Data Analytics with SAS

Explore your data and get actionable insights with the power of SAS

Nishant Sidana

www.bpbonline.com

Copyright © 2024 BPB Online

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without the prior written permission of the publisher, except in the case of brief quotations embedded in critical articles or reviews.

Every effort has been made in the preparation of this book to ensure the accuracy of the information presented. However, the information contained in this book is sold without warranty, either express or implied. Neither the author, nor BPB Online or its dealers and distributors, will be held liable for any damages caused or alleged to have been caused directly or indirectly by this book.

BPB Online has endeavored to provide trademark information about all of the companies and products mentioned in this book by the appropriate use of capitals. However, BPB Online cannot guarantee the accuracy of this information.

First published: 2024

Published by BPB Online
WeWork
119 Marylebone Road
London NW1 5PU

UK | UAE | INDIA | SINGAPORE

ISBN 978-93-55515-971

www.bpbonline.com

Dedicated to

My wife, ***Nisha****, my daughter,* ***Anika***
and
my parents, ***Suresh Sidana*** *and* ***Reeta Sidana***

About the Author

Nishant Sidana is an analytical professional. He has been working in the industry for around a decade and delivering analytical solutions to industries from BFS, Sales, Retail, Energy, Oil and Gas, Telecom, etc. Earlier he has worked in Oil and Gas industry as design engineer.

He is currently working as a Senior Manager (Data Science) with Cognizant Technology Solutions. He is a Graduate Mechanical Engineer from National Institute of Technology, Durgapur. He has also done an executive program, PGXPM in General Management from Great Lakes Institute of Management, Chennai. He has also trained professionals and students for data analytical roles.

About the Reviewer

Arimitra Maiti is a seasoned analytics professional with emphasis on leveraging SAS native solutions and tools, as well as open-source platforms to drive data science initiatives. He possesses a keen interest and responsibility toward outcome-driven problem-solving through diverse analytical techniques. Notably, he is currently engaged in his inaugural technical book review, complementing his recurrent contributions to the wider community through insightful blogging and active participation in open forums.

Acknowledgement

I want to express my deepest gratitude to my family and friends for their unwavering support and encouragement throughout the journey. I want to especially thank, my wife Nisha who supported me with enthusiasm throughout this journey.

I am also grateful to BPB Publications for lending their guidance, support and expertise, successfully throughout the journey. I received unconditional support from the reviewers, technical experts, and editors in revising and finalizing the book.

Finally, I would also like to thank the industry as working in different domains helped me bring minute details of analytical concepts into the book.

Preface

Data Analytics has become the most sort after technology today for every industry. Competitive advantage is not possible without examining the data of business operations. This book will help readers build concepts to discover patterns from the data which can help them understand the business operation and acquire a competitive edge.

SAS is an analytical tool. It has got all the functionalities which can help us analyse data easily. This book has tried to help readers with every aspect and function of data analytics with SAS.

Readers will learn about SAS and its components, concepts and functions for data manipulation, data exploration, data visulaization, etc. The book has presented examples for every topic which will help readers to be job ready.

This book is helpful for readers from any domain such as engineering, management, finance, medical etc., who want to build career in Data Analytics, Data Science, Business Intelligence roles or want to explore the power of data.

I hope that readers will find this book helpful, informative and friendly in their analytical journey:

Chapter 1: Introduction to SAS Programming - The chapter will introduce SAS Software which is an analytical software. It is used for analysing data, visualising data. The chapter will focus on the building blocks of SAS. It will help participants understand SAS. SAS Libraries are an important component while working with SAS. It is used to store data files. The chapter will focus on understanding Data and Set statements which help create new data sets after carrying out modifications on existing data sets.

Chapter 2: Overview of SAS Components - The chapter will focus on SAS components. The components are the building blocks of SAS. We can write programs in SAS with the help of these components. The chapter will also focus on importing data sets to SAS.

