

BRAINALYST
A Data Driven Company

FULL STACK DATA SCIENCE MASTER PROGRAM



2022/23

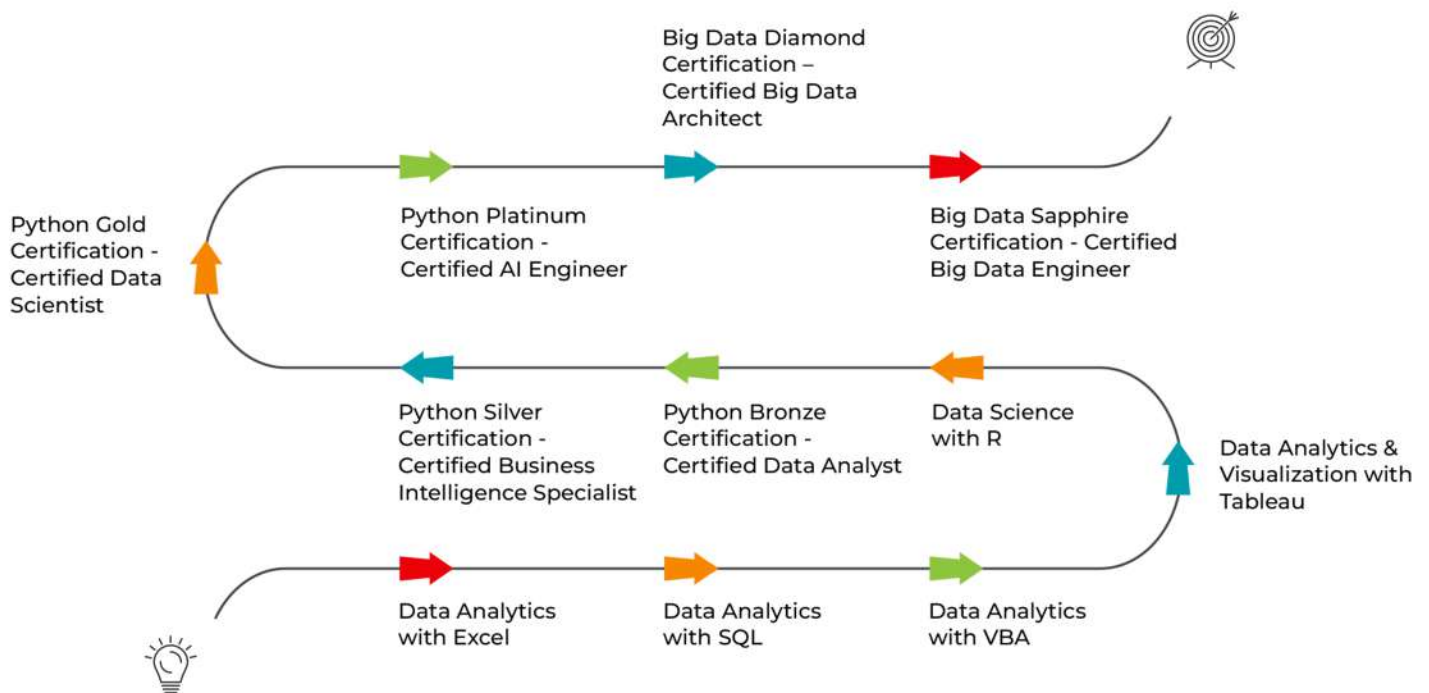
100% Placement Assurance*



Call: +91 - 90501 79137



Email: traning@brainalyst.in



Each module in the full stack course has its assignments, specific duration of hours, assignments and case studies.

Module	No. of Hours
Data Analytics with Excel	18
Data Analytics with VBA	18
Data Analytics with SQL	18
Data Analytics and Visualization with Tableau	18
Data Science with R	108
Python Bronze Certification - Certified Data Analyst	50
Python Silver Certification - Certified Business Intelligence Specialist	50
Python Gold Certification - Certified Data Scientist	50
Python Platinum Certification - Certified AI Engineer	50
Big Data Diamond Certification - Certified Big Data Architect	60
Big Data Sapphire Certification - Certified Big Data Engineer	60
TOTAL	500
Full Stack Course	

Data Analytics with Excel

Excel is a Stepping Stone in a Journey of Data Analytics where you Learn, how to store manipulate and analyze, small well structured data sets.



No. of Hours

18 Hours



**Assignment &
Case Study**

4

Overview :

Data Analytics with Excel :

The advanced Excel training is designed to help you become a skilled Data Analyst and is highly recommended when starting out in the field of Data Science, particularly Data Analytics. By using numerous case studies to ensure hands on experience, the advanced Excel module will teach an individual to make insightful dashboards, and will form the basis to later use VBA macros and SQL. Learning Excel will form the basis for dealing with small structured datasets. their coverage and overall health.



