

Ivy's Big Data Analytics Certification Details (Module I + II+ III + IV + V)
Based on Industry Cases, Live Exercises, & Industry Executed Projects

Module (I) - Analytics Essentials – 81 hrs

1. Statistics Refresher Overview (6 Hrs)

- Population metrics: Mean, Median, Mode, Std Deviation, Variance, Co-variance
- Introduction to types of distributions – Normal, Poisson, Binomial, Gamma
- Hypothesis Testing – t Test, F Test etc
- Ordinary Least Squares method for regression

2. Advanced Analysis & Reporting on Excel using VBA, Macros (45 hrs)

- Introduction to MS Excel, Basic Functions and Usage
- Sorting, Filtering, Advance Filtering, Subtotal
- Pivot Tables
- Data Entry Forms
- What-if-Analysis – Data Tables, Scenario Manager & Goal Seek
- Different Charts Graphs – Which one to use and when
- Vlookup, Hlookup, Match, Index
- Style & Number Formatting, Conditional Formatting
- Worksheet & Workbook Reference
- Smart Printing Options
- Logical Operators & Functions – IF and Nested IF
- Data Validation
- Text Functions
- Relative & Absolute Macro Recording
- VBA Programming
 - Making Macro do Automated Tasks for You
 - Record a Macro
 - Automate a Task using Macro
 - Analysis Using VBA
 - Objects
 - VBA Collection Objects
 - The Application Object
 - Workbook and Window Objects
 - The Range Objects
 - Variable, Data Types, and Constants
 - Types of Variables
 - Promoting for User Input
 - Arrays
 - Interacting with Users
 - VBA Sub, Event and Function Procedures
 - Create Sub Procedures, Event Procedures and Function Procedures

- Ways to Execute these Procedures
- How to Execute Function Procedures
- Control Logic and Loops
 - Conditional Logic
 - Conditional Statements and the If/Then/Else and Select/Case Code Structures
 - Looping
- VBA User Forms Overview
- Industry Projects using VBA

3. SQL based Relational Database Concepts using MS-Access (15 hrs)

- Relational Database Fundamentals
- Steps to Design Efficient Relational Database Models
- Case Studies on Designing Database Models
- Case Study Implementation on Handling Data in MS-Access 2007
- Importing / Exporting Large Amount of Data into/from MS-Access
- Intro to SQL Statements
- Writing Transactional SQL Queries, Merging, joining, sorting, indexing, etc.
- Hands-on Exercises on Manipulating Data Using SQL Queries
- Creating Database Models Using SQL Statements
- Individual Projects on Handling SQL Statements

4. Structured Thinking Workshops (9 Hrs)

- Analytical Problem Solving Skills, Creative Thinking, Root Cause Analysis, Decision Tree Analysis, Fish Bone Analysis, Mind Mapping, Minto Pyramid's Principle, etc.
- Techniques to Handle Case Based Interviews
- Group / Individual Exercises / Case Studies

5. Effective Interviewing Skills Workshop (3 Hrs)

- Myths about Interviews and Common Mistakes
- Convincing the Interviewer
- Correct Preparation for Interview and its Advantages
- The 4 Steps to Interview Success
- Components of Good Interview
- How to face common interview questions

6. Analytics Industry Resume Building Workshop (3 Hrs)

- What do Analytics Recruiters Look for in a Resume?
- Analytics Resume Best Practices
- Analytics Industry Keywords to Include in Resume

- How to Describe Relevant Analytics Projects in Resume?
- Effective Strategies to Get Your Resume Noticed

7. Ivy's Career Center Workshops (These workshop hours are not counted towards the total duration)

- Industry Networking Sessions with At Least Two Industry Guest Speakers
- Mock Interview Sessions with Industry Professionals
- Analytics Company-wise Interview Questions (Collected from our Alumni and Recruiters)
- Comprehensive Analytics Interview Question Banks (SAS, VBA, Behavioral, etc)

(Module II)- Industry Applicable Core Analytics with Case Studies (21 Hrs)

1. Importance of Data for Management Decisions

- Data Types
- Classifying Data to Convey Meaning
- Diagrammatic and Graphic Representation
- Measures of Central Tendency and Dispersion
- Data Modeling
- Ratios and Index Numbers
- Discussion with cases and examples

2. Time Series Analysis

- Mathematical Models for Time Series
 - Measurement of Seasonal, Cyclical and Irregular Variations
 - Discussion with cases and examples
3. Probability Concepts and Applications
 - Bayesian Theory
 - Probability Distributions and Mathematical Expectation
 - Discussion with cases and examples
 4. Basics of Sampling and Sampling Distribution
 5. Theory of Estimation and Testing of Hypothesis
 - Discussion with cases and examples
 6. Correlation and Regression Models
 7. Forecasting
 - Discussion with cases and examples
 8. Theory of Attributes
 - Discussion with cases and examples
 9. Statistical Decision Analysis
 - Discussion with cases and examples
 10. Analysis of Variance
 - Discussion with cases and examples
 11. Multivariate Analysis
 - Discussion with cases and examples

Module (III) - Advanced Analytics using SAS 9 (Base, Advance and Stats Modeling) – 62 hrs

1. SAS Analytics

- Introduction to SAS & Data Processing, SAS Data steps
- SAS Functions
- Types of variables, Variable Formats
- SAS procedures including Proc SQL
- SAS Macros
- Statistical Modelling Concepts and Practical Assignments-
 - Regression
 - ARIMA
 - Clustering

- Logistic
- Time Series Analysis
- Handling Data in SAS – Case Studies I, II, and III
- Assignments / Exercises

2. Industry Executed Analytics Projects on SAS

- Dedicated Practice Sessions on Industry Scenarios in SAS
- Industry Executed **Predictive Modeling** Projects in SAS (5 Projects from Retail, Healthcare, Finance, Education, Aviation industries)

Please Note: For minute details on any of the above modules, please reach out to us through phone (+91 9748 441111 / +91 33 40011221) or email (info@ivyproschoo.com).

