



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

Applicability of SAS to Perform Data Analysis and Its Advantages

Dr. Reena Hooda

Assistant Professor, Department of CSE, Indira Gandhi University Meerpur (Rewari), Haryana, India

ABSTRACT: The topical roundreis mustering of cutting-edgetools where everyone is trying innovatively in achieving different with fewer efforts, time and price, embroiled in examining the web to obtain hidden magical skills for vital stand in competitive environment. Where data is large and complex and outspread the frontiers of data warehouse towards big data, it is requisite to find a tool that is efficient cost effective easy to learn and reliable to proceed. After surfing the Google, a list of software and tools are available provided as free of cost or as an open source tools for instance, Orange, R, Rapid Miner and SAS Studio. The present paper highlighted the applicability of SAS as free tool available in many shots like University Edition or SAS Studio that can be run in cloud environment or on local machine. The paper further elaborates the user friendly environment of SAS for data analysis and functionality with non SAS data formats like Excel. Paper also exemplify the data management, access, coding and presentation options as well as advantage of SAS in writing codes with basic keywords and help boxes, inbuilt procedures of SAS, predefined range of tasks & utilities, libraries for data transforming, loading in data analysis and report generation.

KEYWORDS: Code, Task, CSV, Studio, Browser.

I. INTRODUCTION

SAS is software that was released around 40 years ago in 1976 by SAS institute. [1] It can be used as data mining software to mine and extract from the data stored elsewhere. SAS is user friendly software through which a non-technical user can learn and perform statistical data analysis. It increase overall learning and performance through reduced level of coding and fast learning with wider range of formatting and data cleansing options. It is 4th generation programming that is highly efficient, flexible and extensible. [1] SAS has a complete product documentation help in learning, validation and reengineering process. SAS studio is bouquet of readymade programming procedures for data handling like manipulation, storage, retrieval, analysis & reporting and provides reliable results. [1] SAS has a macro facility for common task facilitating modularized work for easy reuse and reduced programming as a feature of Object Oriented Programming and a non-procedural language that supports SQL. [1] It able to read data from any source whether it is a binary file, Excel file, variable length, missing data. [1]

SAS studio is growing, pure HTML5 web application that can be accessed anywhere with Internet connection and user ID on the laptop or desktop computers or iPad as well. [11] Other than importing files, accessing data and libraries, there is a list of predefined tasks and utilities to generate SAS code, results & output presented as per the choice of the user including graphic representations and analytical procedures. [5] [10] A user friendly environment helps in concentrating on the analysis part rather than spending hours on coding or cleansing. The work on SAS studio can be done in a cloud environment, where all of the uploaded data, coding and results will be saved and communicated to and from accommodated server in cloud through the internet; or a copy of SAS can be installed on local machine, the results can be seen in any major browsers may be its a Google Chrome 27+, Microsoft Internet Explorer 9, Microsoft Internet Explorer 10, and Microsoft Internet Explorer 11, Apple Safari 6+, Mozilla Firefox 21+. [5] [10] [11] It is necessary to mention here that the Studio 3.5 has been verified against WCAG 2.0 (Web Content Accessibility Guidelines) a part of WAI (Web Accessibility Initiative) of the Worldwide Web Consortium and against the accessibility standards for electronic information technology [9]. SAS Studio consist of two kinds of perspectives found on top right-hand side of the toolbar, one is SAS Programmer perspective and other one is Visual Programmer perspective has also Process Flows [12], a set of functionalities to meet the requirements of various people and groups within a society like business, institute, government, NGOs and other agencies. [10] To work in a cost effective environment, faster analysis, customized tasks, various reliable Open Source Tools and free of cost software with a range of extensibility, flexibility,



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

easy to use are offered on net. Everyone is eager to find such tools to make sense to their purpose intellectual level plus financial capabilities, for instance Rapid Miner, SAS Studio, Hadoop etc. are some of the software to manage and mine the big data through collection from many sources accessible in different formats, transformation, loading and analysis, forecasting as well. [6] To predict the hidden facts, knowledge extraction and decision making data warehousing approached has been developed, SAS studio provide a friendly interface to user to analyze the data from different perspectives and mining it to get the target results.

