Case Study on Graphs and Charts using SAS

Fraud is the misappropriation of funds . This case study deals with understanding the data thru graphs and charts to pin point possible fraud cases.

About the data :- Council contracts register; Contract expenditure over $5000

What is this data?

This dataset lists details of all procurement spending over $5,000 including purchase orders and tenders.

Definitions:- The dataset contains the following fields that you should use :

|  |  |
| --- | --- |
| Effective Date | Date of data pull |
| Contract Reference Number | Contract id |
| Service responsible | Area of contract |
| Contract Start Date | Start date |
| End Date | End date . If missing, assume ongoing |
| Review Date | Date of supervisory review |
| Extension Period | No of days |
| Contract Value | Amount |

Problem statement :-

Do an exploratory data analysis to understand the data better. Describe the date and number fields and understand the distributions .

Create contingency tables, graphs and charts to better understand the data.

Do some projects finish significantly earlier that the date expected ?

What is the per day rate at an average for projects ?

Note :- Fraud in these contracts can be of 2 types .

1. If the project can be completed in say 10 days but the contractor quotes 25 days . He then finished the work in 10 days and pockets the money for the 15 days .

2. If the project can be completed @ $1 per day of work and the contractor charges $5 per day of work

Solution in SAS

1. Define :- The project is to create graphs and charts to undertstand cases which could signify fraud defined as

1.1 Lower than normal time to complete the project

1.2 Higher than normal cost per delivery day of project

2. Collect :- The data is given. No other data is required to be referenced

3. Organise :-

3.1 Missing Values

3.2 Outliers

3.3 Non numeric to numeric data

4. Visualise the Y variables across significant segments of 'Service Responsible'

**SAS Code and Solution**

Import data :-

/\* IMPORT DATA USING INFILE INFORMAT STATMENT \*/

/\* THIS CODE WILL GET AUTO GENERATED IF YOU USE THE IMPORT DATA BUTTON

UNDER THE FILE TAB ON TOP LEFT CORNER OF SCREEN\*/

**DATA** WORK.CONTRACTS;

LENGTH

Effective\_Date $ **11**

Contract\_Reference\_Number $ **8**

Title\_of\_agreement $ **94**

Service\_responsible $ **49**

Description\_of\_goods\_and\_service $ **94**

Contract\_Start\_Date **8**

End\_Date $ **10**

Review\_Date $ **10**

Extension\_Period **8**

Contract\_Value **8**

SupplierName $ **40**

NominatedContactPoint $ **37**

;

LABEL

Effective\_Date = "Effective Date"

Contract\_Reference\_Number = "Contract Reference Number"

Title\_of\_agreement = "Title of agreement"

Service\_responsible = "Service responsible"

Description\_of\_goods\_and\_service = "Description of goods and services"

Contract\_Start\_Date = "Contract Start Date"

End\_Date = "End Date"

Review\_Date = "Review Date"

Extension\_Period = "Extension Period"

Contract\_Value = "Contract Value" ;

FORMAT

Effective\_Date $CHAR11.

Contract\_Reference\_Number $CHAR8.

Title\_of\_agreement $CHAR94.

Service\_responsible $CHAR49.

Description\_of\_goods\_and\_service $CHAR94.

Contract\_Start\_Date MMDDYY10.

End\_Date $CHAR10.

Review\_Date $CHAR10.

Extension\_Period BEST2.

Contract\_Value BEST11.

SupplierName $CHAR40.

NominatedContactPoint $CHAR37.

;

INFORMAT

Effective\_Date $CHAR11.

Contract\_Reference\_Number $CHAR8.

Title\_of\_agreement $CHAR94.

Service\_responsible $CHAR49.

Description\_of\_goods\_and\_service $CHAR94.

Contract\_Start\_Date MMDDYY10.

End\_Date $CHAR10.

Review\_Date $CHAR10.

Extension\_Period BEST2.

Contract\_Value BEST11.

SupplierName $CHAR40.

NominatedContactPoint $CHAR37.

;

INFILE '/saswork/SAS\_work6710000127DB\_odaws02-prod-sg/#LN00030'

LRECL=**332**

ENCODING="UTF-8"

TERMSTR=CRLF

DLM='7F'x

MISSOVER

DSD ;

INPUT

Effective\_Date : $CHAR11.

Contract\_Reference\_Number : $CHAR8.

Title\_of\_agreement : $CHAR94.

Service\_responsible : $CHAR49.

Description\_of\_goods\_and\_service : $CHAR94.

Contract\_Start\_Date : ?? MMDDYY10.

End\_Date : $CHAR10.

Review\_Date : $CHAR10.

Extension\_Period : ?? BEST2.

Contract\_Value : ?? COMMA11.

SupplierName : $CHAR40.

NominatedContactPoint : $CHAR37.

;

**RUN**;

/\* EXPLANATION:-

LENGTH – sets the length for each variable ; $ symbol shows it’s a character variable

LABEL – sets the explanation for the variables; can be seen in the PROC CONTENTS output

FORMAT – controls written appearance of the variable values

INFORMAT – reads data into SAS

INFILE – gives details about the datafile that we are importing (external file)

DSD (delimiter-sensitive data)

LRECL=***logical-record-length***specifies the logical record length.

MISSOVER =prevents an INPUT statement from reading a new input data record if it does not find values in the current input line for all the variables in the statement.

ENCODING= specifies the encoding to use when reading from the external file (this is optional )

INPUT – creates input record in input buffer (from where it will be inserted in a sas datafile)\*/

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

NOTE ALTERNATE IMPORT CODE :-

FILENAME REFFILE "/home/subhashini1/my\_content/Contacts register Aug 2015 and purchase order over 5000 April to June 2015.csv" TERMSTR=CR;

**PROC** **IMPORT** DATAFILE=REFFILE

DBMS=CSV

OUT=WORK.IMPORT;

GETNAMES=YES;

**RUN**;

DATA WORK.CONTRACTS;

SET WORK.IMPORT;RUN;

View the contents of the SAS file and understand the variables

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



Note :- You can easily copy paste the outputs from the RESULTS window into an excel sheet !! You can then use it to create graphs and charts in excel – incase you so wish .