Chapter 1.1 : Basic Excel

- Excel Environment
- Key Terminologies
- Short Cuts
- Key Functionalities
- Copy-paste-paste special
- Formatting & conditional Formatting
- Basic Excel Functions – Types of Functions
- Relational operators
- Data Sorting, Filtering and Data Validation
- Understanding of Name Ranges
- Pivot tables – Charts
- Basics of charts

Chapter 1.2 : Basic Programming Elements

- Overview of programming languages
- Basics of programming elements
- variables, data types, data structures, loops, conditional statements, inputs, outputs, functions etc.
- Understanding key terms
- Client/server
- Database
- Hosting/deployment

Chapter 1.3 : Introduction of Basic Statistics

- Introduction to Statistics
- Measures of central tendencies
- Measures of variance
- Measures of frequency
- Measures of Rank
- Basics of Probability, distributions
- Conditional Probability (Bayes Theorem)

Chapter 1.4 : Overview of Dashboard

- What is dashboard & Excel dashboard
- Adding icons and images to dashboards
- Making dashboards dynamic

Chapter 1.5 : Business dashboard Creations

- Management Dashboard for Sales & Services
- Best practices –
Tips and Tricks to enhance dashboard designing

Chapter 1.6 : RDBMS AND SQL (Basic)

- Basic RDBMS Concepts
 - o Introduction to Relational Database management system. Why SQL?
 - o A glance at the tool and its advantages & disadvantages
 - o Understanding Schema, ERDs and Metadata
 - o Introduction to MS SQL Server
 - o What is SQL – A Quick Introduction
 - o Installing MS SQL Server for windows
- Introduction to SQL Server Management Studio
 - o Understanding basic database concepts
 - o Getting started

Chapter 1.7 : Data manipulation using functions

- Descriptive functions
- Logical functions: IF, and, or, not
- Date and Time functions
- Text functions
- Array functions
- Use and application of lookup functions
- Limitations of lookup functions
- Using Index, Match, Offset, reverse vlookup

Chapter 1.8 : Data Visualization in Excel

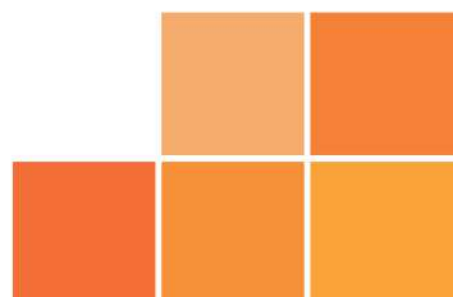
- Overview of chart types – column/bar charts, line/area , pie, doughnut charts, scatter plots
- How to select right chart for your data
- Creating and customizing advance charts – thermometer charts, waterfall charts, population pyramids

Chapter 1.9 : Introduction to Analytics & Data Science

- What is analytics & Data Science?
- Business Analytics vs. Data Analytics vs. Data Science
- Common Terms in Analytics
- Analytics vs. Data warehousing, OLAP, MIS Reporting
- Types of data (Structured vs. Unstructured vs. Semi Structured)
- Relevance of Analytics in industry and need of the hour
- Critical success drivers
- Overview of analytics tools & their popularity

Chapter 1.10 : Create dashboard in Excel - Using pivot controls

- Concept of pivot cache and its use in creating interactive dash- boards in excel
- Pivot table design elements – concept of slicers and timelines
- Designing sample dashboard using Pivot Controls
- Design principles for including charts in dashboards – do's and dont's



Data Analytics with SQL

When Dealing with Data sets, often the data is stored in Database management SQL- a query base language helps in collecting manipulating and analysing such data base



No. of Hours

18 Hours



**Assignment &
Case Study**

4

Overview :

Data Analytics with SQL :

The purpose of this course is to teach you how to extract and analyze data stored in databases using Structured Query Language (SQL). SQL is the most popular data analysis tool for data analysts and data scientists. Most of the data in the world is stored in databases, and learning SQL will make it possible to access and analyze this data easily. This course is suitable for candidates with various quantitative backgrounds who wish to learn job-oriented analytics and reporting skills using MS-Excel, VBA, MS-Access, and SQL, as well as Tableau.



Chapter 2.1 : Intro to RDBMS & Basic SQL

- Basic RDBMS Concepts
 - o Introduction to Relational Database management system. Why SQL?
 - o A glance at the tool and its advantages and disadvantages
 - o Understanding Schema, ERDs and Metadata
 - o Introduction to MS SQL Server
 - o What is SQL – A Quick Introduction
 - o Installing MS SQL Server for windows
- Introduction to SQL Server Management Studio
 - o Understanding basic database concepts
 - o Getting started

Chapter 2.2 : Data Based Object Creation (DDL commands)

- Creating databases and tables. Understanding data types
- Inserting values into the table
- Altering table properties
- Introduction to Keys and constraints
- Creating, Modifying & Deleting Tables
- Create Table & Create Index statements
- Drop & Truncate statements – Uses & Differences
- DDL Statements with constraints
- Import and Export wizard to get the data in SQL server from excel files or delimited files

Chapter 2.3: Data Manipulation

- Data Manipulation statements
- Insert, Update & Delete statements
- Select statement – Sub setting, Filters, Sorting. Removing Duplicates, grouping and aggregations etc.
- Operators, predicates and built in functions(Top, distinct, Limit)
- Where, Group By, Order by & Having clauses
- SQL Functions – Number, Text, Date, etc.
- SQL Keywords – Top, Distinct, Null, etc.
- SQL Operators – Relational (single valued and multi valued), Logical (and, or, not), Use of wildcard operators and wildcard characters, etc.

Chapter 2.4: Accessing data from multiple tables using Select

- Append and Joins
- Union and Union All – Use & constraints
- Intersect and Except statements
- Table Joins – inner join, left join, right join, full join
- Cross joins/cartesian products, self joins, natural joins etc
- Inline views and sub-queries & it's types
- Optimizing your work
- Update operations with and without joins

Chapter 2.5: Advanced SQL

- Creating table copy and database copy
- Views
- Transactions
- Stored Procedures in SQL
- Crud operations using stored procedures
- Window functions in SQL
- Miscellaneous Topics: Rollup and cube

Data Analytics with VBA

Get your first taste in programming with automating task in excel as often operation performed in Excel can be repetitive and here writing here writing macros can help.



