1 Data Science With SAS Training (SAS Programmer)

SAS® Base

1 SAS® Programming 1: Essentials

2 Introduction to SAS® foundation

3 Introduction to SAS® programs

4 Accessing SAS® Data
   - Examining SAS® data sets
   - Accessing SAS® libraries

5 Producing Detail Reports
   - Submitting report data
   - Sorting and grouping report data
   - Enhancing reports

6 Formatting Data Values
   - Using SAS® formats
   - Creating user-defined formats

7 Reading SAS® Data Sets
   - Reading a SAS® data set
   - Customizing a SAS® data set

8 Reading Spreadsheet and Database Data
   - Reading spreadsheet data
   - Reading database data

9 Reading Spreadsheet and Database Data
   - Reading spreadsheet data
   - Reading database data
   - Introduction to reading raw data files
   - Reading standard delimited data
Data Scientist (5 Courses in One) Online Training

- Reading nonstandard delimited data
- Handling missing data

10 Manipulating Data
- Using SAS® functions
- Conditional processing

11 Combining SAS® Data Sets
- Concatenating data sets
- Merging data sets one-to-one
- Merging data sets one-to-many
- Merging data sets with non-matches

12 Creating Summary Reports
- Reading spreadsheet data
- Reading database data

SAS® Programming 2: Data Manipulation Techniques

1 Introduction
- An overview of SAS® foundation
- Course logistics
- Course data files

2 Controlling Input and Output
- Writing observations explicitly
- Writing to multiple SAS® data sets
- Selecting variables and observations

3 Summarizing Data
- Creating an accumulating total variable
- Accumulating totals for a group of data

4 Reading Raw Data Files
- Reading raw data files with formatted input
- Controlling when a record loads

Data Transformations
- Manipulating character values
Data Scientist (5 Courses in One) Online Training

- Manipulating numeric values
- Converting variable type

5 Debugging Techniques
- Using the PUTLOG statement

6 Processing Data Iteratively
- DO loop processing
- Conditional DO loop processing
- SAS® array processing
- Using SAS® arrays

7 Restructuring a Data Set
- Rotating with the DATA step

8 Combining SAS® Data Sets
- Using data manipulation techniques with match-merging

SAS® Advanced

SAS® SQL 1: Essentials

1 Introduction
- Overview of SAS® Foundation
- Course logistics
- Course data files
- Introducing the Structured Query Language
- Overview of the SQL procedure
- Specifying columns
- Specifying rows

2 Displaying Query Results
- Presenting data
- Summarizing data

3 SQL Joins
- Introduction to SQL joins
- Inner joins
- Outer joins
Data Scientist (5 Courses in One) Online Training

- Complex SQL joins

4 Subqueries
- Noncorrelated subqueries
- in-line views

5 Set Operators
- Introduction to set operators
- UNION operator
- OUTER UNION operator
- EXCEPT operator
- INTERSECT operator

6 Creating Tables and Views
- Creating tables with the SQL procedure
- Creating views with the SQL procedure

7 Advanced PROC SQL Features
- Dictionary tables and views
- Using SQL procedure options
- Interfacing PROC SQL with the macro language

SAS® Macro Language 1: Essentials

1 Introduction
- Course logistics
- Purpose of the macro facility
- Program flow

2 Macro Variables
- Introduction to macro variables
- Automatic macro variables
- Macro variable references
- User-defined macro variables
- Delimiting macro variable references
- Macro functions
3 Macro Definitions
- Defining and calling a macro
- Macro parameters

4 DATA Step and SQL Interfaces
- Creating macro variables in the datastep
- Indirect references to macro variables
- Creating macro variables in SQL

5 Macro Programs
- Conditional processing
- Parameter validation
- Iterative processing
- Global and local symbol tables

2 Data Science Certification Training - R Programming

Module 1: Essential to R programming

1 An Introduction to R
- History of S and R
- Introduction to R
- The R environment
- What is Statistical Programming?
- Why use a command line?
- Your first R session