3. Keep required variables

/\* KEEP ONLY THOSE VARIABLES THAT ARE AVAILABLE IN THE

DEFINITIONS GIVEN IN THE PROJECT\*/

**DATA** WORK.CONTRACTS1 (KEEP=Contract\_Reference\_Number

Contract\_Start\_Date

Contract\_Value

End\_Date

Extension\_Period

Review\_Date

Service\_responsible);

SET WORK.CONTRACTS; **RUN** ;

**PROC** **PRINT** DATA=WORK.CONTRACTS1 (FIRSTOBS=**70** OBS=**75**); **RUN** ;

4. Create required Fraud indicator variables (y)

/\* CREATE Y VARIABLES

TAT = END DATE - START DATE

PERDAY = CONTACT AMOUNT / TAT

NOTE - END DATE IS CHARACTER (AS SEEN IN PROC CONTENTS)\*/

**DATA** WORK.CONTRACTS1 ;

SET WORK.CONTRACTS1 ;

END\_DATE1=INPUT(End\_Date, MMDDYY10.);

FORMAT END\_DATE1 MMDDYY10.; **RUN**;

**PROC** **PRINT** DATA=WORK.CONTRACTS1 (OBS=**10**); **RUN**;

**PROC** **CONTENTS** DATA=WORK.CONTRACTS1; **RUN**;



**DATA** WORK.CONTRACTS1;

SET WORK.CONTRACTS1;

TAT= END\_DATE1-Contract\_Start\_Date;

PERDAY = Contract\_Value/TAT; **RUN**;

**PROC** **PRINT** DATA=WORK.CONTRACTS1 (OBS=**10**);

VAR TAT PERDAY; **RUN**;

5. C and O :-

/\* C - NO ACTIVITY REQUIRED

O - MISSING VALUES

CHECK MISSING VALUES FOR NUMERIC VARIABLES USING PROC MEANS

CHECK MISSING VALUES FOR CHARACTER VARIABLES USING PROC FREQ\*/

**PROC** **MEANS** DATA=WORK.CONTRACTS1; **RUN**;

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **The MEANS Procedure** | | | | | | |
| **Variable** | **Label** | **N** | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| Contract\_Start\_Date | Contract Start Date | 194 | 19968.74 | 457.330218 | 17623 | 20362 |
| Extension\_Period | Extension Period | 84 | 0.75 | 1.7276978 | 0 | 14 |
| Contract\_Value | Contract Value | 194 | 702287.9 | 4524217.01 | 0 | 49000000 |
| END\_DATE1 |  | 83 | 20936.33 | 671.611914 | 20331 | 25383 |
| TAT |  | 83 | 1290.34 | 716.242691 | 363 | 5478 |
| PERDAY |  | 83 | 1167.87 | 6794.94 | 0 | 59157.89 |

**PROC** **FREQ** DATA=WORK.CONTRACTS1;

TABLES Service\_responsible; **RUN**;

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **The FREQ Procedure** | | | | |
| **Service responsible** | | | | |
| **Service\_responsible** | **Frequency** | **Percent** | **Cumulative** | **Cumulative** |
| **Frequency** | **Percent** |
| **CHIEF EXECUTIVE** | 1 | 0.52 | 1 | 0.52 |
| **COMMUNITY** | 21 | 10.82 | 22 | 11.34 |
| **Community** | 3 | 1.55 | 25 | 12.89 |
| **DEMOCRATIC LEGAL AND POLICY** | 12 | 6.19 | 37 | 19.07 |
| **Democratic Legal & Policy** | 2 | 1.03 | 39 | 20.1 |
| **ENVIRONMENT** | 10 | 5.15 | 49 | 25.26 |
| **Environment** | 12 | 6.19 | 61 | 31.44 |
| **FINANCE & COMMERCIAL** | 23 | 11.86 | 84 | 43.3 |
| **Finance & Commercial** | 31 | 15.98 | 115 | 59.28 |
| **HOUSING** | 5 | 2.58 | 120 | 61.86 |
| **HUMAN RESOURCES ICT & SHARED SUPPORT** | 19 | 9.79 | 139 | 71.65 |
| **Human Resources ICT/CSC & Shared Support Services** | 25 | 12.89 | 164 | 84.54 |
| **PLANNING** | 8 | 4.12 | 172 | 88.66 |
| **PROPERTY** | 11 | 5.67 | 183 | 94.33 |
| **Planning & Sustainability** | 11 | 5.67 | 194 | 100 |

**DATA** WORK.CONTRACTS1 (DROP= End\_Date

Extension\_Period

Review\_Date);

SET WORK.CONTRACTS1; **RUN**;

/\* IN THE PROC FREQ OUTPUT WE CAN SEE THAT BECAUSE OF CASE DIFFERENCE

SOME SEGMENTS ARE COMING TWICE EG :- COMMUNITY / Community;

CONVERT ALL THE DATA TO UPPER CASE TO STANDARDISE OUTPUT

WE CAN ALSO SEE THAT SOME MINOR SPELLING DIFFERENCE IN CAUSING DIFFERENT

SEGMENTS EG:- DEMOCRATIC LEGAL AND POLICY /Democratic Legal & Policy.

CREATE SAME SPELLING ACROSS \*/

**DATA** WORK.CONTRACTS1;

SET WORK.CONTRACTS1;

IF Service\_responsible IN ('Democratic Legal & Policy')

THEN Service\_responsible= 'DEMOCRATIC LEGAL AND POLICY';

IF Service\_responsible IN ('Human Resources ICT/CSC & Shared Support Services')

THEN Service\_responsible= 'HUMAN RESOURCES ICT & SHARED SUPPORT';

IF Service\_responsible IN ('Planning & Sustainability')

THEN Service\_responsible= 'PLANNING';

**RUN**;

**DATA** WORK.CONTRACTS1;

SET WORK.CONTRACTS1;

Service\_responsible= UPCASE(Service\_responsible); **RUN**;

**PROC** **FREQ** DATA=WORK.CONTRACTS1;

TABLES Service\_responsible; **RUN**;

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **The FREQ Procedure** | | | | |
| **Service responsible** | | | | |
| **Service\_responsible** | **Frequency** | **Percent** | **Cumulative** | **Cumulative** |
| **Frequency** | **Percent** |
| **CHIEF EXECUTIVE** | 1 | 0.52 | 1 | 0.52 |
| **COMMUNITY** | 24 | 12.37 | 25 | 12.89 |
| **DEMOCRATIC LEGAL AND POLICY** | 14 | 7.22 | 39 | 20.1 |
| **ENVIRONMENT** | 22 | 11.34 | 61 | 31.44 |
| **FINANCE & COMMERCIAL** | 54 | 27.84 | 115 | 59.28 |
| **HOUSING** | 5 | 2.58 | 120 | 61.86 |
| **HUMAN RESOURCES ICT & SHARED SUPPORT** | 44 | 22.68 | 164 | 84.54 |
| **PLANNING** | 19 | 9.79 | 183 | 94.33 |
| **PROPERTY** | 11 | 5.67 | 194 | 100 |

