Case study :-

A Retail company has a tracker to track sales names “Sales”. However, there are some returns also in the file names “Sales\_returns” . The company has sales managers who handle different regions in the file “Sales\_manager”.

The company wants to understand the following :-

1. What is the net sales after returns have been factored in ?
2. What % of the total sales do the returns form?
3. Is the returns linked to value of items (cost) ?

**PROC** **IMPORT** DATAFILE='/home/subhashini1/my\_content/Sales.csv'

DBMS=CSV

OUT=WORK.SALES;

**RUN**;

/\* EXPLORE THE DATA \*/

**PROC** **CONTENTS** DATA=WORK.SALES; **RUN** ;

/\* C & O - GET THE RETURNS DATA TOGETHER WITH THE SALES DATA\*/

**PROC** **IMPORT** DATAFILE='/home/subhashini1/my\_content/Sales\_returns.csv'

DBMS=CSV

OUT=WORK.RETURNS;

**RUN**;

**PROC** **CONTENTS** DATA=WORK.RETURNS; **RUN** ;

/\* SORT THE DATASETS ON PRIMARY KEY \*/

**PROC** **SORT** DATA=WORK.SALES;

BY "Order ID"N; **RUN** ;

**PROC** **SORT** DATA=WORK.RETURNS;

BY "Order ID"N; **RUN** ;

**DATA** WORK.TOTAL;

MERGE WORK.SALES WORK.RETURNS;

BY "Order ID"N; **RUN** ;

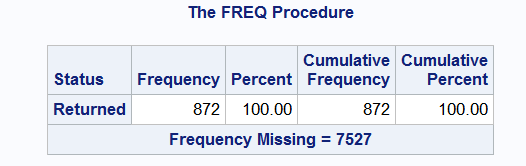
**PROC** **PRINT** DATA=WORK.TOTAL (OBS=**10**) ; **RUN** ;

/\* V - RUN FREQ TABLE TO UNDERSTAND DATA SPLIT IN 'STATUS' BETWEEN

RETURN AND NOTRETURN\*/

**PROC** **FREQ** DATA=WORK.TOTAL ;

TABLES STATUS ; **RUN** ;



**DATA** WORK.TOTAL;

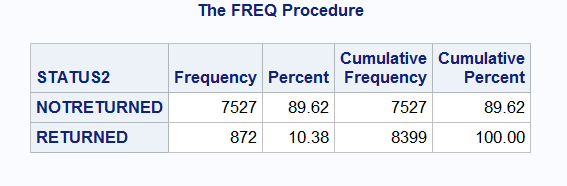
SET WORK.TOTAL ;

IF STATUS = 'Returned' THEN STATUS2= 'RETURNED';

ELSE STATUS2 = 'NOTRETURNED'; **RUN** ;

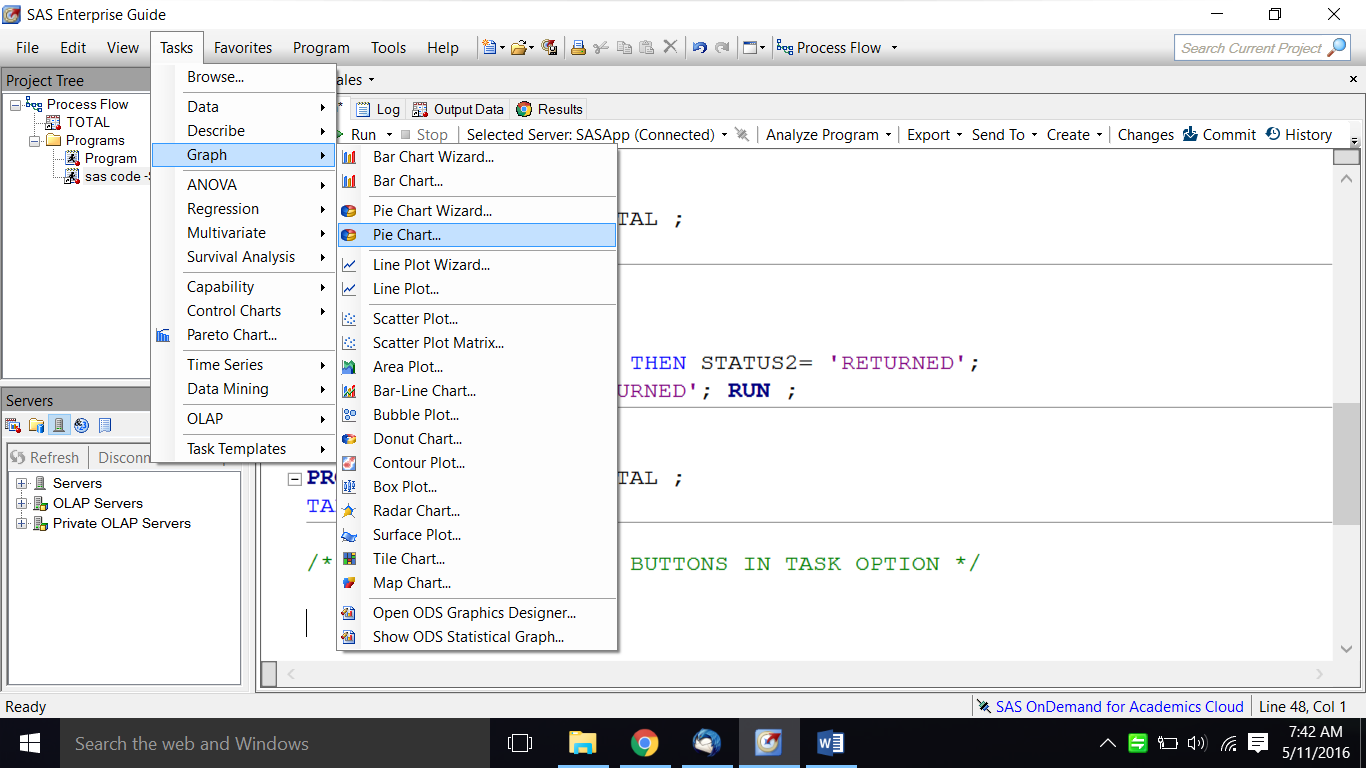
**PROC** **FREQ** DATA=WORK.TOTAL ;