2 Introduction to the R language
A) Starting and quitting R
- Recording your work

B) Basic features of R
- Calculating with R
- Named storage
- Functions
- Exact or approximate?
- R is case-sensitive
C) Built-in functions and online help
- Built-in examples
- Finding help when you don’t know the function name
- Built-in graphics functions
- Additional elementary built-in functions

D) Logical vectors and relational operators
- Boolean algebra
- Logical operations in R
- Relational operators
- Data input and output
- Changing directories
- dump() and source()
- Redirecting R output
- Saving and retrieving image files
- Data frames and the read.table function

3 Programming statistical graphics
A) High-level plots
- Bar charts and dot charts
- Pie charts
- Histograms
- Box plots
- Scatterplots
- QQ plots
B) Choosing a high-level graphic

C) Low-level graphics functions
- The plotting region and margins
- Adding to plots
- Setting graphical parameters

4 Programming with R

A) Flow control
- The for () loop
- The if () statement
- The While () loop
- Newton’s method for root finding
- The repeat loop, and the break and next statements

B) Managing complexity through functions
- What are functions?
- Scope of variables

C) Miscellaneous programming tips
- Using fix ()
- Documentation using#

D) Some general programming guidelines
- Top-down design

E) Debugging and maintenance
- Recognizing that a bug exists
- Make the bug reproducible
- Identify the cause of the bug
- Fixing errors and testing
- Look for similar errors elsewhere
- The browser () and debug () functions

F) Efficient programming
- Learn your tools
- Use efficient algorithms
- Measure the time your program takes
- Be willing to use different tools
5 Simulation

A) Monte Carlo simulation

B) Generation of pseudorandom numbers

C) Simulation of other random variables
   - Bernoulli random variables
   - Binomial random variables
   - Poisson random variables
   - Exponential random numbers
   - Normal random variables

D) Monte Carlo integration
   - Advanced simulation methods
   - Rejection sampling
   - Importance sampling

6 Computational linear algebra

A) Vectors and matrices in R
   - Constructing matrix objects
   - Accessing matrix elements; row and column names
   - Matrix properties
   - Triangular matrices
   - Matrix arithmetic

B) Matrix multiplication and inversion
   - Matrix inversion
   - The LU decomposition
   - Matrix inversion in R
   - Solving linear systems

C) Eigenvalues and eigenvectors
   - Advanced topics
   - The singular value decomposition of a matrix
7 Numerical optimization

A) The golden section search method

B) Newton–Raphson

C) The Nelder–Mead simplex method

D) Built-in functions

E) Linear programming
   - Solving linear programming problems in R
   - Maximization and other kinds of constraints
   - Special situations
   - Unrestricted variables
   - Integer programming
   - Alternatives to lp ()
   - Quadratic programming

Module 2: Data Manipulation Techniques using R programming

1 Data in R
Modes and Classes
Data Storage in R
Testing for Modes and Classes
Structure of R Objects
Conversion of Objects
Missing Values
Working with Missing Values

2 Reading and Writing Data
Reading Vectors and Matrices
   A) Data Frames: read.table
B) Comma-and Tab-Delimited Input Files

C) Fixed-Width Input Files

D) Extracting Data from R Objects

E) Connections

F) Reading Large Data Files

G) Generating Data
   - Sequences
   - Random Numbers
   - Permutations
   - Random Permutations
   - Enumerating All Permutations

H) Working with Sequences

I) Spreadsheets
   - The RODBC Package on Windows
   - The gdata Package (All Platforms)

J) Saving and Loading R Data Objects

K) Working with Binary Files

L) Writing R Objects to Files in ASCII Format
   - The write Function
   - The write.table function
   - Reading Data from Other Programs

3 R and Databases

A) A Brief Guide to SQL
   - Navigation Commands
   - Basics of SQL
Data Scientist (5 Courses in One) Online Training