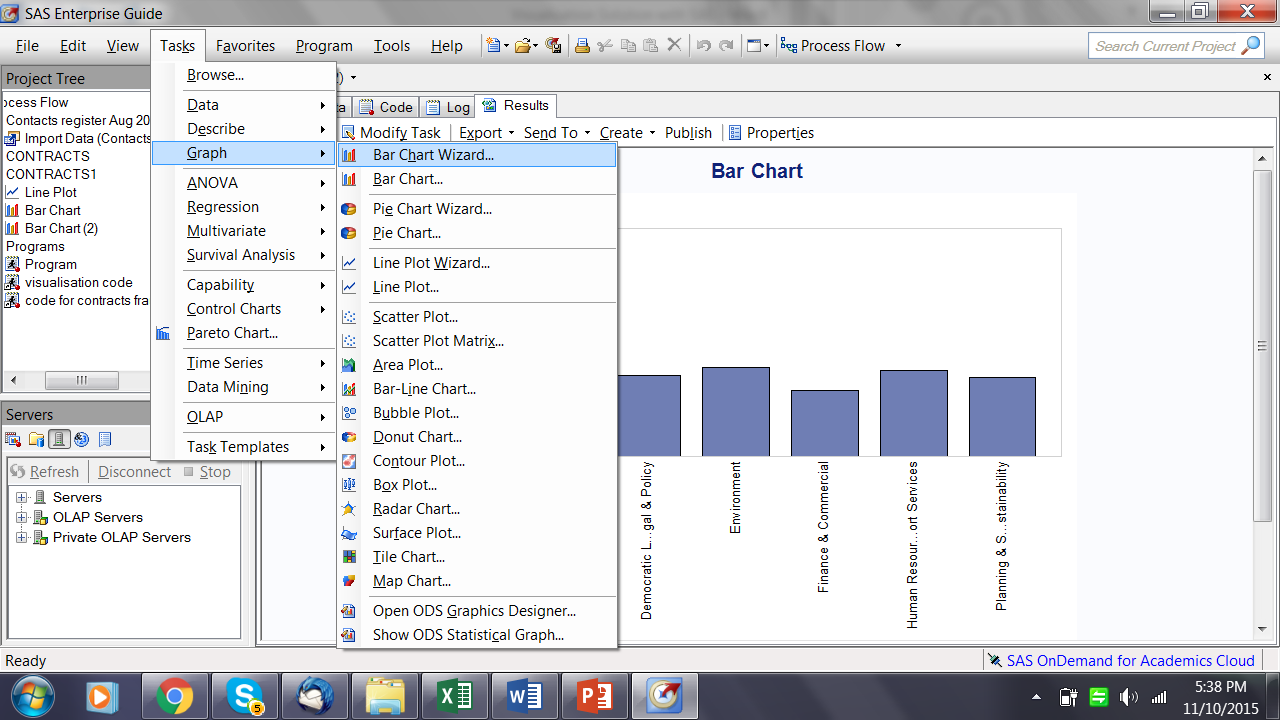
6. Visualisation :-

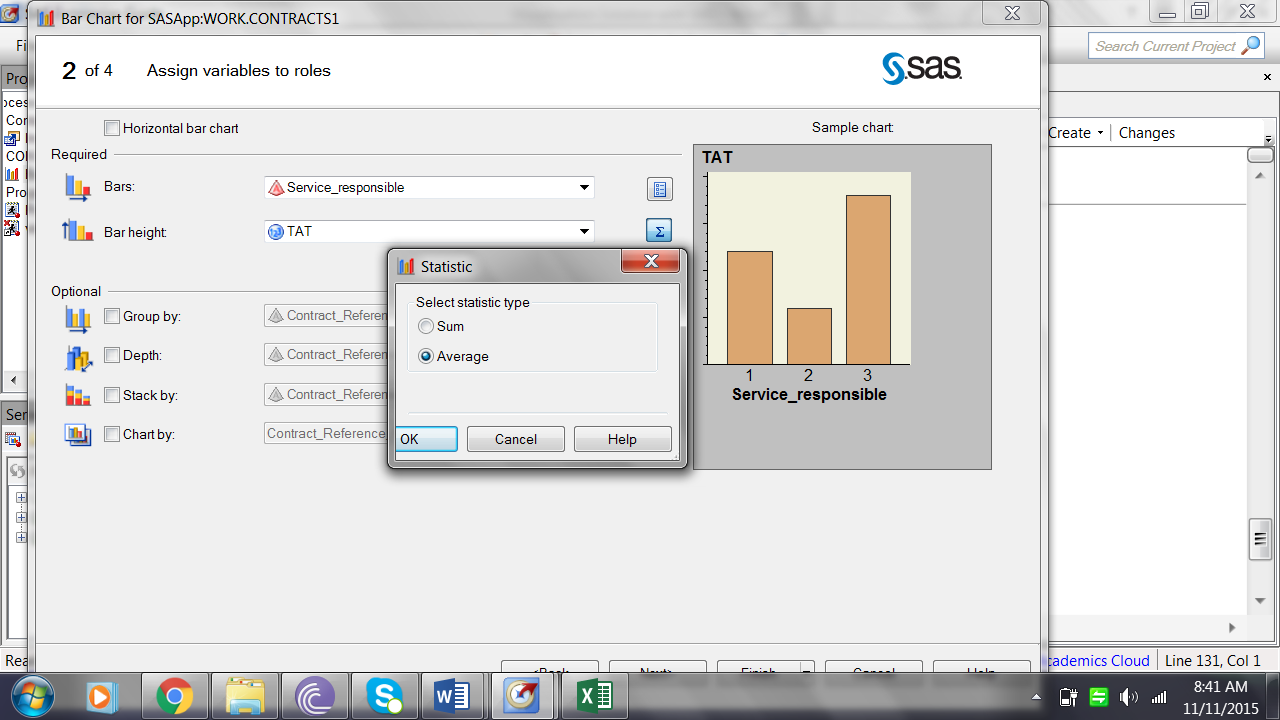
a. Bar Chart :- To understand Average value of TAT across different segments of Service\_responsible

Method 1 :- Use the buttons on top of the screen

Tasks > Graph> Bar Chart Wizard

This will open the wizard . Just use the drag and drop functionality to create the graph





The Bar chart appears Case Study on Graphs and Charts using SAS

Fraud is the misappropriation of funds . This case study deals with understanding the data thru graphs and charts to pin point possible fraud cases.