No. of Hours

18 Hours



**Assignment &
Case Study**

2

Overview :

Data Analytics with VBA :

The Microsoft Excel VBA programming language allows you to automate, analyze, and manipulate data in Excel. Excel VBA is widely used for financial modeling and validation in high finance applications and frequency trading applications. The course introduction of various topics such as Logic, VBA, & Macro Recordings, IDE. Additionally, to multiple case studies used in class sessions, the course includes various assignments and projects for self-study and hands-on experience.



Chapter 3.1 : Data Analytics with VBA

- What is Logic?
- What Is VBA?
- Introduction to Macro Recordings, IDE

Chapter 3.2 : A look at some commonly used code snippets

- Explore few code snippets to understand the syntax
- Identify the commonly used codes

Chapter 3.3 : Programming Constructs in VBA

- Control Structures
- Looping Structures
- The With- End with Block

Chapter 3.4 : Functions & Procedures in VBA-Modularizing your programs

- Worksheet & workbook functions
- Automatic Procedures and Events
- Arrays

Chapter 3.5 : Objects & Memory Management in VBA

- The NEW and SET Key words
- Destroying Objects – The Nothing Keyword

Chapter 3.6 : Error Handling

- Explore numerous methodologies for handling errors in VBA
- Learn to read error message

Chapter 3.7 : Controlling accessibility of your code - Access Specifiers

- Understanding the access specifiers to set the accessibility of classes, methods, and other members
- Understand syntax used to facilitate the encapsulation of components

Chapter 3.8 : Code reusability - Adding references and components of your code

- Learning to use references to better deal with objects aiding in code reusability

Chapter 3.9 : How VBA works with Excel

- Working in the Visual Basic Editor
- Introducing the Excel Object Model
- Using the Excel Macro Recorder
- VBA Sub and Function Procedures

Chapter 3.10 : Key Component of programming language

- Essential VBA Language Elements
- Keywords & Syntax
- Programming statements
- Variables & Data types
- Comments
- Operators
- Working with Range Objects

Chapter 3.11 : Communicating with your users

- Simple Dialog Boxes
- User Form Basics
- Using User Form Controls
- Add-ins
- Accessing Your Macros through the User Interface
- Retrieve information through Excel from Access Database using VBA



Data Analytics and Visualization with Tableau

Visualisation is an important aspect of Data analytics as it helps in analytics and presenting the facts visually and an interactive data visualization software like tableau certainly helps.



No. of Hours

18 Hours



**Assignment &
Case Study**

3

Overview :

Data Analytics & Visualization with Tableau

A data visualization is a visual representation of information and data. A data visualization tool employs visual elements such as charts, graphs, and maps to let users view and understand trends, outliers, and patterns in data. As part of this course, you'll be able to learn skills such as MIS Reporting Analyses, Data Mining & Analysis, Data Visualization, and Data Blending & Manipulation. Additionally, to multiple case studies used in class sessions, the course includes various assignments and projects for self-study and hands-on experience.

Chapter 4.1 : Getting Started

- What is Tableau?
- Tableau product suite
- How Does Tableau Work?
- Tableau Architecture
- Connecting to Data & data source concepts
- Understanding the Tableau workspace
- Dimensions and Measures
- Data Types & Default Properties
- Tour of Shelves & Marks Card
- Using Show Me
- Saving and Sharing your work-overview

Chapter 4.2 : Data handlings & Summaries – I

- Date Aggregations and Date parts
- Cross tab & Tabular charts
- Totals & Subtotals
- Bar Charts & Stacked Bars
- Line Graphs with Date & Without Date
- Tree maps/Scatter Plots
- Individual Axes, Blended Axes, Dual Axes & Combination chart/Edit axis
- Parts of Views
- Sorting
- Trend/Reference Lines/Forecasting
- Filters/Context filters

Chapter 4.3 : Data handlings & Summaries – II

- Sets (In/Out Sets/Combined Sets)
- Grouping/Bins/Histograms
- Drilling up/down – drill through
- Hierarchies
- View data
- Actions (across sheets)

Chapter 4.4 : Building Advanced Reports/Maps

- Explain latitude and longitude
- Default location/Edit locations
- Building geographical maps
- Using Map layers

Chapter 4.5 : Calculated Fields

- Aggregate vs. Disaggregate data
- Explain – #Number of Rows
- Basic Functions (String/Date/Numbers etc)
- Usage of Logical conditions

Chapter 4.6 : Table Calculations

- Explain scope and direction
- Percent of Total, Running / Cumulative calculations
- Introduction to LOD (Level of Detail) Expressions
- User applications of Table calculations

Chapter 4.7 : Parameters

- Using Parameters in
 - o Calculated fields
 - o Bins/Reference Lines
 - o Filters/Sets
- Display Options (Dynamic Dimension/Measure Selection)
- Create What-If/ Scenario analysis

Chapter 4.8 : Building Interactive Dashboard

- Combining multiple visualizations into a dashboard (overview)
- Making your worksheet interactive by using actions
- Filter/URL/Highlight
- Complete Interactive Dashboard for Sales & Services

Chapter 4.9 : Building Stories

- Story Points
- Options in Formatting your Visualization
- Working with Labels and Annotations
- Effective Use of Titles and Captions

Chapter 4.10 : Work with data

- Multiple Table Join
- Data Blending
- Difference between joining and blending data, and when we should do each
- Toggle between to Direct Connection and Extracts

Chapter 4.11 : Sharing work with others

- Sharing Workbooks
- Publish to Reader/PDF
- Publish to Tableau Server and sharing on the web



Data Analytics and R

The language R can help you to perform advanced data analytics and being a statistical language assists you in creating statistical analytical scripts and predictive models.