TABLES STATUS2 ; **RUN** ;

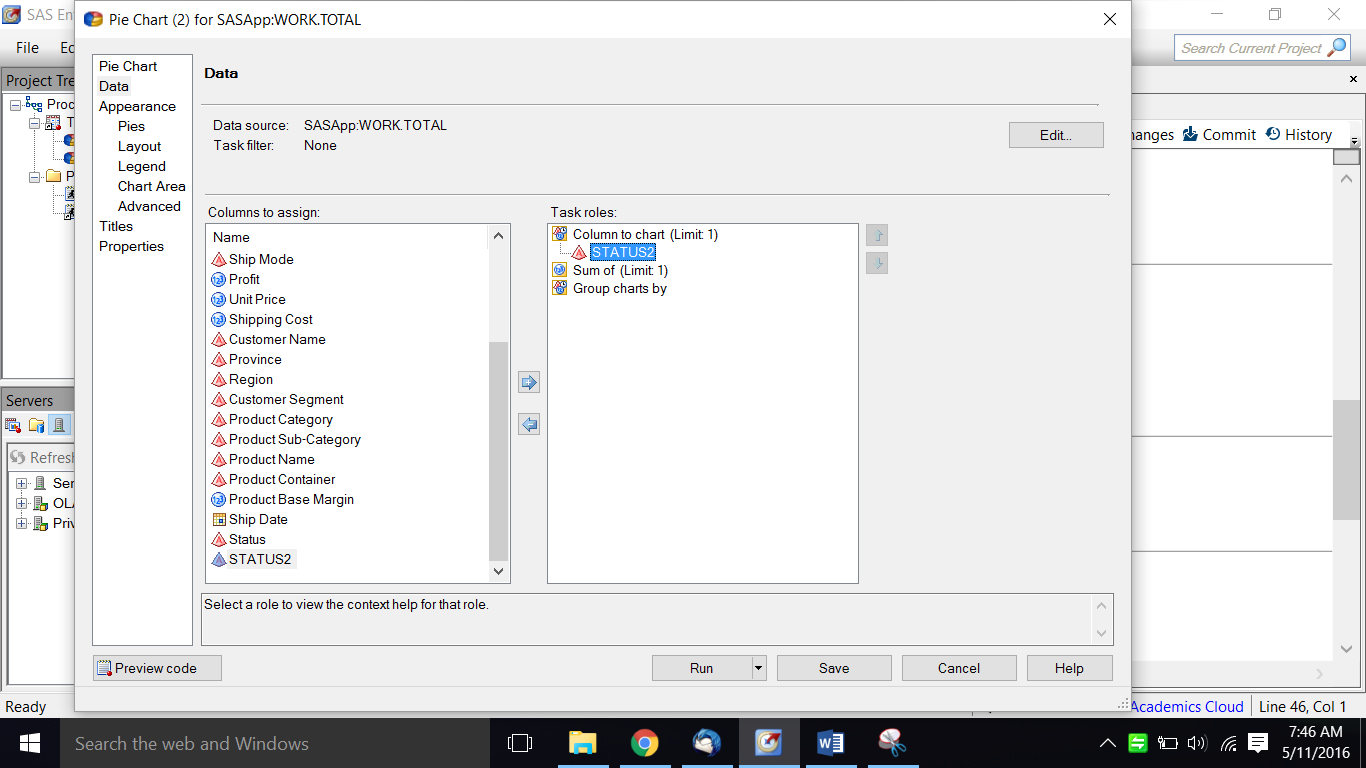


USE THE BUTTON DRIVEN MENUE TO EXECUTE THE PIE CHART

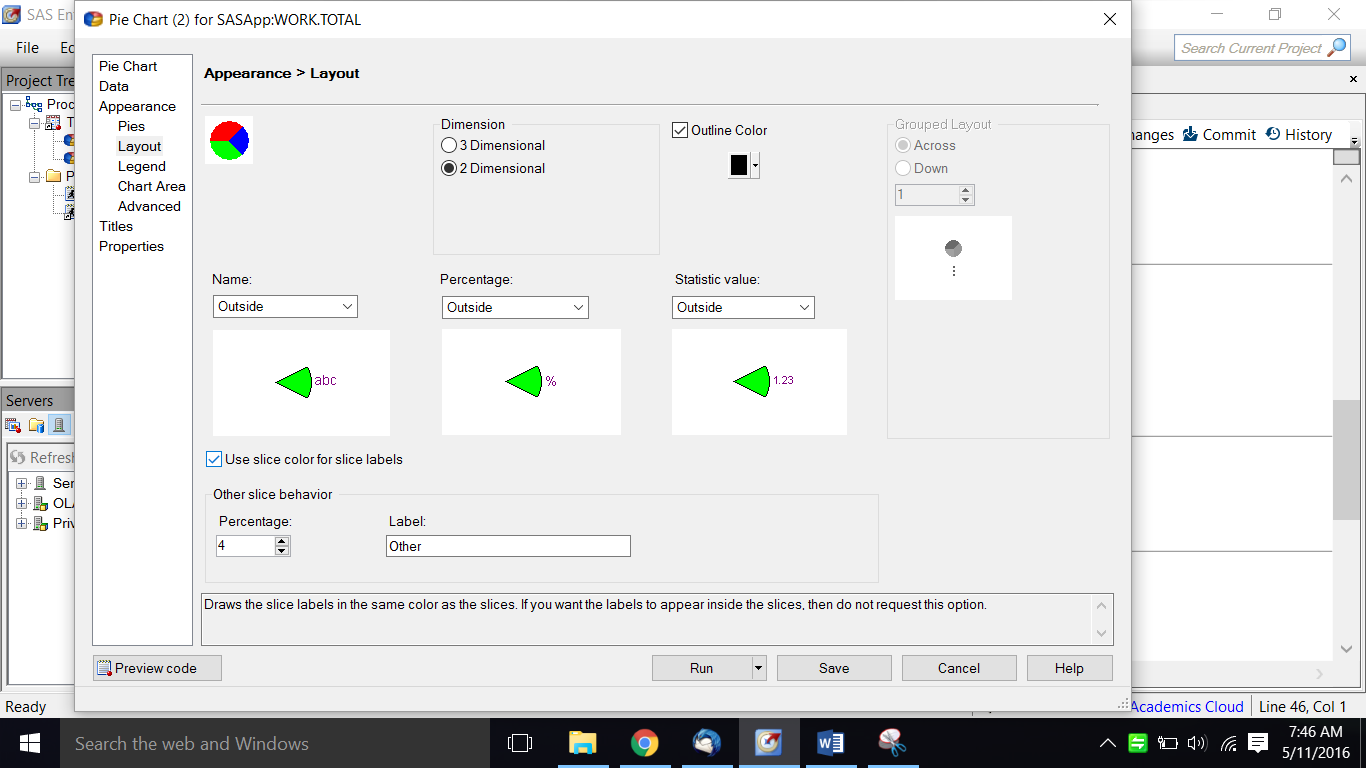
TASKS > GRAPHS > PIE CHART

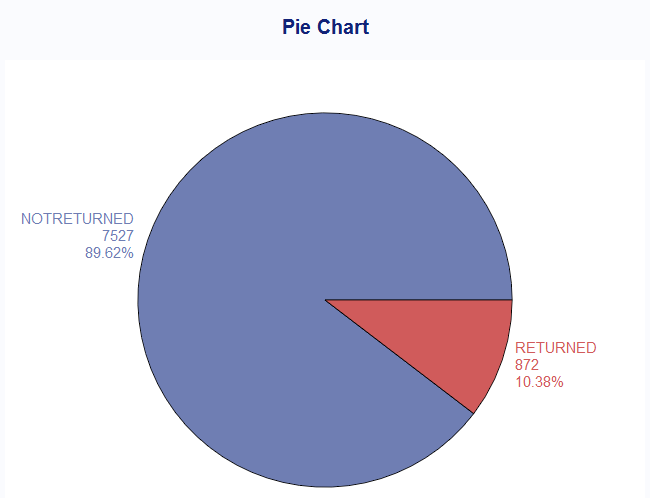


SELECT THE DATA



SELECT THE LAYOUT





Thus 10.38% is the count of returns to total Sales

/\* ALTERNATE METHOD - RUN PIE CHART USING CODE\*/

/\* Set the graphics environment \*/

goptions reset=all cback=white border htitle=**12**pt htext=**10**pt;

title1 "RETURNS IN SALES";