About the data :- Council contracts register; Contract expenditure over $5000

What is this data?

This dataset lists details of all procurement spending over $5,000 including purchase orders and tenders.

Definitions:- The dataset contains the following fields that you should use :

|  |  |
| --- | --- |
| Effective Date | Date of data pull |
| Contract Reference Number | Contract id |
| Service responsible | Area of contract |
| Contract Start Date | Start date |
| End Date | End date . If missing, assume ongoing |
| Review Date | Date of supervisory review |
| Extension Period | No of days |
| Contract Value | Amount |

Problem statement :-

Do an exploratory data analysis to understand the data better. Describe the date and number fields and understand the distributions .

Create contingency tables, graphs and charts to better understand the data.

Do some projects finish significantly earlier that the date expected ?

What is the per day rate at an average for projects ?

Segregate the cases that should be put up for investigation for fraud .

Note :- Fraud in these contracts can be of 2 types .

1. If the project can be completed in say 10 days but the contractor quotes 25 days . He then finished the work in 10 days and pockets the money for the 15 days .

2. If the project can be completed @ $1 per day of work and the contractor charges $5 per day of work

Solution in SAS

1. Define :- The project is to create graphs and charts to undertstand cases which could signify fraud defined as

1.1 Lower than normal time to complete the project

1.2 Higher than normal cost per delivery day of project

2. Collect :- The data is given. No other data is required to be referenced

3. Organise :-

3.1 Missing Values

3.2 Outliers

3.3 Non numeric to numeric data

4. Visualise the Y variables across significant segments of 'Service Responsible'

**SAS Code and Solution**

Import data :-

/\* IMPORT DATA USING INFILE INFORMAT STATMENT \*/

/\* THIS CODE WILL GET AUTO GENERATED IF YOU USE THE IMPORT DATA BUTTON

UNDER THE FILE TAB ON TOP LEFT CORNER OF SCREEN\*/

**DATA** WORK.CONTRACTS;

LENGTH

Effective\_Date $ **11**

Contract\_Reference\_Number $ **8**

Title\_of\_agreement $ **94**

Service\_responsible $ **49**

Description\_of\_goods\_and\_service $ **94**

Contract\_Start\_Date **8**

End\_Date $ **10**

Review\_Date $ **10**

Extension\_Period **8**

Contract\_Value **8**

SupplierName $ **40**

NominatedContactPoint $ **37**

;

LABEL

Effective\_Date = "Effective Date"

Contract\_Reference\_Number = "Contract Reference Number"

Title\_of\_agreement = "Title of agreement"

Service\_responsible = "Service responsible"

Description\_of\_goods\_and\_service = "Description of goods and services"

Contract\_Start\_Date = "Contract Start Date"

End\_Date = "End Date"

Review\_Date = "Review Date"

Extension\_Period = "Extension Period"

Contract\_Value = "Contract Value" ;

FORMAT

Effective\_Date $CHAR11.

Contract\_Reference\_Number $CHAR8.

Title\_of\_agreement $CHAR94.

Service\_responsible $CHAR49.

Description\_of\_goods\_and\_service $CHAR94.

Contract\_Start\_Date MMDDYY10.

End\_Date $CHAR10.

Review\_Date $CHAR10.

Extension\_Period BEST2.

Contract\_Value BEST11.

SupplierName $CHAR40.

NominatedContactPoint $CHAR37.

;

INFORMAT

Effective\_Date $CHAR11.

Contract\_Reference\_Number $CHAR8.

Title\_of\_agreement $CHAR94.

Service\_responsible $CHAR49.

Description\_of\_goods\_and\_service $CHAR94.

Contract\_Start\_Date MMDDYY10.

End\_Date $CHAR10.

Review\_Date $CHAR10.

Extension\_Period BEST2.

Contract\_Value BEST11.

SupplierName $CHAR40.

NominatedContactPoint $CHAR37.

;

INFILE '/saswork/SAS\_work6710000127DB\_odaws02-prod-sg/#LN00030'

LRECL=**332**

ENCODING="UTF-8"

TERMSTR=CRLF

DLM='7F'x

MISSOVER

DSD ;

INPUT

Effective\_Date : $CHAR11.

Contract\_Reference\_Number : $CHAR8.

Title\_of\_agreement : $CHAR94.

Service\_responsible : $CHAR49.

Description\_of\_goods\_and\_service : $CHAR94.

Contract\_Start\_Date : ?? MMDDYY10.

End\_Date : $CHAR10.

Review\_Date : $CHAR10.

Extension\_Period : ?? BEST2.

Contract\_Value : ?? COMMA11.

SupplierName : $CHAR40.

NominatedContactPoint : $CHAR37.

;

**RUN**;

/\* EXPLANATION:-

LENGTH – sets the length for each variable ; $ symbol shows it’s a character variable

LABEL – sets the explanation for the variables; can be seen in the PROC CONTENTS output

FORMAT – controls written appearance of the variable values

INFORMAT – reads data into SAS

INFILE – gives details about the datafile that we are importing (external file)

DSD (delimiter-sensitive data)

LRECL=***logical-record-length*** specifies the logical record length.

MISSOVER =prevents an INPUT statement from reading a new input data record if it does not find values in the current input line for all the variables in the statement.

ENCODING= specifies the encoding to use when reading from the external file (this is optional )

INPUT – creates input record in input buffer (from where it will be inserted in a sas datafile)\*/

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

NOTE ALTERNATE IMPORT CODE :-

FILENAME REFFILE "/home/subhashini1/my\_content/Contacts register Aug 2015 and purchase order over 5000 April to June 2015.csv" TERMSTR=CR;

**PROC** **IMPORT** DATAFILE=REFFILE

DBMS=CSV

OUT=WORK.IMPORT;

GETNAMES=YES;

**RUN**;

DATA WORK.CONTRACTS;

SET WORK.IMPORT;RUN;

View the contents of the SAS file and understand the variables

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



Note :- You can easily copy paste the outputs from the RESULTS window into an excel sheet !! You can then use it to create graphs and charts in excel – incase you so wish .