Chapter 3: Data Manipulation - Raw data available cannot be used directly. It needs to be clean, and noise needs to be removed before using it for further analysis. Data Manipulation is a critical step for every analyst. It basically helps shape the data. Complete data is not important to us. It helps remove unwanted data before processing it.

Chapter 4: Advanced Data Manipulation - The chapter will focus on advanced techniques for data manipulation.

Chapter 5: SAS Functions and Options - The chapter will focus on SAS Functions which is a SAS component that accepts an argument, perform an operation on data and return a value. Functions help us perform many operations such as arithmetic, text, date and time. SAS options also control SAS operations.

Chapter 6: Data Exploration-I - After cleaning the data, the first step in data analysis is Exploratory Data Analysis (EDA). It is a critical step in data analysis. It helps us understand the data. It helps us know the volume of data we are working on, data types, etc. It helps get insights from data about the business operations held in past.

Chapter 7: Data Exploration-II - The chapter will focus on advance concepts of data exploration.

Chapter 8: Importing Raw Data Files - A raw data file is an external text file whose records contain data values that are organized in fields. They can be read by a variety of software programs. This chapter focuses on concepts for importing raw data files to SAS.

Chapter 9: Advanced SAS: Proc SQL - SQL is a standardized language that is widely used to retrieve and update data in tables and in views that are based on those tables. It can be used with any SAS data set. It enables you to combine data from two or more different types of data sources and present them as a single table. It has all the functions for data analysis.

Chapter 10: Macro Programming for Faster Data Manipulation - Macros allows us to avoid repetitive sections of code and to use them again and again when needed. Used for reducing the amount of text that must be entered to do common tasks. It enables to assign of a name to character strings or groups of SAS programming statements.

Chapter 11: Data Visualization - Data Visualization is a critical element in analytics to get insights from data. Graphs and SGPLOT will help us create data visualization with Bar Charts, Pie-Charts, Histograms, Box plots etc.

Code Bundle and Coloured Images

Please follow the link to download the
Code Bundle and the ***Coloured Images*** of the book:

https://rebrand.ly/y7vclmu

The code bundle for the book is also hosted on GitHub at
https://github.com/bpbpublications/Data-Analytics-with-SAS.
In case there's an update to the code, it will be updated on the existing GitHub repository.

We have code bundles from our rich catalogue of books and videos available at **https://github.com/bpbpublications**. Check them out!

Errata

We take immense pride in our work at BPB Publications and follow best practices to ensure the accuracy of our content to provide with an indulging reading experience to our subscribers. Our readers are our mirrors, and we use their inputs to reflect and improve upon human errors, if any, that may have occurred during the publishing processes involved. To let us maintain the quality and help us reach out to any readers who might be having difficulties due to any unforeseen errors, please write to us at :

errata@bpbonline.com

Your support, suggestions and feedbacks are highly appreciated by the BPB Publications' Family.

Did you know that BPB offers eBook versions of every book published, with PDF and ePub files available? You can upgrade to the eBook version at www.bpbonline.com and as a print book customer, you are entitled to a discount on the eBook copy. Get in touch with us at :

business@bpbonline.com for more details.

At **www.bpbonline.com**, you can also read a collection of free technical articles, sign up for a range of free newsletters, and receive exclusive discounts and offers on BPB books and eBooks.

Piracy

If you come across any illegal copies of our works in any form on the internet, we would be grateful if you would provide us with the location address or website name. Please contact us at **business@bpbonline.com** with a link to the material.

If you are interested in becoming an author

If there is a topic that you have expertise in, and you are interested in either writing or contributing to a book, please visit **www.bpbonline.com**. We have worked with thousands of developers and tech professionals, just like you, to help them share their insights with the global tech community. You can make a general application, apply for a specific hot topic that we are recruiting an author for, or submit your own idea.