No. of Hours

18 Hours



**Assignment &
Case Study**

6

Overview :

Data Science with R :

The comprehensive R training program will provide you with hands-on skills in Data Science with R. Since recent years, R has gained enormous popularity among Data Science practitioners and it is no surprise that the R language is often referred to as a lingua franca for Data Science. Through various examples and projects, this Data Science R course provides an effective introduction to basic data analytics, statistical prediction modelling, and machine learning.

Chapter 5.1 : Introduction to R

- Introduction to R language
- R for Data Science and Data Analytics
- Introduction to R Editors & IDE's (R Studio)

Chapter 5.2 : Installation of R Software

- Instructions for R & R-Studio Installation
- Differences between R-GUI and R Studio
- Overview of R-Studio & Customize Settings

Chapter 5.3 : Basics of R Programming

- Basics of R-Programming - Syntax Rules
 - o Case Sensitivity
 - o Use of Comments
 - o Object Naming Rules
 - o Operators
 - Assignment
 - Arithmetic
 - Relational
 - Logical
 - Package Reference
- Basic R-Commands & Functions
 - o Shortcuts
 - o Base functions (View, print etc)
- Debugging of code
 - o Reading preliminary errors

Chapter 5.4 : Modular Programming – Packages

- Introduction to Modular Programming & Packages
 - o Difference between System and User libraries
- Important packages in R for Data Science/Data Analytics
 - o Important Packages for data manipulation
 - o Important Packages for statistics
 - o Important Packages for visualization
 - o Important Packages for predictive modelling
- Installing & loading Packages
 - o Different methods of installing packages and pros and cons of them
 - Manual
 - `install.packages()`
 - GUI mechanism
- Different methods of loading packages and pros and cons of them
 - o `require()`
 - o `library()`
 - o GUI mechanism
- Name Spaces & Managing work environment
 - o Concept of Name Space
 - o Use of changing working directory
 - o Creating .RData data files
 - o Saving particular objects
 - o Saving all objects
 - o Auto Replacement of objects upon loading multiple .RData files
 - o Default Working Directory v/s Personalized Working Directory
 - o Removing objects

Chapter 5.5 : Data Types & Data structures

- R-Data types - strings, factors, numeric, date & time
 - o Different type of numeric based data types and differences between them
 - o Character v/s factor data type
 - o Logical data type
 - o Hierarchy of data types
 - o Dates
 - o Converting to date data type using `as.Date()`
 - o strp codes
 - o epoch and unix time 1970 concept
 - o Problems with date with year without century
 - o Using lubridate for converting to date data type
- Basic Operations: mathematical – string – date
 - o Concept of class
 - o Generic v/s Class based functions
 - o Exploring basic functions
- R-Data Structures - vectors, matrices, lists, data frames etc.
 - o Homogeneous v/s Heterogeneous data structures
 - o 1D v/s 2D v/s 3D data structures
 - o Vectors
 - Definition, Methods of creation and best practices
 - sub setting of elements
 - Concept of indexing, filtering
 - passing index numbers using `where()`
 - passing logical input using conditions
 - find and replace
 - o Matrix and List
 - Definition
 - Method of Creation
 - sub setting of elements



- o data.frame and data.table
 - Definition
 - Methods of Creation
 - sub setting of elements and best use practices

Chapter 5.6 : Other Programming Elements

- Control flow & conditional statements
 - o range based loops
 - o condition based loops
 - o making non-vectorised operations vectorised
- User defined functions
 - o Creation numerous UDF
- Concept of apply functions
 - o apply family of functions, difference and best use practices

Chapter 5.7 : Importing & Exporting data

- Importing Data from various sources (Csv, text, excel, xml, etc.)
 - o Use of changing directory
 - o Understanding delimited files
 - o importing files in different format
- Viewing Data objects
 - o Using View() function
 - o Exploring dataset
- Exporting Data to various formats
 - o Exporting dataset in multiple formats

Chapter 5.8 : Understanding of data

- Quick checks of data quality using various R-Functions
- View first and last few rows
 - o Column and Row Names
 - o Number of rows and columns
 - o The frequency, distinct values and proportion of the categories in a text based columns
 - o Arithmetic and Statistical properties of the Numerical columns
- Identifying key problems if there are any
 - o Wrong Data Type of the columns (Solution: Typecasting of the columns)
 - o Columns names not being as per R object naming rules (Solution: Renaming of the columns)
 - o Observations having Duplicates (Solution: Duplicate Removal)
 - o Missing Value in columns (Solution: Missing Value Treatment)
 - o Columns having Outliers (Solution: Outlier Capping)

Chapter 5.9 : Data Preparation/manipulation of data

- Overview of data preparation/Data Manipulation
 - o Basic EDA v/s Structure Based Data Manipulation v/s Content Based Data Manipulation
- Data preparation using single table
 - o Extraction of Columns
 - o Appending Columns
 - o Dropping Columns
 - o Renaming Columns
 - o Rearranging Columns
 - o 3 methods of filtering and pros, cons, best use practices