3. Keep required variables

/\* KEEP ONLY THOSE VARIABLES THAT ARE AVAILABLE IN THE

DEFINITIONS GIVEN IN THE PROJECT\*/

**DATA** WORK.CONTRACTS1 (KEEP=Contract\_Reference\_Number

Contract\_Start\_Date

Contract\_Value

End\_Date

Extension\_Period

Review\_Date

Service\_responsible);

SET WORK.CONTRACTS; **RUN** ;

**PROC** **PRINT** DATA=WORK.CONTRACTS1 (FIRSTOBS=**70** OBS=**75**); **RUN** ;

4. Create required Fraud indicator variables (y)

/\* CREATE Y VARIABLES

TAT = END DATE - START DATE

PERDAY = CONTACT AMOUNT / TAT

NOTE - END DATE IS CHARACTER (AS SEEN IN PROC CONTENTS)\*/

**DATA** WORK.CONTRACTS1 ;

SET WORK.CONTRACTS1 ;

END\_DATE1=INPUT(End\_Date, MMDDYY10.);

FORMAT END\_DATE1 MMDDYY10.; **RUN**;

**PROC** **PRINT** DATA=WORK.CONTRACTS1 (OBS=**10**); **RUN**;

**PROC** **CONTENTS** DATA=WORK.CONTRACTS1; **RUN**;



**DATA** WORK.CONTRACTS1;

SET WORK.CONTRACTS1;

TAT= END\_DATE1-Contract\_Start\_Date;

PERDAY = Contract\_Value/TAT; **RUN**;

**PROC** **PRINT** DATA=WORK.CONTRACTS1 (OBS=**10**);

VAR TAT PERDAY; **RUN**;

5. C and O :-

/\* C - NO ACTIVITY REQUIRED

O - MISSING VALUES

CHECK MISSING VALUES FOR NUMERIC VARIABLES USING PROC MEANS

CHECK MISSING VALUES FOR CHARACTER VARIABLES USING PROC FREQ\*/

**PROC** **MEANS** DATA=WORK.CONTRACTS1; **RUN**;

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **The MEANS Procedure** | | | | | | |
| **Variable** | **Label** | **N** | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| Contract\_Start\_Date | Contract Start Date | 194 | 19968.74 | 457.330218 | 17623 | 20362 |
| Extension\_Period | Extension Period | 84 | 0.75 | 1.7276978 | 0 | 14 |
| Contract\_Value | Contract Value | 194 | 702287.9 | 4524217.01 | 0 | 49000000 |
| END\_DATE1 |  | 83 | 20936.33 | 671.611914 | 20331 | 25383 |
| TAT |  | 83 | 1290.34 | 716.242691 | 363 | 5478 |
| PERDAY |  | 83 | 1167.87 | 6794.94 | 0 | 59157.89 |

**PROC** **FREQ** DATA=WORK.CONTRACTS1;

TABLES Service\_responsible; **RUN**;

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **The FREQ Procedure** | | | | |
| **Service responsible** | | | | |
| **Service\_responsible** | **Frequency** | **Percent** | **Cumulative** | **Cumulative** |
| **Frequency** | **Percent** |
| **CHIEF EXECUTIVE** | 1 | 0.52 | 1 | 0.52 |
| **COMMUNITY** | 21 | 10.82 | 22 | 11.34 |
| **Community** | 3 | 1.55 | 25 | 12.89 |
| **DEMOCRATIC LEGAL AND POLICY** | 12 | 6.19 | 37 | 19.07 |
| **Democratic Legal & Policy** | 2 | 1.03 | 39 | 20.1 |
| **ENVIRONMENT** | 10 | 5.15 | 49 | 25.26 |
| **Environment** | 12 | 6.19 | 61 | 31.44 |
| **FINANCE & COMMERCIAL** | 23 | 11.86 | 84 | 43.3 |
| **Finance & Commercial** | 31 | 15.98 | 115 | 59.28 |
| **HOUSING** | 5 | 2.58 | 120 | 61.86 |
| **HUMAN RESOURCES ICT & SHARED SUPPORT** | 19 | 9.79 | 139 | 71.65 |
| **Human Resources ICT/CSC & Shared Support Services** | 25 | 12.89 | 164 | 84.54 |
| **PLANNING** | 8 | 4.12 | 172 | 88.66 |
| **PROPERTY** | 11 | 5.67 | 183 | 94.33 |
| **Planning & Sustainability** | 11 | 5.67 | 194 | 100 |

**DATA** WORK.CONTRACTS1 (DROP= End\_Date

Extension\_Period

Review\_Date);

SET WORK.CONTRACTS1; **RUN**;

/\* IN THE PROC FREQ OUTPUT WE CAN SEE THAT BECAUSE OF CASE DIFFERENCE

SOME SEGMENTS ARE COMING TWICE EG :- COMMUNITY / Community;

CONVERT ALL THE DATA TO UPPER CASE TO STANDARDISE OUTPUT

WE CAN ALSO SEE THAT SOME MINOR SPELLING DIFFERENCE IN CAUSING DIFFERENT

SEGMENTS EG:- DEMOCRATIC LEGAL AND POLICY /Democratic Legal & Policy.

CREATE SAME SPELLING ACROSS \*/

**DATA** WORK.CONTRACTS1;

SET WORK.CONTRACTS1;

IF Service\_responsible IN ('Democratic Legal & Policy')

THEN Service\_responsible= 'DEMOCRATIC LEGAL AND POLICY';

IF Service\_responsible IN ('Human Resources ICT/CSC & Shared Support Services')

THEN Service\_responsible= 'HUMAN RESOURCES ICT & SHARED SUPPORT';

IF Service\_responsible IN ('Planning & Sustainability')

THEN Service\_responsible= 'PLANNING';

**RUN**;

**DATA** WORK.CONTRACTS1;

SET WORK.CONTRACTS1;

Service\_responsible= UPCASE(Service\_responsible); **RUN**;

**PROC** **FREQ** DATA=WORK.CONTRACTS1;

TABLES Service\_responsible; **RUN**;

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **The FREQ Procedure** | | | | |
| **Service responsible** | | | | |
| **Service\_responsible** | **Frequency** | **Percent** | **Cumulative** | **Cumulative** |
| **Frequency** | **Percent** |
| **CHIEF EXECUTIVE** | 1 | 0.52 | 1 | 0.52 |
| **COMMUNITY** | 24 | 12.37 | 25 | 12.89 |
| **DEMOCRATIC LEGAL AND POLICY** | 14 | 7.22 | 39 | 20.1 |
| **ENVIRONMENT** | 22 | 11.34 | 61 | 31.44 |
| **FINANCE & COMMERCIAL** | 54 | 27.84 | 115 | 59.28 |
| **HOUSING** | 5 | 2.58 | 120 | 61.86 |
| **HUMAN RESOURCES ICT & SHARED SUPPORT** | 44 | 22.68 | 164 | 84.54 |
| **PLANNING** | 19 | 9.79 | 183 | 94.33 |
| **PROPERTY** | 11 | 5.67 | 194 | 100 |