- Aggregation
- Joining Two Databases
- Subqueries
- Modifying Database Records

B) ODBC
C) Using the RODBC Package
D) The DBI Package
E) Accessing a MySQL Database
F) Performing Queries
G) Normalized Tables
H) Getting Data into MySQL
I) More Complex Aggregations

4 Dates
- as.Date
- The chron Package
- POSIX Classes
- Working with Dates
- Time Intervals
- Time Sequences

5 Factors
- Using Factors
- Numeric Factors
- Manipulating Factors
- Creating Factors from Continuous Variables
- Factors Based on Dates and Times
- Interactions

6 Subscripting
- Basics of Subscripting
- Numeric Subscripts
- Character Subscripts
- Logical Subscripts
- Subscripting Matrices and Arrays
- Specialized Functions for Matrices
- Lists
- Subscripting Data Frames
7 Character Manipulation
- Basics of Character Data
- Displaying and Concatenating Character
- Working with Parts of Character Values
- Regular Expressions in R
- Basics of Regular Expressions
- Breaking Apart Character Values
- Using Regular Expressions in R
- Substitutions and Tagging

8 Data Aggregation
- Table
- Road Map for Aggregation
- Mapping a Function to a Vector or List
- Mapping a function to a matrix or array
- Mapping a Function Based on Groups
- There shape Package
- Loops in R

9 Reshaping Data
- Modifying Data Frame Variables
- Recoding Variables
- The recode Function
- Reshaping Data Frames
- The reshape Package
- Combining Data Frames
- Under the Hood of merge

Module 3: Statistical Applications using R programming

1 Basics
   A) First steps
      - An overgrown calculator
      - Assignments
      - Vectorized arithmetic
      - Procedures
      - Graphics
B) R language essentials
- Expressions and objects
- Functions and arguments
- Vectors
- Quoting and escape sequences
- Missing values
- Functions that create vectors
- Matrices and arrays
- Factors
- Lists
- Data frames
- Indexing
- Conditional selection
- Indexing of data frames
- Grouped data and data frames
- Implicit loops
- Sorting

2 The R environment
A) Session management
- The workspace
- Textual output
- Scripting
- Getting help
- Packages
- Built-in data
- attach and detach
- subset, transform, and within

B) The graphics subsystem
- Plot layout
- Building a plot from pieces
- Using par
- Combining plots

C) R programming
- Flow control
- Classes and generic functions
D) Data entry
- Reading from a text file
- Further details on `read.table`
- The data editor
- Interfacing to other programs

3 Probability and distributions
A) Random sampling

B) Probability calculations and combinatorics

C) Discrete distributions

D) Continuous distributions

E) The built-in distributions in R
- Densities
- Cumulative distribution functions
- Quantiles
- Random numbers

4 Descriptive statistics and graphics
A) Summary statistics for a single group

B) Graphical display of distributions
- Histograms
- Empirical cumulative distribution
- Q–Q plots
- Boxplots

C) Summary statistics by groups

D) Graphics for grouped data
- Histograms
- Parallel boxplots
- Stripcharts
E) Tables
- Generating tables
- Marginal tables and relative frequency

F) Graphical display of tables
- Barplots
- Dotcharts
- Piecharts

5 One-and two-sample tests
- One-sample t test
- Wilcoxon signed-rank test
- Two-sample t test
- Comparison of variances
- Two-sample Wilcoxon test
- Paired t test
- Matched-pairs Wilcoxon test

6 Regression and correlation
- Simple linear regression
- Residuals and fitted values
- Prediction and confidence bands
- Correlation
- Pearson correlation
- Spearman’s $\rho$
- Kendall’s $\tau$

7 Analysis of variance and the Kruskal–Wallis test

A) One-way analysis of variance
- Pairwise comparisons and multiple testing
- Relaxing the variance assumption
- Graphical presentation
- Bartlett’s test