II. LITERATURE SURVEY

SAS (Statistical Analysis System) is a growing application with limited data survey. Most of the references contributed in the documentation and example data sets with worthy efforts to work with SAS Studio for statistical data analysis. In [1] various features of SAS have been provided. [5], [11], [12] delivered the introduction about SAS, working of SAS in data analysis including data representations in different formats, while [6] highlighted the applications of SAS and its usability. [8] [9] [10] [12] [13] offered the complete documentation of the SAS studio that helps the user in data analysis. In [14] provided the example data base in excel on which various operation can be performed while learning and performing data analysis.

III. DATA EXTRACTION IN SAS AND RESULTS

As mentioned earlier, data inputs can be in any format. To analyze the data in SAS Studio, the file must be saved in CSV (comma separated value) format. If one is using the Excel file then it must be saved in .csv extension in save as dialog box and after that comma separated values can be seen by opening the file in notepad. To import the file in SAS Studio, the file must be uploaded to the server before performing any task or coding, for this click the upload button in navigation pane and give user name to newly uploaded file. By default the file is saved as work. After signing in the SAS Studio, the Studio window will be opened divided in two main parts: Navigation Pane and Work Area. Navigation pane has five sub panes: Server Files and Folders, File Shortcuts, Tasks, Snippets and Libraries and can be minimized. Work area is program area that always opened and cannot hide like navigation pane; it has main parts such as code, log, results and output after running the program [12]. Fig.1 shows the Studio main window after signing in using user id and password.

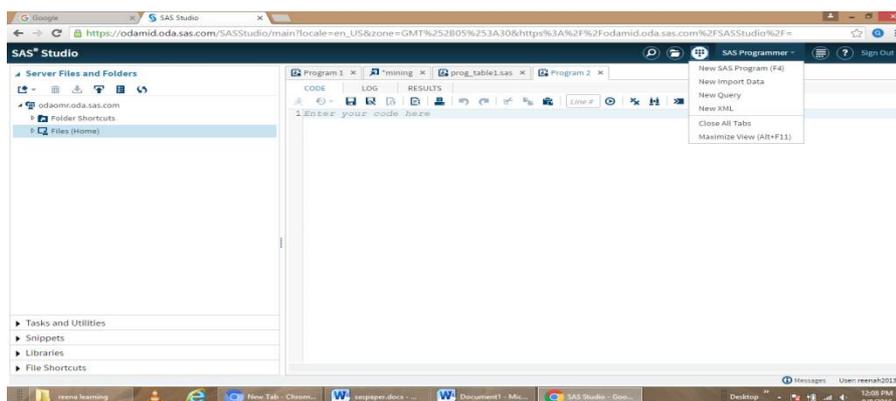


Fig.1. Shows the Main Window of SAS

To work with SAS Studio, first go to folder shortcuts of navigation pane and then choose the file. If there is no file or wanted to add a new file then click the upload button on navigation pane or it can be selected from the drop down menu of SAS programmer on right-hand side of toolbar, click on import and give the file path. In present paper Excel file titled mining has been used for the analysis in SAS [14]. Now right click the file just uploaded and shown under home file in navigation pane and check the path in the properties option that displays the changed location of the file to specify that the file has been uploaded. Put on this file source in the file address of 'infile' keyword. Write the code and title the SAS table for example, mining that will be shown as work.mining in output data window. The order of the

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

columns in input keyword of the coding part must be same as the source table, and run the file. The SAS table can be seen in output data menu of the program on the top. Table work.mining has been displayed that can be filtered or values can be sorted or can the table properties to show different information like General, Columns, Extended attributes etc. as shown in Fig.2.