/\* WE SEE IN THE PROC MEANS OUTPUT - GENERATED ABOVE- THAT THE END

DATE OF THE PROJECT IS AVAILABLE ONLY FOR 83 CASES. SINCE BOTH THE

Y VARIABLES NEED END DATE FOR CALCULATIONS, WE SHOULD REMOVE ALL THE

OBSERVATIONS WHERE END DAT IS MISSING \*/

**DATA** WORK.CONTRACTS1;

SET WORK.CONTRACTS1;

WHERE END\_DATE1 NE **.**; **RUN**;

**PROC** **MEANS** DATA=WORK.CONTRACTS1; **RUN**;

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | |  |  | | --- | --- | | |  | | --- | | **The MEANS Procedure** | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | | **Variable** | **Label** | **N** | **Mean** | **Std Dev** | **Minimum** | **Maximum** | | --- | --- | --- | --- | --- | --- | --- | | |  | | --- | | Contract\_Start\_Date | | Contract\_Value | | END\_DATE1 | | TAT | | PERDAY | | |  | | --- | | Contract Start Date | | Contract Value | |  | |  | |  | | |  | | --- | | 83 | | 83 | | 83 | | 83 | | 83 | | |  | | --- | | 19645.99 | | 1499100.41 | | 20936.33 | | 1290.34 | | 1167.87 | | |  | | --- | | 531.4227680 | | 6817042.50 | | 671.6119144 | | 716.2426908 | | 6794.94 | | |  | | --- | | 17623.00 | | 0 | | 20331.00 | | 363.0000000 | | 0 | | |  | | --- | | 20362.00 | | 49000000.00 | | 25383.00 | | 5478.00 | | 59157.89 | | | | | |

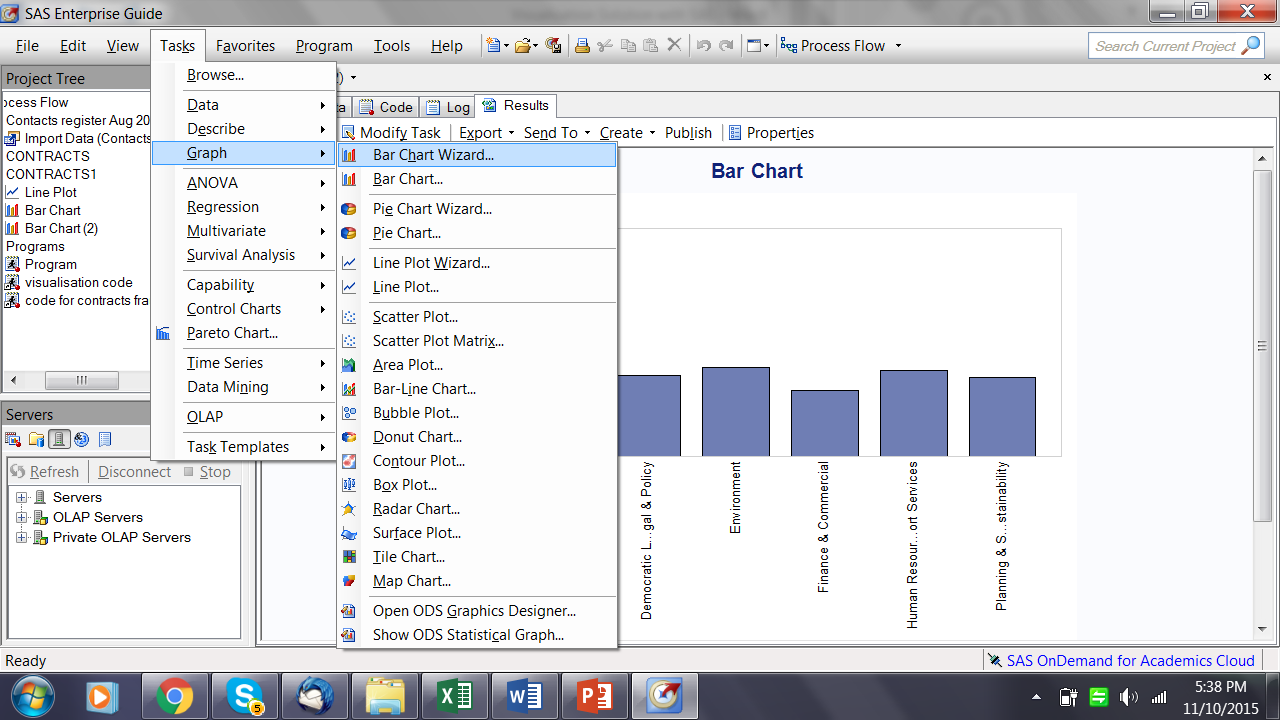
6. Visualisation :-

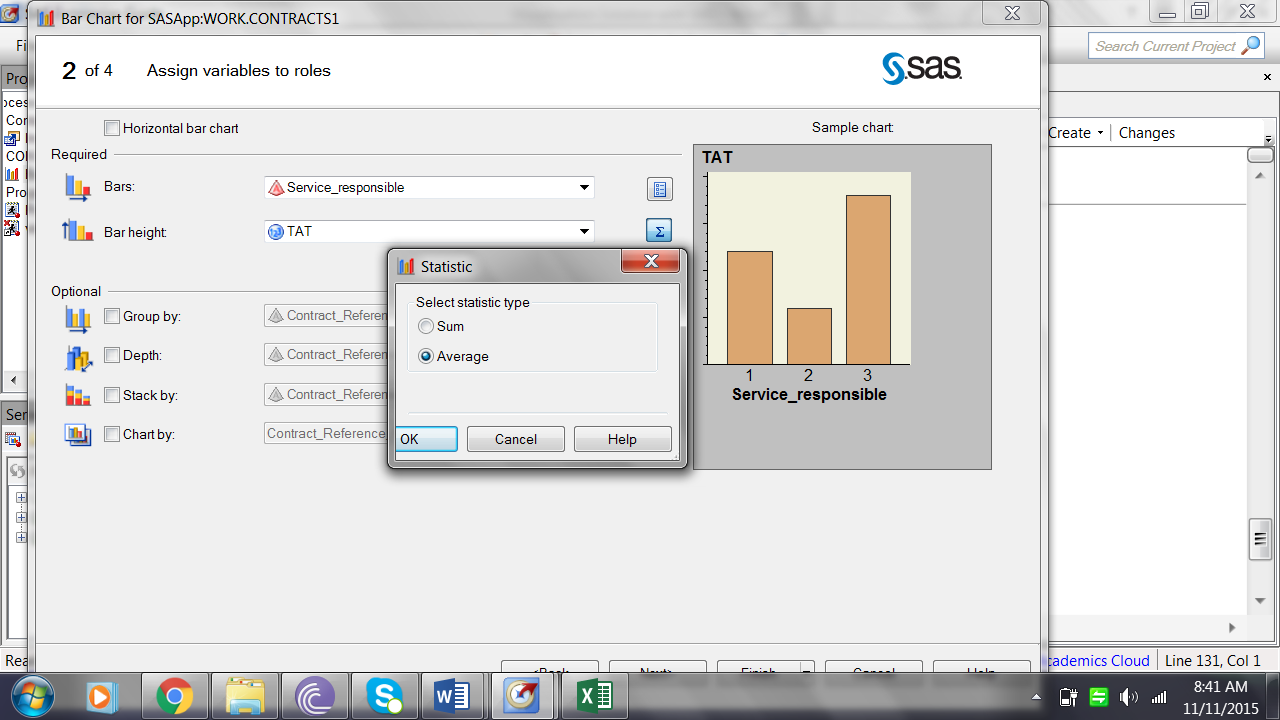
a. Bar Chart :- To understand Average value of TAT across different segments of Service\_responsible