B) Kruskal–Wallis test

C) Two-way analysis of variance
- Graphics for repeated measurements
D) The Friedman test

E) The ANOVA table in regression analysis

8 Tabular data
- Single proportions
- Two independent proportions
- k proportions, test for trend
- r × c tables

9 Power and the computation of sample size
A) The principles of power calculations
   - Power of one-sample and paired t tests
   - Power of two-sample t test
   - Approximate methods
   - Power of comparisons of proportions

B) Two-sample problems

C) One-sample problems and paired tests

D) Comparison of proportions

10 Advanced data handling
A) Recoding variables
   - The cut function
   - Manipulating factor levels
   - Working with dates

B) Recoding multiple variables

C) Conditional calculations

D) Combining and restructuring data frames
   - Appending frames
   - Merging data frames
   - Reshaping data frames
   - Per-group and per-case procedures
   - Time splitting
11 Multiple Regression
- Plotting multivariate data
- Model specification and output
- Model search

12 Linear models
A) Polynomial regression
B) Regression through the origin
C) Design matrices and dummy variables
D) Linearity over groups
E) Interactions
F) Two-way ANOVA with replication
G) Analysis of covariance
   - Graphical description
   - Comparison of regression lines
H) Diagnostics

13 Logistic regression
A) Generalized linear models
B) Logistic regression on tabular data
   - The analysis of deviance table
   - Connection to test for trend
C) Likelihood profiling
D) Presentation as odds-ratio estimates
E) Logistic regression using raw data
F) Prediction
G) Model checking

14 Survival analysis
- Essential concepts
Data Scientist (5 Courses in One) Online Training

- Survival objects
- Kaplan–Meier estimates
- The log-rank test
- The Cox proportional hazards model

15 Rates and Poisson regression

A) Basic ideas
   - The Poisson distribution
   - Survival analysis with constant hazard
B) Fitting Poisson models
C) Computing rates
D) Models with piecewise constant intensities

16 Nonlinear curve fitting

- Basic usage
- Finding starting values
- Self-starting models
- Profiling
- Finer control of the fitting algorithm

3 Big Data Hadoop And Spark Developer

1 Big Data Hadoop

1. Introduction
   - About this Course
   - About Big Data
   - Course Logistics
   - Introductions

2. Hadoop Fundamentals
   - The Motivation for Hadoop
   - Hadoop Overview
   - HDFS
   - MapReduce
   - The Hadoop Ecosystem
   - Lab Scenario Explanation
   - Hands-On Exercise: Data Ingest with Hadoop Tools
3. Introduction to Pig
- What Is Pig?
- Pig’s Features
- Pig Use Cases
- Interacting with Pig

4. Basic Data Analysis with Pig
- Pig Latin Syntax
- Loading Data
- Simple Data Types
- Field Definitions
- Data Output
- Viewing the Schema
- Filtering and Sorting Data
- Commonly-Used Functions
- Hands-On Exercise: Using Pig for ETL Processing

5. Processing Complex Data with Pig
- Storage Formats
- Complex/Nested Data Types
- Grouping
- Built-in Functions for Complex Data
- Iterating Grouped Data
- Hands-On Exercise: Analyzing Ad Campaign Data with Pig

6. Multi-Dataset Operations with Pig
- Techniques for Combining Data Sets
- Joining Data Sets in Pig
- Set Operations
- Splitting Data Sets
- Hands-On Exercise: Analyzing Disparate Data Sets with Pig

7. Extending Pig
- Adding Flexibility with Parameters
- Macros and Imports
- UDFs
- Contributed Functions
- Using Other Languages to Process Data with Pig
8. Pig Troubleshooting and Optimization
- Troubleshooting Pig
- Logging
- Using Hadoop’s Web UI
- Optional Demo: Troubleshooting a Failed Job with the Web UI
- Data Sampling and Debugging
- Performance Overview
- Understanding the Execution Plan
- Tips for Improving the Performance of Your Pig Jobs