Reviews

Please leave a review. Once you have read and used this book, why not leave a review on the site that you purchased it from? Potential readers can then see and use your unbiased opinion to make purchase decisions. We at BPB can understand what you think about our products, and our authors can see your feedback on their book. Thank you!

For more information about BPB, please visit **www.bpbonline.com**.

Join our book's Discord space

Join the book's Discord Workspace for Latest updates, Offers, Tech happenings around the world, New Release and Sessions with the Authors:

https://discord.bpbonline.com

Table of Contents

Chapter 1
Introduction to SAS Programming

Introduction

This chapter will introduce **Statistical Analysis System (SAS)**, an analytical software, used for analyzing and visualizing data. It will focus on the building blocks of SAS and help the readers understand SAS libraries which are an important component while working with SAS. It is used to store data files. This chapter will focus on understanding Data and Set statements which help create new data sets after carrying out modifications on existing data sets.

Structure

In this chapter we will be discussing the following topics:

- SAS
- GUI
- SAS libraries
- Data and set statements
- Executing data steps
- Variable attributes

Objectives

By the end of this chapter, you will be able to understand how SAS executes data steps. You would be able to understand the concept of SAS libraries as well as variable attributes. You will also learn the purpose of different windows of SAS will be understood.

Statistical Analysis System (SAS)

Competition in every sector is increasing rapidly and data analytics is the science which can help companies beat the competition. Companies today wants to utilize the power of analytics on data to explore new ways for increasing profits, expanding product base, dominating the competition, and so on. To perform data analysis, we require analytical tools. SAS is a tool used for data analytic activities.

SAS is a programming language used for statistical analysis. It is developed by SAS institute. It is used for:

- Data mining
- Data management
- Business intelligence
- Exploratory Data Analysis (EDA)
- Data visualization
- Data analysis
- Predictive analytics
- Advanced analytics

Graphical User Interface (GUI)

Graphical User Interface (**GUI**) helps in data preparation, integration, visualization, reporting, and analytics. SAS has 4 important windows which help in easy access to data activities:

- **Code/Editor window**: It is used for writing and editing codes in SAS, as per *Figure 1.1*:

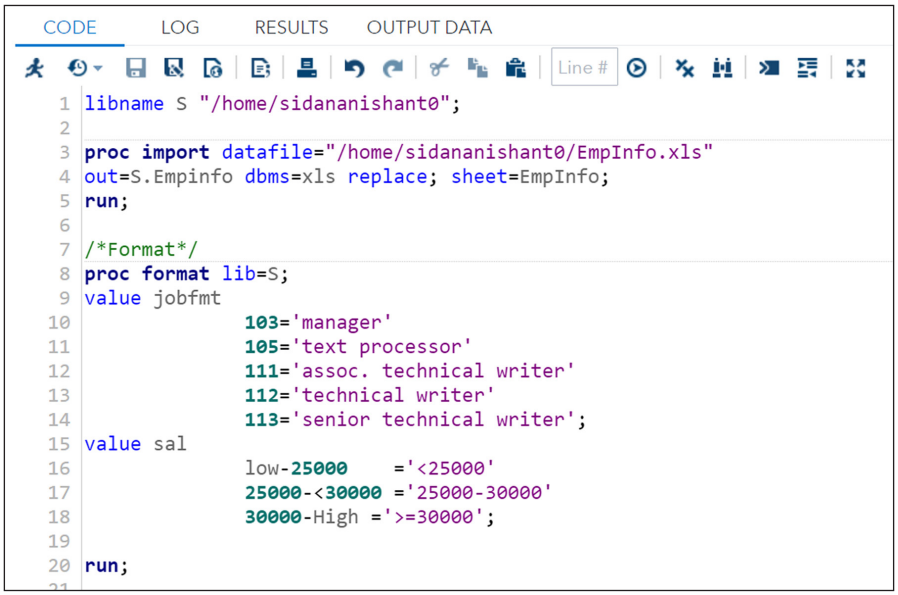


***Figure 1.1**: Code/Editor window*

- **Log window**: It provides information about the execution of a program, and also lists the errors, warnings, or notes that come after executing the code, as per *Figure 1.2*:

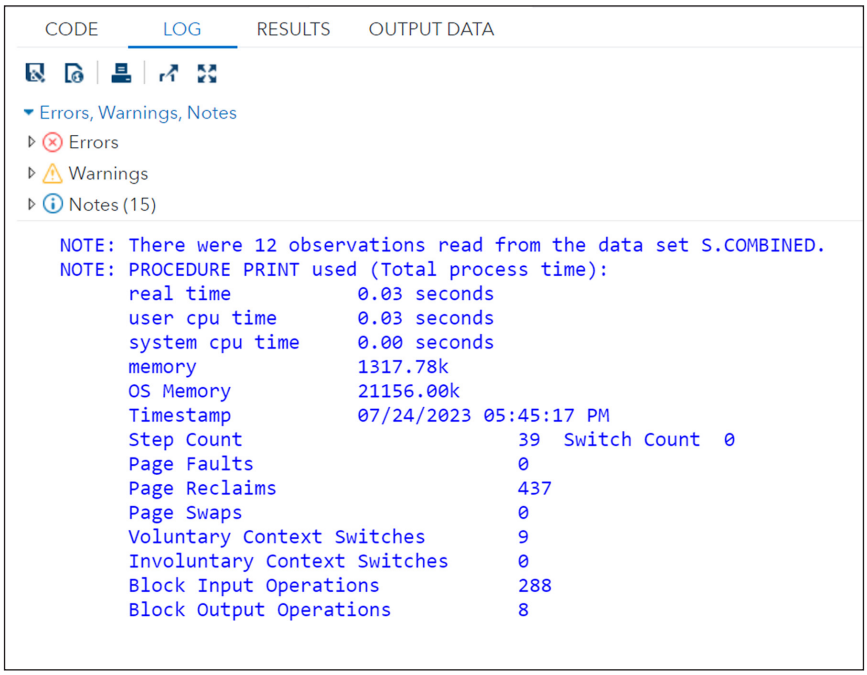
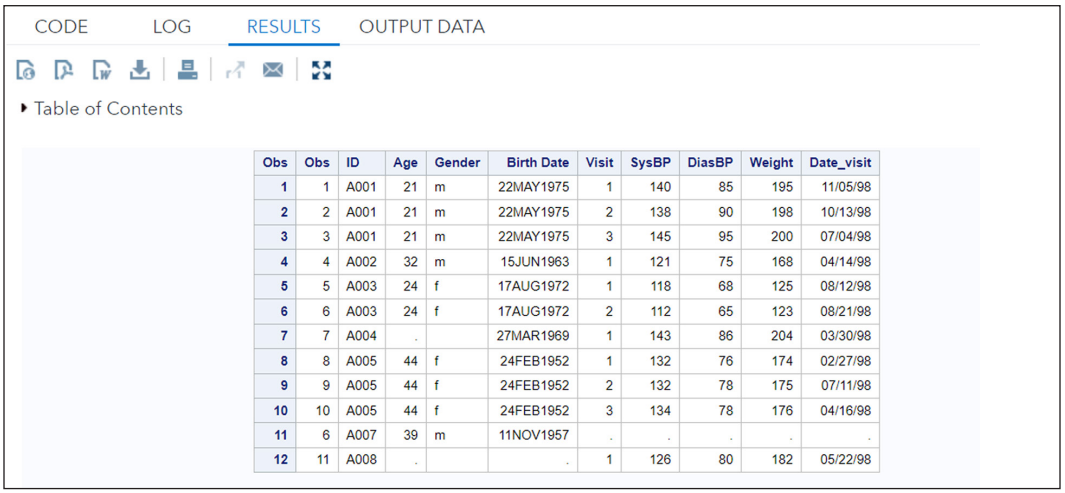