Method 1 :- Use the buttons on top of the screen

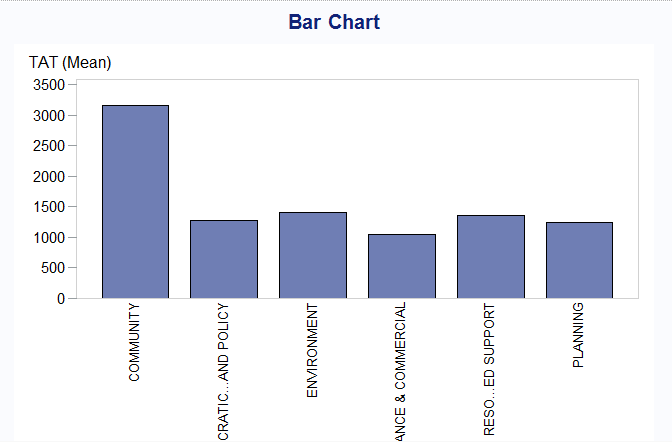
Tasks > Graph> Bar Chart Wizard

This will open the wizard . Just use the drag and drop functionality to create the graph





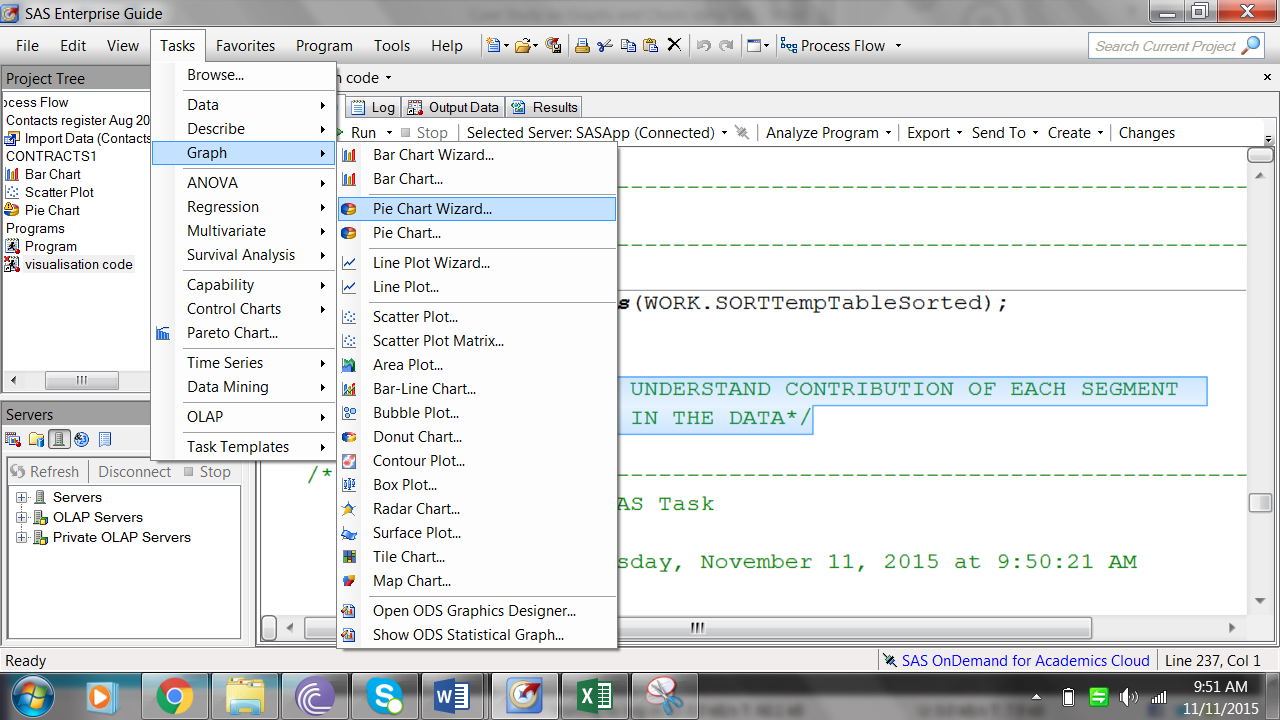
The Bar chart appears

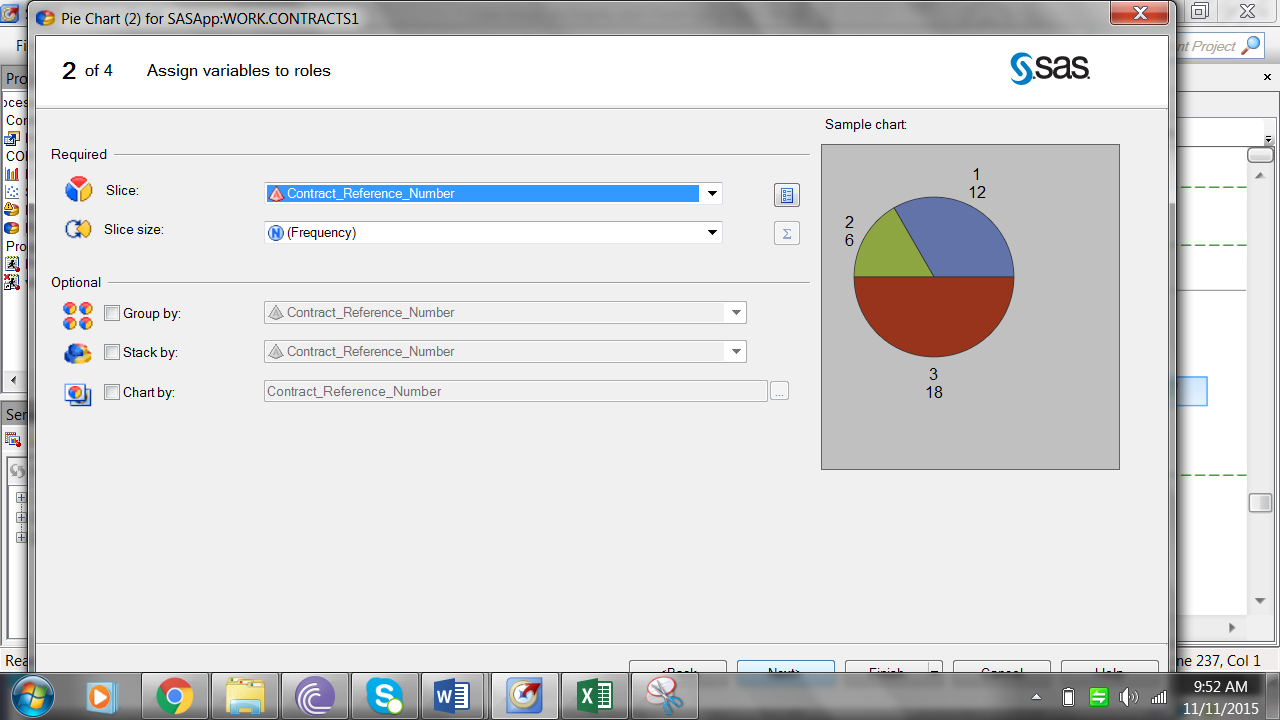


/\* CREATE PIE CHART TO UNDERSTAND CONTRIBUTION OF EACH SEGMENT

OF Service\_responsible IN THE DATA\*/

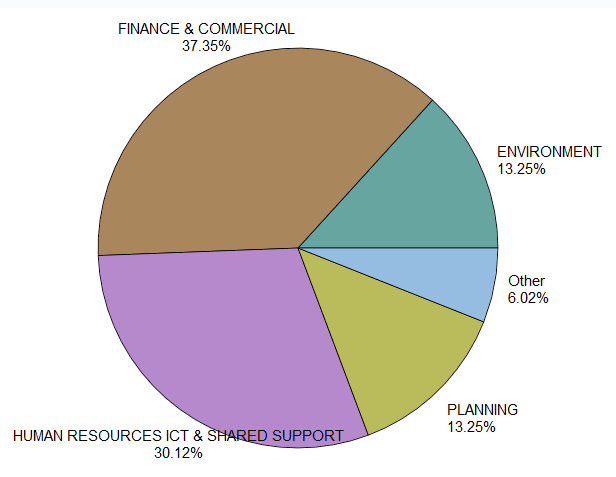