9. Introduction to Hive
- What Is Hive?
- Hive Schema and Data Storage
- Comparing Hive to Traditional Databases
- Hive vs. Pig
- Hive Use Cases
- Interacting with Hive

10. Relational Data Analysis with Hive
- Hive Databases and Tables
- Basic HiveQL Syntax
- Data Types
- Joining Data Sets
- Common Built-in Functions
- Hands-on Exercise: Running Hive Queries on the Shell, Scripts, and Hue

11. Hive Data Management
- Hive Data Formats
- Creating Databases and Hive-Managed Tables
- Loading Data into Hive
- Altering Databases and Tables
- Self-Managed Tables
- Simplifying Queries with Views
- Storing Query Results
- Controlling Access to Data
- Hands-on Exercise: Data Management with Hive
12. Text Processing with Hive
   - Overview of Text Processing
   - Important String Functions
   - Using Regular Expressions in Hive
   - Sentiment Analysis and N-Grams
   - Hands-on Exercise (Optional): Gaining Insight with Sentiment Analysis

13. Hive Optimization
   - Understanding Query Performance
   - Controlling Job Execution Plan
   - Partitioning
   - Bucketing
   - Indexing Data

14. Extending Hive
   - SerDes
   - Data Transformation with Custom Scripts
   - User-Defined Functions
   - Parameterized Queries
   - Hands-on Exercise: Data Transformation with Hive

15. Introduction to Impala
   - What is Impala?
   - How Impala Differs from Hive and Pig
   - How Impala Differs from Relational Databases
   - Limitations and Future Directions
   - Using the Impala Shell

2 Apache Spark

1 An Introduction to Spark - Getting started
   - What is Spark and what is its purpose?
   - Components of the Spark unified stack
   - Resilient Distributed Dataset (RDD)
   - Downloading and installing Spark standalone
   - Scala and Python overview
   - Launching and using Spark’s Scala and Python shell
2 About Resilient Distributed Dataset and DataFrames
- Understand how to create parallelized collections and external datasets
- Work with Resilient Distributed Dataset (RDD) operations
- Utilize shared variables and key-value pairs

3 The Spark application programming
- Understand the purpose and usage of the SparkContext
- Initialize Spark with the various programming languages
- Describe and run some Spark examples
- Pass functions to Spark
- Create and run a Spark standalone application
- Submit applications to the cluster

4 An Introduction to Spark libraries
- Understand and use the various Spark libraries

5 About Spark configuration, monitoring and tuning
- Understand components of the Spark cluster
- Configure Spark to modify the Spark properties, environmental variables, or logging properties
- Monitor Spark using the web UIs, metrics, and external instrumentation
- Understand performance tuning considerations

4 Business Analytics with Excel (with VBA)

1. Macro Introduction
- What is Macro?
- How to create and Execute Macro?
- What is the structure of an Excel workbook from VBA point of view?
- Important Coding platforms you should remember
- Getting started with Recording Macros -Some Example
- Key things to know about MS Excel Objects like Applications, Work sheets, Workbook, etc
2. Learn About Programming
- Objectives and start and end points
- Identify Sequential Steps of Programming
- Concepts of Create Programming
- Understanding of Object-Oriented Programming
- Understanding of Life Cycle of Programming

3. Introduction to Flow Chart & Diagram
- What is Flow Chart?
- What is Process?
- What is Attribute?
- Symbols used in DFD

4. Getting Started with VBA
- Definition
- Properties window
- Use project explorer
- VBA Code Security
- Compile & Debug Applications
- Tips and Tricks
- Option Explicit

5. VBA Excel Object
- Excel VBA Objects - Workbook Object, Worksheet Object & Range Object
- Working with Excel VBA Controls
- Procedures: Sub Procedure, Function Procedure & String Handling
- Date and Time Function