- o 2 methods of sorting, multi column sorting, basic use practices
- o transposing vs reshaping
- o wide to long and long to wide format
- o ID v/s numerical v/s categorical variable and confusions associated with it
- o Binning
- o One hot encoding
- o Label encoding
- o Aggregation
- Data preparation using multiple tables
 - o Merging, theory of joins, one to one, one to many and many to many relationship
 - o appending- rows v/s columns and common mistake associated with appending



Chapter 5.10 : Basic Statistics

- Basic Terms
 - o Samples v/s Population
 - o Statistic v/s Parameter
 - o Need of creating sample
 - o Methods of creating samples
 - o Descriptive v/s Inferential Statistics
- Descriptive Statistics
 - o Measure of Frequency
 - Visualization method
 - Frequency Table
 - o Measure of Central Tendency
 - Mean, Median, Mode
 - Pros and Cons
 - o Measure of Variability
 - Variability /s Central Tendency
 - Range v/s IQR v/s Sum of Absolute Deviation v/s Sum of Squared Deviation v/s Variance v/s Standard Deviation
 - o Measure of Shape
 - Symmetrical v/s Asymmetrical shape
 - Normal Distribution – characteristics etc
 - Different form of skewed distribution with examples
 - Kurtosis- Leptokurtic, Platykurtic, Mesokurtic with examples
 - Unimodal v/s Bimodal v/s Multimodal distributions
- Inferential Statistics
 - o Central Limit Theorem
 - o 3 Sigma Rule
 - o Finding Probability for a raw score
 - o Standard Normal Distribution
 - o Learning to read Z table for +ve and –ve z scores



o Z Test

- Difference between finding probability for a raw score v/s hypothesis testing
- Standard Error
- Null v/s Alternative Hypothesis
- Significance Level/Alpha Value determination
- Confidence Level
- p-value
- One tailed v/s Two Tailed Hypothesis Testing
- Standard Error: definition and usage

o t test

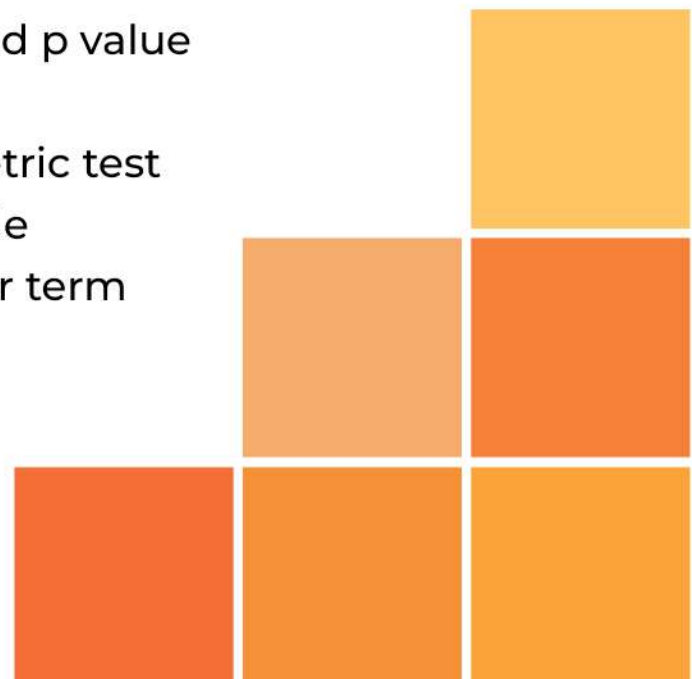
- One Sample t-test
 - Signal to Noise ratio formula of t formula
 - Best Case Use
- Dependent t test
 - Understanding Overlapping distributions
 - Understanding paired samples
- Independent t test
 - difference between equal and unequal variance

o f test

- ANOVA
 - Understanding different ANOVA
 - Understanding One Way ANOVA
 - Relationship between f and p value

o chi-squared test

- Parametric v/s Non-parametric test
- Expected v/s Observed Table
- Chi Square value as an error term



- Case Study
 - o Identification of correct test
 - o Running multiple test for a single problem to understand the involvement of time
- Others
 - o 6 sigma rule
 - o Industries working at different standard deviation
 - o Point Estimate v/s Confidence Interval

Chapter 5.11 : Data Visualization

- Overview of Data Visualization
 - o Business Driven Graphs
 - o Data Driven Graphs
 - o Univariate v/s Bivariate v/s Multivariate Graphs
 - o Use of different libraries
- Mapping Key graphs for various analysis
 - o Histogram
 - o Boxplot
 - o Line Chart
 - o Line of Best Fit
 - o Scatterplot
 - o Bar Chart
 - o Stacked Bar Chart
 - o Dodged Bar Chart
 - o Facet Bar Chart
 - o Pie Chart
 - o Cross Tab
 - o Word Cloud
 - o Radar Chart
 - o Waffle Chart
 - o Maps
 - o Introduction to Shiny



