# SAS INTERVIEW QUESTION

**1) What is the default library in SAS?**

Ans:

Work is a default library in SAS. If we do not specify the library name while creating a dataset in the Data step , SAS saves the dataset in the Work Library. Datasets stored in the Work library are temporary so long as the session remains. Once you log out from the session dataset in the work library are permanently deleted.

**2) Explain Proc Sort Procedure.**

Ans :   Proc Sort procedure is used to sort the dataset in ascending or descending order according to the variable specified in the BY statement.

Syntax:
 Proc sort data=Libraryname.datasetname;
 By variable_name;

**3) What is the difference between NODUPKEY and NODUP options in Sort Procedure.**
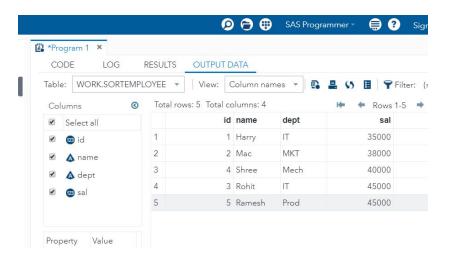
Ans:  **NODUP**

This will remove the duplicate records if all the variables in the records have identical Values.

**NODUPKEY**

This will remove the duplicates By value(where every By variable is duplicated) but does not look at the other fields in the observations.

Output: Observation no 4 is duplicated in all the variables in the observation. Thus it will be removed and record will be sorted in ascending order of sal variable
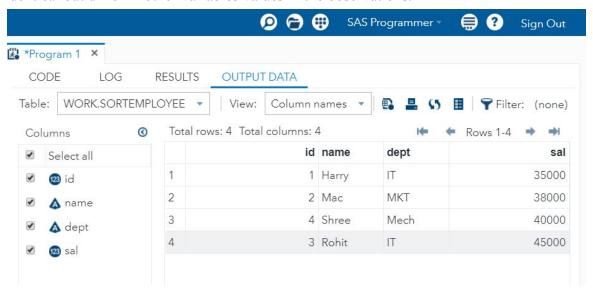


With NODUPKEY: In the following dataset Ramesh and Rohit are having same salary but differ in other variables values

**Output:**

Thus it will remove Ramesh Observation from the Dataset which has only by variable sal values identical but differ in other variables values in the observations.



**4) Describe the Basic structure of the SAS program.**

Ans: Structure of SAS program is divided into two parts Data Step and Proc Step.

    i) In Data step operations such as reading data from external file, creating dataset and manipulation on the data is performed.

    ii) In proc step built in procedure in the sas such are proc print, proc sort etc are implemented in the dataset.

**5) What is the difference between datalines and cards in SAS?**

Ans: When we want to store  the data into dataset entered in the SAS program instead of reading from the external file we use DATALINES and CARDS with the INPUT statement. There is no significant difference between the DATALINES and CARDS.  CARDS is defined as alias for DATALINES(according to the SAS documentation).

**6)  What is the use of  DATALINES4 ?**

Ans: Semicolon is considered as a terminator in the SAS programming, it is the end of your instruction. Whenever an semicolon is encountered by the SAS compiler it considered it as Instruction. While entering data in the sas program in DATALINES semicolon is give at the end of the data as shown below.

Data student;
Input id name$ percentage
Datalines;
1      abc    98.7
2      pqr    99.7
3     xyz    85.3
;

Thus you cannot have any semicolon in between your data. If your data contains semicolon in  between then use DATALINES4 instead of DATALINE.

**7)  How do we combine two or more datasets in SAS?**

Ans: Combining the dataset in SAS can be done in two ways:

1) Appending i.e joining one dataset below another dataset. This is also called as Column or variable combining. Appending is possible only when both the combining dataset has the same  structure  i.e  same  number  of  variables(columns).  We  use  SET  statements  for combining two or more datasets with columns.

   syntax:
   
           data libraryname.datasetname;
           SET  DatasetA  Dataset B;
           run;

2) Merging:
   If there are two datasets containing different variables (i.e number of column) with different datatype and there is a need to combine two dataset to perform analysis on the single large dataset, we use Merging.

**Before Merging:**
1) Both the datasets must contain a common variable having the same datatype.
2) Before Merging, sort both the dataset on the basis of common variables.

**General form of Merging:**
data  libraryname.datsetname;
Merge dataset1 dataset 2……. dataset n;
By variablename; /*Common variable name */
Run;

## 8) What is the difference between SET and MERGE?

Ans: SET is used to CONCATENATE the dataset or APPEND the dataset,while MERGE
matches the observations of the data sets.

## 9)  Explain about SAS automatic _ERROR_ variable?

Ans:
_ERROR_ is a SAS system variable. Its default value is 0. When an error occurs in the sas
Data step this _ERROR_  is automatically set to 1.

## 10) WHat are the datatypes supported in SAS?

Ans: SAS supports only two Types of data types:  1) Numeric (int and float)  2) String
10) What is the difference between proc content and proc summary?

## 11) WHat is the difference between an INPUT and INFILE statement?

Ans:
INFILE statement is used to read the external file into the SAS system while INPUT
statement is used to declare the variables in the to store the data.

## 12) What is the difference between FORMAT and INFORMAT in SAS?

**Ans:**

**Informat**
Informats tell the SAS how to read the data into SAS variables. This is required when we
are importing external data into the SAS system with the input statement in the Data step
in the SAS program.

**Format**
Format tells the SAS how to display or print the  data in the variable on the Output.
Format can be used in the Data step while creating a dataset or can be used in the proc
step while printing the output.

Format assigned in the Data step  while creating the dataset is permanent, while that is assigned in the proc step is temporary.

## 13) How are the dates stored in the SAS?

Ans :

date values in SAS are stored as the number of days since January 1, 1960. This means that stored date values can be negative (if the date is before January 1, 1960) or positive (if the date is after January 1, 1960). For example, the date January 2, 1960 will be stored in SAS as the number 1 and if it is 31 december 1959 then it will be stored as -1.

## 14) How can you limit the variables written to the output dataset in DATA STEP?

Ans:

With the help of KEEP and DROP statement we can limit the number of variables written in the output dataset.
KEEP : Tells the SAS to keep the variables in the output dataset  listed in the KEEP Statement.
DROP : Tells the SAS to DROP the variable or do not include the variables listed in the DROP statement.

## 15) How would you create multiple observations from a single observation?

Ans: With the help of  double trailing @@ we can create multiple observations from a single Observation.

## 16) Mention how to limit decimal places for the variable using PROC MEANS?

Ans: Using the Option MAXDEC=  you can limit the decimal places for the variable using PROC  MEANS.

## 17)  How can we implement the code Reusability in SAS?

Ans:

With the Help of SAS MACROS  we implement the code Reusability in SAS.

## 18) What is SAS?

Ans: SAS (Statistical Analytics System)

SAS is a software used for  advanced analytics, multivariate analyses, business intelligence, data management and predictive analytics.

## 19) What is the difference between PROC MEANS and PROC Summary?

Ans:

The difference between the two procedures is that PROC MEANS produces a report by default on the other hand to produce a report in PROC SUMMARY, you must include a PRINT option in the PROC SUMMARY statement.

**20) Where do you use PROC MEANS over PROC FREQ?**

Ans: PROC MEANS  is used for numerical variables and the PROC FREQ is used for Categorical Variables.