**proc** **gchart** data=work.total;

pie Status2 / other=**0**

midpoints="RETURNED" "NOTRETURNED"

value=none

percent=arrow

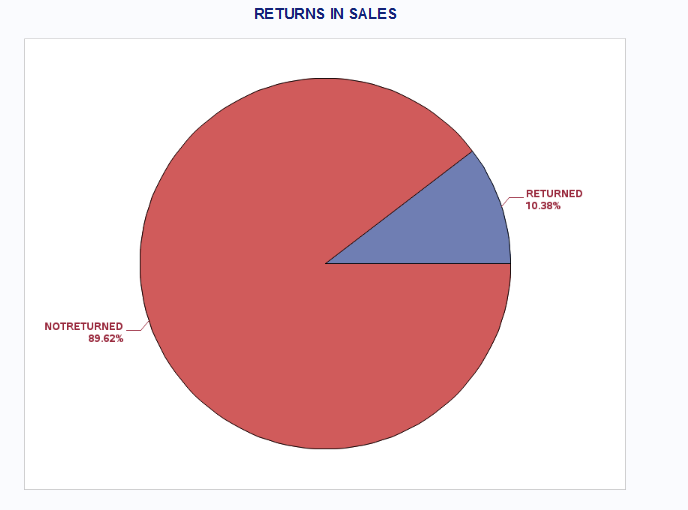
slice=arrow

noheading

plabel=(font='Albany AMT/bold' h=**1.3** color=depk);

**run**;

**quit**;



/\* CHECK BY VALUE OF SALES\*/

/\* Set the graphics environment \*/

goptions reset=all cback=white border htitle=**12**pt htext=**10**pt;

title1 "RETURNS IN SALES";

**proc** **gchart** data=work.total;

pie Status2 / SUMVAR=SALES

midpoints="RETURNED" "NOTRETURNED"

value=none

percent=arrow

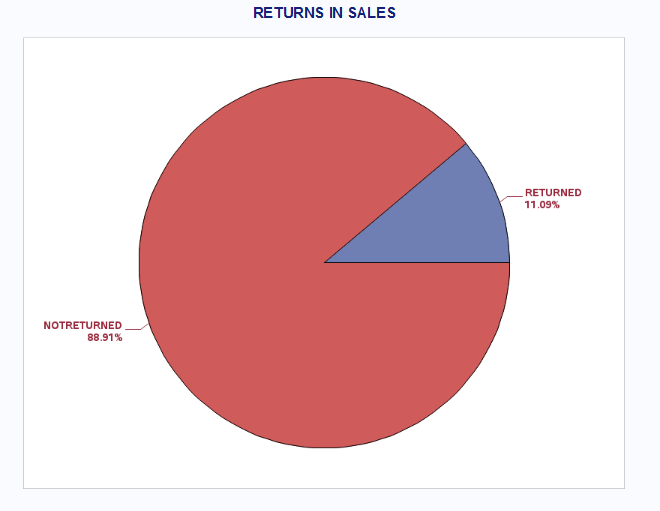
slice=arrow

noheading

plabel=(font='Albany AMT/bold' h=**1.3** color=depk);

**run**;

**quit**;



Thus there are returns totalling to 11.09% of the Sales values

/\*A - CORRELATION \*/

**DATA** WORK.TOTAL ;

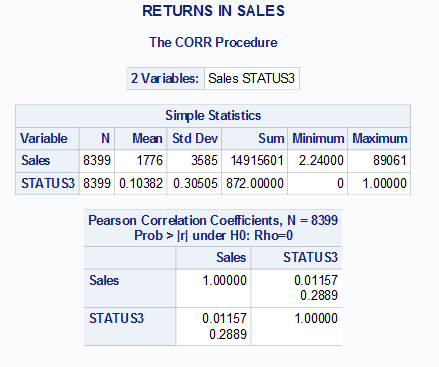
SET WORK.TOTAL ;

IF STATUS2 = 'RETURNED' THEN STATUS3 = **1**;

ELSE STATUS2 = **0** ; **RUN** ;

**PROC** **CORR** DATA=WORK.TOTAL;

VAR SALES STATUS3; **RUN** ;



# Insight

1. 10% of the total count of goods sold is returned
2. 11% of the goods is returned by value.
3. The correlation between higher sales value of goods and returns is very low (.011) . Since the p value of the correlation is high (.22) we cannot conclude that this value of correlation is not by coincidence . Therefore all evidence points to no correlation between value of goods and returns