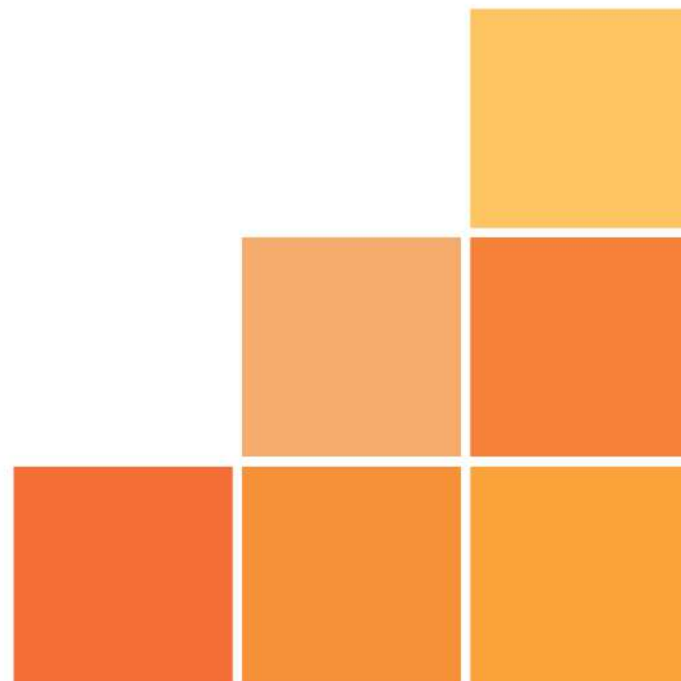
Chapter 5.12 : Shiny

- Intro to R
- UI-SERVER Connection
- Integration of Input Widgets in UI
- Integration of Input Widgets in Server
- Understanding observe, render and reactive functions
- Exploring different web layout using fluidPage
- Using HTML Codes in Shiny
- Applying themes
- Integration of tables
- Downloading from Shiny app
- Creating an EDA Tool

Chapter 5.13 : Introduction to Modelling

- Evolution of Predictive Models
 - o Stage 1 (Rule Based)
 - o Stage 2 (Statistical Models)
 - o Stage 3 (Machine Learning Models)
 - o Stage 4 (Deep Learning Models)
- Definitions
 - o Model, algorithm, dependent variable, independent variable, estimates/beta's etc.
- Types of Models
 - o Business/Statistical Problem
 - Regression
 - Classification
 - Segmentation
 - Forecasting
 - Optimization

- o Learning Setup
 - Supervised
 - Semi Supervised
 - Un-Supervised
 - Re-Inforcement
- o Business Objective
 - Strategic
 - Objective
- o Nature of Algorithm
 - Statistical
 - Mathematical
 - Rule Based
 - Probabilistic
 - Ensemble
- o Linear v/s Non-Linear
- o Parametric v/s Non-Parametric

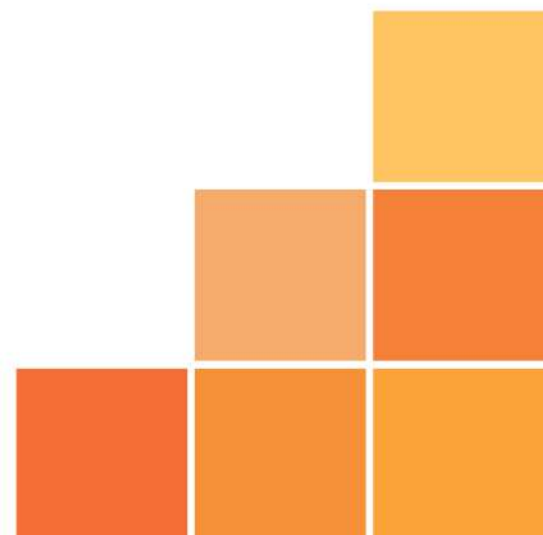


Chapter 5.14 : Linear Regression

- Understanding Equation
- Line of Best Fit
- Statistical Method of Finding Unknowns
 - o Manually Calculating beta, constant, SSE, SSR, SST, R², Adjusted R², MAPE, RMSE and understanding them visually
- Machine Learning Method of Finding Unknowns
- Running Univariate v/s Bivariate v/s Multivariate Regression
- Understanding output: p-value, t-value, f-value
- Understanding beta's impact on Y
- Impact of Standardized data on beta's magnitude
- Understanding one hold out cross validation, under fitting, over fitting
- Case Study
 - o Pre Modeling
 - Basic EDA
 - Data Quality Checks
 - o Modeling
 - Data Preparation 1
 - Correcting Column Names
 - Cleaning Data for Type Casting
 - Uni-variate Analysis
 - Dropping of rows based on missing value in Y variable
 - Creating KPI and other derived variables
 - Dropping ID variable
 - Dividing data into numerical, categorical and dependent datasets
 - Performing Missing and Outlier Treatment on Numerical datasets
 - Performing Missing (Mode Value) Imputation on Categorical dataset.



- Dividing categorical dataset into ordinal and nominal datasets.
- Performing label encoding on Ordinal and Dummy Variable Creation on Nominal datasets
- Combining datasets together
- Data Preparation 3
 - Feature Reduction methods
 - o RFE
 - o Stepwise
 - o Lasso
 - o VIF
 - o Cooks Distance
 - o Corr X and Y; Corr X and X
 - Picking important variables
- Data Preparation 4
 - Splitting data into train and test
- Developing model and removing less predictive features
- Running Cook's Distance and removing observations
- Implementing model on train and test
- o Post Modeling
 - Evaluating Model (train v/s test)
 - R²
 - MAPE
 - RMSE/RMSE
 - Creating simple Calculator in MS Excel using final equation