Module (IV): Data Management & Analysis using “R” – 25 hrs

Course Outline

Data Handling, Visualization and Statistical Modeling in R

- Data Handling and Visualization in R
 - Introduction and Overview of R Package
 - Data Cleaning and Management in R
 - Logic Building in R
 - Data Visualization in R
 - **Exercise: Data Summarization using Financial Retail Datasets**
- Data Modeling & Statistics Refresher
 - Data Modeling Techniques Overview
 - Missing Imputations
 - Multi-collinearity Check
 - Hypothesis Testing
 - **In-Case Study: Academic Performance Case Study**
 - **Self-Case Study: Health Care Case Study**
- Linear Regression
 - Regression
 - BLUE Property
 - Residual Analysis
 - Multiple Regression

- Model Building
- **In-class Case Study: Predict Academic Performance of School Students**
- **Self Case Study: Predict Customer Value for an Insurance Firm**
- Logistic Regression
 - Model theory, Model Fit Statistics
 - Reject Reference, Binning, Classing
 - Dummy Creation, Dummy Correlation
 - Model Development (Multicolinearity, WOE, IV, HLT, Gini KS, Rank Ordering, Clustering Check)
 - Model Validation (Rerun, Scoring)
 - Final Dashboard
 - **In-class Case Study: Predict Customer Churn for a Telecom firm**
 - **Self Case Study: Predict Propensity to Buy Financial Product among Existing Bank Customers**
- Factor Analysis & Clustering
 - Factor Analysis
 - Cluster Analysis
 - **In-class & Self Case Study: Loan Dataset**
- Time Series / ARIMA Forecasting
 - Univariate Time Series
 - ARIMA
 - **In-class Case Study: Forecast Gold Prices**
 - **Self Case Study: Forecast US Treasury Bond Prices**
- **Text Mining & Analysis**
 - Extract & Process Unstructured Text Data from Social Media Channels
 - How to Choose Correct Statistical Tools for Text Analysis
 - **In-class Case Study: Brand Specific Twitter Stream Data**
 - **Self Case Study: Brand Impact of Maggi Disaster on Nestle**
- **Introduction to Machine Learning**
 - How does Machine Learning Work?
 - ML using Random Forest
 - In-class Case Study: Create Ideal Home Buyer Profile

Module (V) – Hadoop – 24 Hrs

- **Introduction to Hadoop**
 - Distributed computing
 - Parallel computing
 - Concurrency
 - Cloud Computing
 - Data Past, Present and Future
 - Computing Past, Present and Future
 - Hadoop
 - NoSQL
- **Hadoop Stack**
 - MapReduceNoSQL
 - CAP Theorem
 - Databases: Key Value, Document, Graph
 - Hive and Pig
 - HDFS
- **Hadoop Hands-on**
 - Installing Hadoop Single Node cluster
 - Understanding Hadoop configuration files
- **MapReduce Introduction**
 - Functional □ Concept of Map
 - Functional □ Concept of Reduce
 - Functional □ Ordering, Concurrency, No Lock, Concurrency
 - Functional – Shuffling
 - Functional □ Reducing, Key, Concurrency
 - MapReduce Execution framework
 - MapReduce Partitioners and Combiners
 - MapReduce and role of distributed filesystem
 - Role of Key and Pairs
 - Hadoop Data Types
- **MapReduce Exercises**
 - Understanding Sample MapReduce code
 - Executing MapReduce code
- **HDFS Introduction**

- Architecture
- File System
- Data replication
- Name Node
- Data Node
- **Hive**
 - Architecture
 - Data Model
 - Physical Layout
 - DDL DML SQL Operations
- **Hive Hands on**
 - Installation
 - Setup
 - Exercises
- **Pig**
 - Rationale
 - Pig Latin
 - Input, Output and Relational Operators
 - User Defined Functions
 - Analyzing and designing using Pig Latin
- **Pig Hands on**
 - Installation
 - Setup
 - Executing Pig Latin scripts on File system
 - Executing Pig Latin scripts on HDFS
 - Writing custom User Defined Functions
- **Introduction to Sqoop**
- **Hadoop Multi node Cluster Setup**
 - Installation and Configuration
 - Running MapReduce Jobs on Multi Node cluster
- **Hadoop Project – Connecting Hadoop ecosystem with ETL Tool**
 - Moving data from local file system to HDFS
 - Writing Map Reduce Code
 - Moving data from HDFS to Database
 - Writing Pig Query
 - Writing Hive Query