6. Operators, Data Types & Variables
- Definition
- Key types of Variables & Operators
- How and when to use variables to store information
- Typical Variable naming prototype you should learn
- When to use constants or Variables more on Excel Macros - II Loops
- Types of Datatypes

7. Conditional Statement
- Types of IF Conditions: - IF Then Construct
- IF Then Statement Else Statement
IF Then Statement Elseif Statement Else Statement
Using of IIf (expr, truepart, falsepart)
Select Case Statement

8. Loop Statement
- Definition
- For-Next Loop
- Do While Loop, Do Loop, Do Until Loop

9. VBA User Forms-Part-1
- Introduction of User Forms
- User Form Controls you must learn
- Make the most of text boxes, label, drop down, list box and image box
- Use Message Box & Input Box in VBA
- Use Forms events

10. VBA User Forms-Part-2
- Displaying a progress indicator
- Creating a wizard — an interactive series of dialog boxes
- Add, Update, Next & Previous Data using of VBA
- Searching Data using VBA

11. Advance VBA Programming
- Displaying a Chart using VBA
- Working with Pivot Tables
- How to Add Own Controls in Excel Ribbon
- Shell Functions
- Error Handling
- What is ADO
- Connecting VBA to MS-Access

12. Dashboard
- Objective of dashboard
- Define KPIs (Key performance Indicator)
- Dashboard reports based on Tables and Number or Charts/Graphs or Both

5 Data Science with Python
1 The Data Science: An Overview
- Introduction to the Data Science
- Different Sectors Using Data Science
- The Purpose and Components of Python

2 Data Analytics Overview
- The Data Analytics Process
- Exploratory the Data Analysis (EDA)
- EDA-Quantitative Technique
- EDA - Graphical Technique
- The Data Analytics Conclusion or Predictions
- The Data Analytics Communication
- The Data Types for Plotting

3 Statistical Analysis and Business Applications
- Introduction to the Statistics
- About Statistical and Non-statistical Analysis
- The Major Categories of Statistics
- About the Statistical Analysis Considerations
- The Population and Sample
- What is the Statistical Analysis Process?
- The Data Distribution
- Dispersion

4 Python Environment Setup and Essentials
- About the Anaconda
- The Installation of Anaconda Python Distribution
- Data Types in the Python
- Basic Operators and Functions

5 What is Mathematical Computing with Python (NumPy)?
- An Introduction to the Numpy
- The Activity-Sequence it Right
- Class and Attributes of ndarray
- All About the Basic Operations
- Activity-Slice It
- Copy and Views
- About the Mathematical Functions of Numpy
6 The Scientific computing with Python (Scipy)
- Introduction to the SciPy
- About the SciPy Sub Package - Integration and Optimization
- What is SciPy sub package?
- Know About the SciPy Sub Package - Statistics, Weave and IO

7 The Data Manipulation with Pandas
- Introduction to the Pandas
- Understanding DataFrame
- The Missing Values
- The Data Operations
- About File Read and the Write Support
- What is Pandas Sql Operation?

8 The Natural Language Processing with Scikit Learn
- NLP: An Overview
- What are NLP Applications?
- About NLP Libraries-Scikit
- The Extraction Considerations
- The Scikit Learn-Model Training and Grid Search

9 The Data Visualization in Python using matplotlib
- Introduction to the Data Visualization
- What are Line Properties?
- (x,y) Plot and Subplots
- The Types of Plots

10 Web Scraping with BeautifulSoup
- Web Scraping and Parsing
- Understanding and Searching the Tree
- Know the Navigating options
- Know About Modifying the Tree
- How to Parse and Print the Document?

11 Python integration with Hadoop MapReduce and Spark
- Know Why Big Data Solutions are provided for Python?
- Describing Hadoop Core Components
- The Python Integration with HDFS using Hadoop Streaming
The Python Integration with Spark using PySpark