Chapter 5.15 : Logistic Regression

- Understanding different classification problems
- Linear Regression v/s Logistic Regression
- Logistic Regression Equation
- Imbalanced Class Problems and effect on accuracy calculation
- Threshold (Predicted Class) Based Accuracy Metrics
 - o Confusion Matrix
 - o Sensitivity
 - o Specificity
 - o Precision
 - o F1 Score
 - o Accuracy
 - o False Positive Rate
- Non-threshold based Accuracy Metrics
 - o Concordance
 - o AUC (Area Under the Curve) Score
 - o KS Statistic
 - o Decile Analysis
 - o Lift and Gain Chart
- Identification of threshold
 - o Business Scenario
 - o Ready Made rules based on class proportion
 - o ROC Curve
 - o KS Statistic
- Impact of threshold of Sensitivity and Specificity
- Running Logistic Regression on a Dataset
 - o Calculating accuracy metrics using 0.5 threshold value
 - o Calculating accuracy metrics using different threshold values
 - o Applying model on unseen data
 - o Creating simple Calculator in MS Excel using final equation

Chapter 5.16 : Segmentation

- Applications
 - o Customer Segmentation
 - o Market Segmentation
 - o Risk Profiling
 - o Profiling for Advt.
- Types of Segmentation
 - o Heuristic
 - o Scientific
- Type of Heuristic Segmentation
 - o Value Based Segmentation
 - o Life Stage Segmentation
 - o Loyalty Segmentation
 - o RFM Segmentation
- Type of Scientific Segmentation
 - o K-means
 - What is Similarity
 - How this Similarity is found
 - Using Distance metric
 - Using Cosine Similarity
 - What are the things to take care of
 - Standardization of Data
 - Randomness Involved
 - Algorithm: K-means
 - Value of K
 - Centroid
 - Process
 - How to Evaluate the model
 - Profiling: Pros and Cons
 - Size of the Clusters
 - Technical Approach
 - o Silhouette Metric
 - o Pseudo-F value

- Understanding Output
 - Centroid Value
 - Class Labels
- Feature Reduction
 - Correlation between the X vars
 - Principal Component
- DBSCAN
 - Definition
 - Epsilon v/s Min Pts
 - Core v/s Border v/s Noise Point
- Hereditary Clustering
 - AGNES and DIANA
 - Different Links
 - How dendrogram gets created and calculation related to it



Chapter 5.17 : Forecasting

- Definition
- Regression v/s Forecasting
- Definition
 - o Cross Sectional Data
 - o Time Series Data
 - o Panel Data
- Different Time Periods
- Types of Forecasting
 - o Weather, Demand, Macro Economics, Trading
 - o Short, Medium, Long Term
 - o Univariate, Bivariate, Multivariate
- Pre-requisites for performing Forecasting
- Important Concepts
 - o Lag
 - o Lead
 - o Differencing
 - o Stationary of Data
 - o Auto-Correlation
 - o Partial-Auto Correlation
 - o Random Walk /White Noise
 - o Trend
 - o Seasonality
 - o Cyclicity
 - o Irregularity



- The difference b/w Model - Forecasting and Predictive Models
 - o Difference in sampling
 - o Difference in Missing Value Imputation
 - Consider the 1 past or 1 future value
 - Using a predictive model (KNN Model)
 - o Outlier Treatment
 - o Additional methods of evaluating the model
 - AIC
 - BIC
- Basic Techniques
 - o MA, WMA
- Intermediate Statistical Techniques
 - o Auto Regressive
 - o smoothing (Holt Winters / ETS Models)
 - Single Exponential
 - Double Exponential
 - Triple Exponential
 - o Decomposition
- Advanced Statistical Techniques
 - o ARIMA Family
 - AR:Auto Regressive Model
 - MA:Moving Average Model
 - ARMA:Auto Regressive Moving Average Model
 - ARIMA:Auto Regressive Integrated Moving Average Model
 - o SARIMA
 - o ARIMAX, SARIMAX
 - o Finding p,d,q values- manual v/s automated
 - o How to use Regression for solving forecasting models?

Chapter 5.18 : Machine Learning

- Intro
 - o Definition
 - o Hyper-parameters and identification of them
 - o Bias v/s Variance
 - o K-Fold Cross Validation
 - o Concept of optimization & Optimizers (Gradient Descent/Stochastic Gradient Descent/Mini Batch Gradient Descent)
 - o Understanding loss functions
 - o Over and Under Sampling
- K Nearest Neighbor
 - o Characteristics
 - o Intuition behind KNN
 - o KNN for regression/Classification/Imputation of missing's
 - o Pre-requisites and Assumptions
 - o Hyper-parameters and its effect on complexity of model
 - K
 - Distance Metric
 - o Pros and Cons of KNN
- Naïve Bayes
 - o Traditional v/s Bayesian Stats
 - o Bayesian Theorem
 - o Meaning of Naïve Bayes and Assumption
 - o Important Terms
 - Probability (Class Probability)
 - Conditional Probability
 - Evidence
 - Prior, posterior probabilities
 - o Advantages and Disadvantages



Chapter 5.19 : Deep Learning Basics

- Introduction to Deep Learning
- Types of Deep Learning Algorithms
- History and Stages of Deep Learning algorithms
- NN Mimicking human brain cell
- Global Features
- Simple Neural Network and the XOR Problem
- Architecture of ANN
 - o Neuron processing
 - o Hidden Layer
 - o Activation Function & types of activation function
 - o Weights and Bias
 - o Optimization Function
 - o Backward Propagation
 - o Feed Forward Network
 - o Epoch/Iteration
 - o Batch Size
 - o Learning Rate
 - o Dense Network/Fully connected network
 - o Loss Function