Using buttons Tasks > Graph > Pie Chart Wizard





The code gets auto generated in the code window.

The output looks like the following



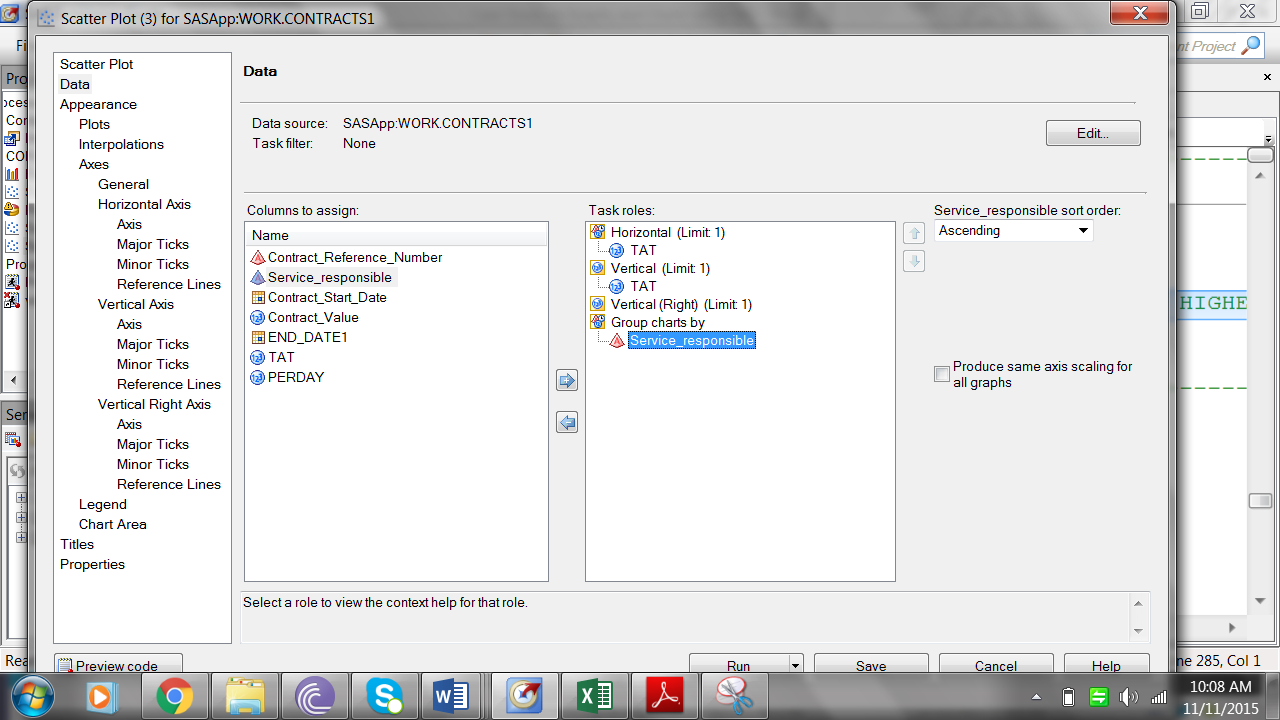
/\* DRAW A SCATTE PLOT TO SEE IF SOME CASES HAVE SIGNIFICANTLY HIGHER TAT

OR HIGHER RATE PER DAY VALUES\*/

Note – I am doing it for TAT . You should do it for Per Day Rate as your exercise

Use the buttons and plot Tasks > Grpahs > Scatter plot

Choose the same variable for a 2 D scatter plot



Export the result as pdf