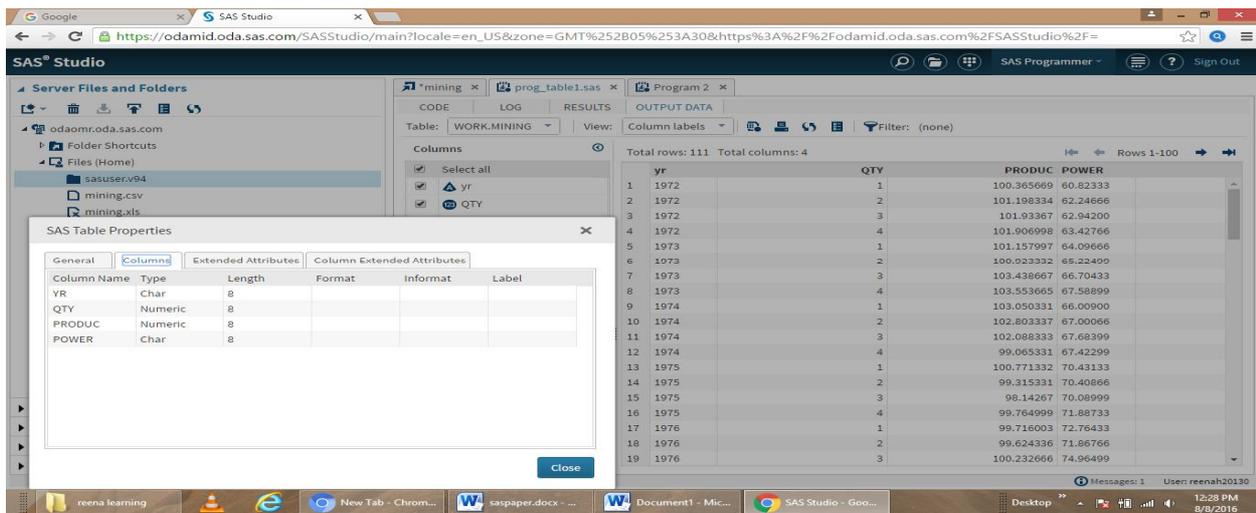


Fig. 2. Shows the Output Data Table having 111 records and 4 fields.

The number of columns in SAS tables can be increased or decreased based on the view required. It is must to mention that if during the task the network connection will go off or disconnected all the coding has been lost so save the coding for future use and run it or copy it to avoid rewriting. Table is a run time entity, every time a code is to be run, it will be displayed in work area. Task and Utilities contains a list of the operations that can be performed on the data like List Table Attributes, Characterize Data, List data, Filter Data, Sorting Data, partitioning and so on that is required to select the data source with customized options, for instance apply the filter by inserting the values in required fields and click on 'Output Data Set' to display the results, if only subset of the data output has been required then also enter the number of observations to be displayed on the pane. The advantage is that all the required code is automatically generated for the specified task as shown in Fig.3, and output data set is shown in table 1 having 8 observations.

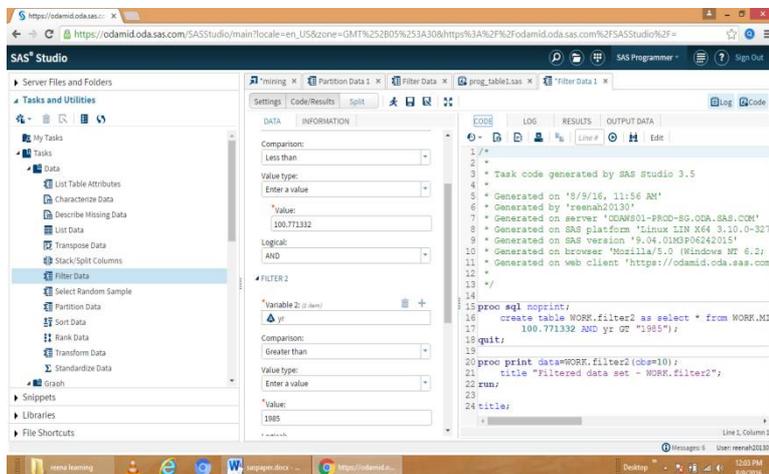


Fig. 3. Shows the filter task and auto generated code.

