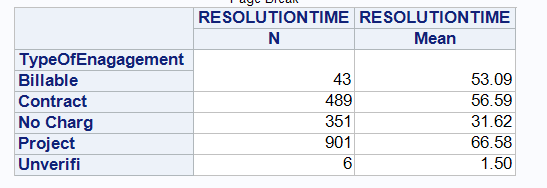
CLASS Product;

TABLE Product, RESOLUTIONTIME\*N RESOLUTIONTIME\*MEAN;

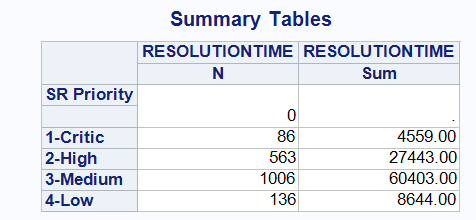
RUN;



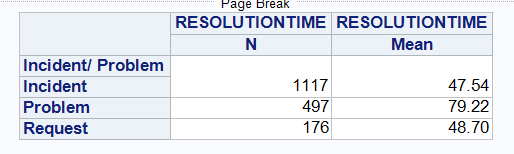
Type Of Engagement for Resolution Time :-



Priority level and Resolution Time



Incident/Problem vs Resolution Time



11. To create the final Project Datamart , the following data manipulations should be done

11.1 Remove Geography as a field since all cases are for Asia-Pacific

11.2 Remove Service Request Status as all status is closed

11.3 Remove SR Open Date as it is used to calculate the Y variables

11.4 Remove SR Close Date as it is used to calculate the Y variable

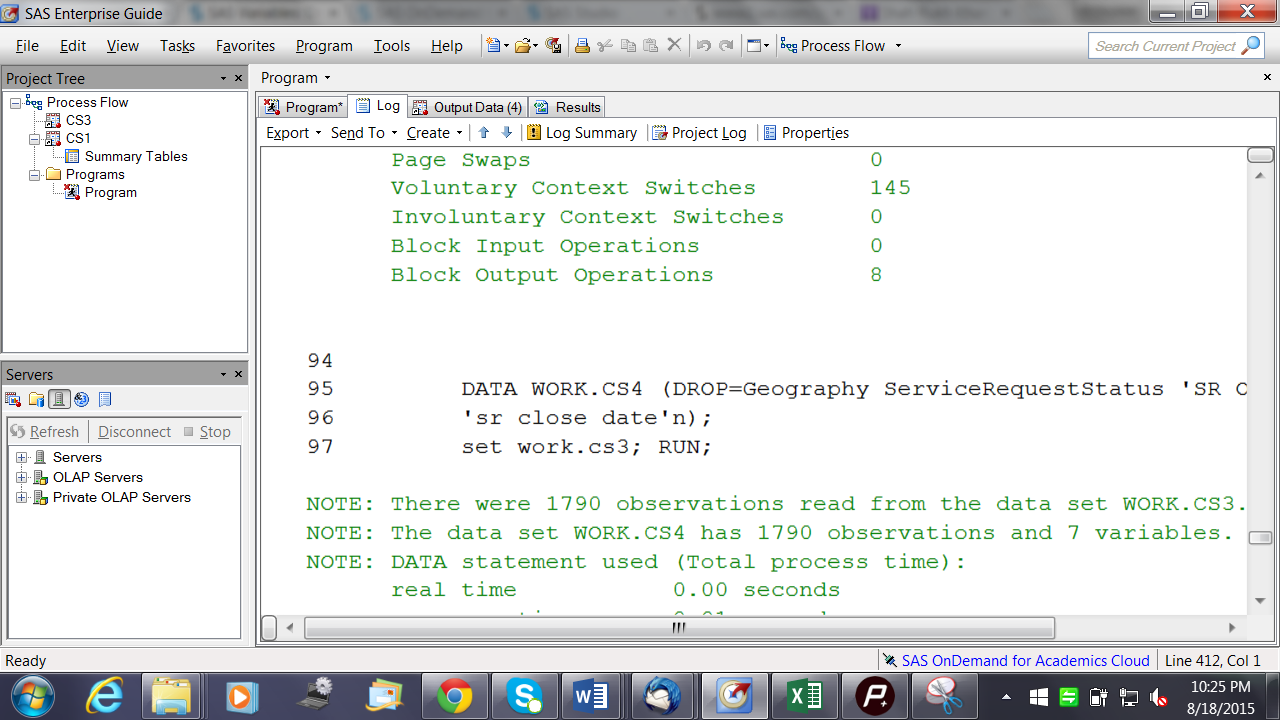
11.5 Convert TypeOfEnagagement, Incident..Problem, SR.Priority, Product, Country into numeric fields (dummy and derived variables).

Dropping variables

DATA WORK.CS4 (DROP=Geography ServiceRequestStatus 'SR OPEN DATE'n

'sr close date'n);

set work.cs3; RUN;



Creating Dummy variables :-

Tip :- Use an array to create dummy variables in SAS .

A SAS array is a convenient way of temporarily identifying a group of variables for processing within a data step. Once the array has been defined the same tasks can be performed for a series of related variables, the array elements.

DATA WORK.CS5 (DROP = i);

SET WORK.CS4;

ARRAY A {\*} COUNTRY COUNTRY1-COUNTRY18 ;

DO i = 1 TO 18;

A(i) = (COUNTRY=i);

END;

PROC FREQ DATA=CS5;

TABLE COUNTRY1-COUNTRY18;

RUN;

Tasks for you :-

a. Create the Dummy Variables for the rest of the variables

b. For SR Priority field , use the code

DATA WORK.NewDataFileName;

SET WORK.OldDataFileName;

PRIORITY = SUBSTRN('SR PRIORITY'n,1,1);

RUN;

c. Drop the variables from which you have created the Dummy Variables by using code

DATA WORK.NewDataFileName (Drop= Var1 var2 …. Varn);

SET WORK.OldDataFileName;

RUN;

d. Save the Project Datamart on the Desktop using the File < Export option