

Data Science with SAS

What are we offering!

Curriculum:

✓ **Phase 1: SAS programming**

Learn how to do data analysis with SAS language- Data cleaning, Data Exploration, reporting and Automation

✓ **Phase 2: Statistics-**

No previous Background in stats? Don't worry! We are here to make maths simple for you.

✓ **Phase 3: Predictive modelling/Machine learning-**

- What is the probability that your best friend is going to leave the company in next 1 year?
- Help banks to understand who can be a likely defaulter!
- Honda is going to launch a new car? What should be the optimal price they should set for that?

What else?

It's a complete online/offline analytical course not only making you trained on data science using SAS but making you equipped with job ready skills on day of your new analytics job.

We are offering:

- ✓ 6 Assignments
- ✓ 3 real time industry projects on healthcare, banking and retail
- ✓ Resume that will help you merge your previous experience with data analytics
- ✓ Placement preparation- Question bank, PDF's, E-books, Sharing your CV with our Industry links
- ✓ Certification at end of the course

Starting packages with SAS skills- 4 lacs

Starting packages with SAS+ Stats +Predictive modeling Skills- As much as you would love to get – **Still 6 lacs to 8 lacs is the industry average**

Duration- 3 month's weekend course (3 hours a week) - Fees- 25K (\$500- International students)

Don't want to do predictive modelling? Try (SAS programming + SQL) - Fees-20K

Session 1- Introduction to Analytics

- What is Data Analytics & Data Science
- Different types of Data Analytics (Descriptive, Predictive, Prescriptive)
- What is Artificial Intelligence
- Machine Learning (Supervised & Unsupervised Learning)
- Deep Learning (Artificial Neural Networks, CNN)
- Overview of Banking, Healthcare, Telecom domain
- Real world Applications of Machine Learning & Deep Learning
- What to expect from this course (Salary, Market trends, job roles, Domain)

Phase 1: SAS programming

Base SAS

Session 1-2- Reading data in SAS

- Introduction to SAS interface and library structure and definition
- Reading data using datalines and importing and exporting datasets
- Types of datafiles to be read in SAS
- Infile statement - reading raw data
- Formats and Informat
- Variable attributes and data modification using Data and Set statements
- Infile vs proc import
- Missover, trunccover and pad

Assignment 1 for practicing reading data in SAS

Session 3-4 Data Preparation

- Using conditional statements to modify data - Where, If and Nested IF
- SAS Functions for data manipulation
- Numeric functions in SAS- Max(), Min(), Avg(), standard dev, variance
- Character function in SAS- RAND(), LENGTH(),TRIM(), COMPRESS(), and more
- Date functions in SAS- Intck(), Intnx()
- How to create new variables in SAS
- How to impute missing values in data
- How to look for outliers in data

Assignment 2 for practicing data preparation session

Session 5-6 Data Manipulation

- Retain Function
- FIRST. / LAST. In SAS
- How to use Array?
- Loops in SAS- while, until, Case statements
- Merging data in SAS- Inner Join, Left join, right Join, Full join
- One to one join, Many to Many, Many to One, One to Many joins
- Appending the datasets –Union, Intersect, Except
- Enhancing Output with ODS

Session 7-8 Data Analysis

- Procedures in SAS (Procs)
- Proc Print – Printing Data
- Proc Sort- How to sort data
- Procs for how to remove duplicates from your data?
- Proc Means- Summarizing data
- Proc Summary
- Proc Tabulate
- Proc Transpose
- Proc Append
- Proc Contents
- Proc Dataset
- Proc Freq

Assignment 3 for practicing data preparation session

Advance SAS

Session 9-10 Proc SQL (Structured Query Language)

- Introduction to SQL
- Select clause
- Data set Manipulations (creating, updating, deleting) using SAS SQL
- Use of conditional statements (where statement, Between, And, etc.)
- Group By, Order By, Having expressions in SQL
- Dataset Join -Inner Join, Left Join, Right Join, Full Join

- Subqueries In SQL
- SAS vs SQL many to many join- How they differ
- Comparing two tables
- Find records only exist in one table
- Random Sampling with PROC SQL
- Alternative to _N_ in PROC SQL
- NODUPKEY with PROC SQL
- Use DISTINCT in CASE WHEN

Session 11-12 SAS Macros

- Introduction to SAS Macro, Macro Variable & Macro Program
- Macro Code Constituents
- Macro Functions
- Macro Processing
- Ways to create macro variable
- Concatenation of Macro Variables
- How to store list of values in a macro variable
- How to debug SAS Macros

Phase 2: Statistics

Session 13-14 Introduction to Business Analytics

- Relevance in industry and need of the hour
- Types of analytics – Marketing, Risk, Operations etc.
- Future of analytics and critical requirement

Fundamentals of Statistics

- Basic statistics; descriptive and summary
- Inferential statistics
- Statistical tests

Basic Analytics

- Statistics Basics Introduction to Data Analytics and Statistical Techniques
- Types of Variables, measures of central tendency and dispersion
- Variable Distributions and Probability Distributions
- Normal Distribution and Properties
- Correlation
- Hypothesis Testing Null/Alternative Hypothesis formulation
- P Value Interpretation

Phase 3: Machine Learning/ Statistical Modeling

Session 15-16 Linear Regression Model

- Basics of regression analysis
- Correlation, VIF, missing value imputations and outliers
- Create Linear regression model
- Interpretation of results
- Performance metrics for model.

Hands on project for implementing Linear Regression

Session 17-18 Logistic Regression Model

- Use cases of Logistic regression model.
- Create a logistic regression model in R
- Churn prediction models and management
- Sensitivity, specificity, Confusion matrix.
- ROC curve.
- Performance metrics of logistic regression

Hands on project for implementing Logistic Regression

Session 19-20 Segmentation/Cluster Analysis

- What is K-means clustering model
- Create a clustering model in R.
- Interpreting results to select numbers of clusters for model.
- Checking accuracy of the model.

Hands on project for implementing Clustering

Session 21 Other Machine Learning techniques

- Time series
- Survival Analysis

Real Time Industry Projects

Project 1: Predicting who will default and who will not for a bank loan or Credit Card

Project 2: Predicting the optimal price of a car to be launched

Project 3: Segmentation of customer base for a retail chain

Project 4: (Case study in any one domain, 3 session)

Specialization in Banking

Specialization in Healthcare

Specialization in Telecom