Tab. 1. Filtered data set - work.filter2

Obs	Yr	QTY	PRODU	POWER
1	1986	3	99.502	86.237
2	1986	4	99.994	84.31366
3	1987	1	100.516	85.28766
4	1992	1	99.733	100.6449
5	1992	2	100.144	99.40399
6	1992	3	99.897	100.7186
7	1992	4	100.226	99.23266
8	1993	1	99.326	98.22166
9	1993	2	99.361	98.01333
10	1993	3	99.503	94.41866



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

The Libraries section of the navigation pane contains all the contents of mining table created. The table can be expanded columns can be viewed. Besides viewing the column names, they can be rearranged; a green ok sign shows the valid arrangement. [12]

IV. ADVANTAGE OF SAS STUDIO AND APPLICATIONS

As it's a web based application that is just requires a browser and no need to install anything on the system as it requires just login and go. It is so flexible, interactive even simple and easy to comprehend-programming instructions, keywords, syntax that suits every data analyzer. New version SAS 9.4 [13] meets the increasing demand for capacity to handle data variety, quantity, and integration from different sources, data sharing and more is the reusability of code. The snippets provide the templates for user code that are meant for most common tasks, for example different types of graphs. SAS's new version includes an embedded middle tier server known as web application server eliminating the third party support say JBoss, WebSphere etc. It support in management of relational data through SAS FedSQL in a high performance mode i.e. working in SAS studio is not time consuming as it quickly produce the results and output data set, based on user's preferences. Applicability of SAS analytics ranges from a beginner to complex analysis in various fields [5] like pharmaceuticals, engineering, education, manufacturing to develop the concrete results and anticipating future. [4] SAS business solutions empowered with data management, analytics, Business Intelligence using flexible, extensible programming language and benefits in data integration from various formats, validation, cleaning and compiled data, moreover presentation in form of reporting, summary and also graphical presentations. [5]

V. CONCLUSION

SAS Studio is like a gift namely free of cost for the user who is fascinating toward searching of new insights in data analysis rapidly yielding in interactive graphical presentations. SAS has a wide-ranging documentation and learning facility furthermore well expressive tools having embedded SQL support that enhance the programming to produce precise results. It is a drag & drop technology that automatically create the codes, offer predefined tasks & utilities, templates and flexibility in handling data tables comprising graphical arrangements swiftly without the requirement of additional provisions. If a network connection is functioning well, it can be started on any major browsers and windows. The further scope of the paper is to deploy data mining methods in SAS Studio to identify the hidden patterns and forecasting the target.

REFERENCES

1. [https://en.wikipedia.org/wiki/SAS_\(software\)](https://en.wikipedia.org/wiki/SAS_(software))
2. https://www.sas.com/content/dam/SAS/en_us/doc/factsheet/sas-university-edition-107140.pdf
3. <http://support.sas.com/training/tutorial/studio/create-table-csv-file.html>
4. <http://www.sas.com/apps/webnet/video-sharing.html?bcid=2415818414001>
5. <http://support.sas.com/software/products/sasstudio/>
6. <http://www2.sas.com/proceedings/sugi31/099-31.pdf>
7. http://www.sas.com/en_us/software/university-edition.html
8. <http://support.sas.com/documentation/onlinedoc/sasstudio/>
9. <http://support.sas.com/documentation/prod-p/webeditor/3.5/en/HTML/webeditor11y.htm>
10. <http://support.sas.com/documentation/cdl/en/webeditorug/68828/HTML/default/viewer.htm#n1lynjknzrukln1cxencems3iec.htm>
11. Michael A. Monaco, Marie Dexter, Jennifer Tamburro, 'Introduction to SAS® Studio'. Paper SAS302-2014. Retrieved from: <http://support.sas.com/resources/papers/proceedings14/SAS302-2014.pdf>
12. <http://support.sas.com/documentation/cdl/en/webeditorug/68828/HTML/default/viewer.htm#p092daf4a5ypjcn138maiooc8wsn.htm#n04giewxrgzjr2n1ng6mue98up6e>
13. <http://support.sas.com/documentation/cdl/en/whatsnew/64788/PDF/default/whatsnew.pdf>
14. <http://www.principlesofeconometrics.com/excel.html>

BIOGRAPHY

Reena Hooda is Assistant Professor in the Department of Computer Science & Engineering, Indira Gandhi University Meerpur (Rewari) Haryana, India. She Received her Master of Computer Application (MCA) degree in 2005, MBA in 2007 and Ph.D in 2012 from MDU Rohtak, Haryana (India). Her research interests are DBMS, Computer Networks and Data Mining Techniques (Learning and Applicability).