***Figure 1.2:** Log Window*

- **Results window**: It is used for displaying the output after executing the code, as per *Figure 1.3*:



Obs	Obs	ID	Age	Gender	Birth Date	Visit	SysBP	DiasBP	Weight	Date_visit
1	1	A001	21	m	22MAY1975	1	140	85	195	11/05/98
2	2	A001	21	m	22MAY1975	2	138	90	198	10/13/98
3	3	A001	21	m	22MAY1975	3	145	95	200	07/04/98
4	4	A002	32	m	15JUN1963	1	121	75	168	04/14/98
5	5	A003	24	f	17AUG1972	1	118	68	125	08/12/98
6	6	A003	24	f	17AUG1972	2	112	65	123	08/21/98
7	7	A004	.		27MAR1969	1	143	86	204	03/30/98
8	8	A005	44	f	24FEB1952	1	132	76	174	02/27/98
9	9	A005	44	f	24FEB1952	2	132	78	175	07/11/98
10	10	A005	44	f	24FEB1952	3	134	78	176	04/16/98
11	6	A007	39	m	11NOV1957	.	.	.	.	.
12	11	A008	.		.	1	126	80	182	05/22/98

Figure 1.3: *Results window*

- **Server files and folders**: It is used to represent the physical storage locations of data, as per *Figure 1.4:*

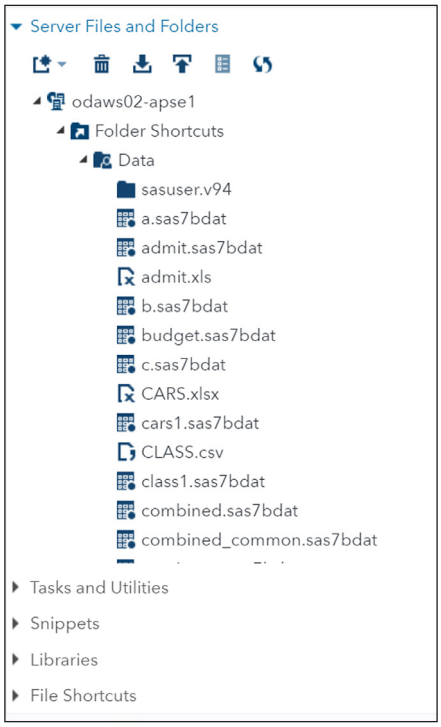


Figure 1.4*: Server files and folders*

SAS libraries

Output of the data pre-processing work done on data with SAS tool is stored in SAS in the form of SAS files. SAS can store data files in the form of a library.

A SAS library is defined as a collection of SAS files and SAS data sets, recognized by SAS.

The files can be stored as a unit. SAS libraries do not physically store data files. They just reference the location on computer where the files are physically stored.

Types of libraries in SAS are:

- Temporary library
- Permanent library

Temporary library

The output of data pre-processing work create many subsets of original dataset before getting the final dataset after the completion of data pre-processing. There is no requirement for storing these subsets permanently as we require them only during the working session till we get final dataset after pre-processing.

A temporary SAS library is used as a temporary storage location for data files. Important points:

- It stores data files only for the current SAS session and the data files are deleted once we end the SAS session.
- The name of the temporary library in SAS is work.

There is no specific library name that we can provide while storing a data file in a temporary library.

Permanent library

Once data pre-processing is completed, the final dataset needs to be stored permanently as we need it further for carrying out analysis to finally arrive at a solution for the problem.

A permanent SAS library is used as a permanent storage location of data files as the following:

- The data files stored here are available in every SAS session.
- We can specify a library name while creating SAS permanent library.
- The **`libname`** statement is used to create a permanent library.
- Permanent library creates a reference to the path where SAS files are physically stored.
- Syntax for creating a permanent library,