Chapter 5.20 : Text Mining

- Introduction to text functions
- REGEX
 - o Meta Characters
 - o Literals
- Feature Reduction
 - o Basic Techniques
 - Regex
 - Remove the numbers
 - Remove symbols from the data
 - Making the text in lower case
 - Remove the commonly occurring words
 - o Advanced Techniques (NLP based techniques)
 - POS Tagging
 - Stemming
 - Lemmatization
 - Stop Words
 - Business Specific Words
- Conversion of Text to Numerical Data
 - o Count Vectorizer
 - o TF-IDF
 - o Word2Vec
- Web Scrapping
- Sentiment Analysis
- Text Classification - Sentiment Analysis – Intent Analysis
 - o Naïve Bayes
 - o SVM
- Text Segmentation - Topic Mining
 - o K-means
 - o LDA



- o Types
 - Bernoulli NB
 - Multinomial NB
 - Gaussian NB
- Decision Trees
 - o Rule / Tree Based Algorithms
 - o Features
 - o Pros and Cons
 - o Assumptions
 - o Splitting Criteria
 - o Regression v/s Classification
 - o Types of decision trees
 - CHAID/CART/C5.0
 - o Hyper-parameters and their effects
 - Splitting criteria of tree
 - MSE
 - F-value
 - Gini
 - Entropy
 - Information Gain
 - Stopping criteria of tree
 - Depth of the Tree
 - Number of Nodes
 - Minimum number of obs in the node
 - Number of features
 - Min Samples in Leaf
- Ensemble Learning
 - o Homogeneous v/s Heterogeneous models
 - o Bagging
 - Process
 - Bootstrap Aggregating
 - Hyper Parameters and their effect on complexity of the model



- o Random Forest
 - Bootstrapping + Random Subspaces
 - Hyper Parameters and their effect on complexity of the model
- o Boosting
 - Ada-Boost
 - Gradient Boosting
 - XGBoosting
- o Sequential v/s Parallel Process
- Support Vector Machine
 - o History of algorithm and its development
 - o concept of margin
 - o Optimization + Constraint problem
 - o Loss Function Explanation
 - o Margin v/s Soft Margin
 - o Kernel Trick / Transformation
 - o Hyper parameters and its effect on complexity of the model
 - constraint
 - gamma
 - o Pros and Cons of SVM
 - o Classifier v/s Regressor SVM
- Regularization
 - o Ridge
 - o Lasso
 - o Elastic Net
 - o Regularization for modeling, feature reduction



Python Bronze Certification

Certified Data Analyst

Become a Certified data analyst by learning to import, manipulate & visualize data using the programming language Python.



No. of Hours

50 Hours



**Assignment &
Case Study**

6

Overview :

Python Bronze Certification - Certified Data Analyst :

Using Python as a tool for data science, this course introduces the language to beginners. When you have completed the lab exercises, you will be ready to write your first Python script on your own! The course objectives are to give you an introduction to Python as a language for programming, and to give you a taste of how to work with data in Python. The course will prepare you to apply these skills to business problems and to effectively communicate these understandings to internal and external stakeholders.



Chapter 6.1 : Introduction to Science and Python

- History and Evolution of Data Science
- Modern day role of Data Science in Companies
- Components of Data Science
- Role of Python in implementation of Data Science
- Python Notebooks and IDEs
- Understanding Jupyter Notebook

Chapter 6.2 : Data Types and Structures of Core Python

- Fundamental Syntax Rules
- Basic Data Types found in Core Python v/s Pandas
- Type Casting methods and best practices
- Understanding the concept of class
- Exploring associates attributes
- Handling Data Structures found in Core Python

Chapter 6.3 : Role of Modular Programming in Python

- Basic of Programming: Loops and Conditions
- Creating User Defined Functions
- Understanding Modules and Packages
- Installing and Loading Python Packages
- Creating and Using own modules and packages
- Dealing with directories and anaconda environments using yml file

Chapter 6.4 : Vectorized Data Structures: Numpy Array and Pandas Series

- numpy's functions
- Using Numpy's Array
- Using Pandas Series (indexing, functions etc.)
- Comparing Array and Series with traditional python data structure (Difference between non-vectorized and vectorized structure)

Chapter 6.5 : Data Mining (Basic)

- Data Importing and Exporting
- Basic Data Exploration

Chapter 6.6 : Data Mining (Intermediate)

- Extraction of Columns
- Adding of Columns
- Renaming of Columns
- Dropping of Columns
- Rearranging of Columns
- Typecasting of Columns
- Filtering and Sorting of Rows

Chapter 6.7 : Data Mining (Advanced)

- Aggregation
- Appending
- Merging
- Transposing

Chapter 6.8 : Data Quality check

- Missing Value Treatment
- Outlier Capping

Chapter 6.9 : Data Visualization

- Histogram
- Boxplot
- Line Chart
- Scatterplot
- Bar Chart
- Stacked Bar Chart
- Dodged Bar Chart
- Pie Chart
- Cross Tab



Chapter 6.10 : Descriptive Statistics

- Understanding difference in Samples and Population
- Understanding difference in Statistic and Parameter
- Understanding the need of creating samples
- Exploring methods of creating samples
- Understanding difference in Descriptive and Inferential Statistics
- Measure of Frequency
- Measure of Central Tendency
- Measure of Variability
- Measure of Shape

Chapter 6.11 : Understanding Probability Distribution

- Central Limit Theorem
- 3 Sigma Rule
- Finding Probability for a raw score
- Standard Normal Distribution

Chapter 6.12 : Hypothesis Testing

- Z-test
- T-test
- F-test
- Chi-Squared test
- Pearson's Correlation test

Chapter 6.13 : Finding Business Insights using Statistics

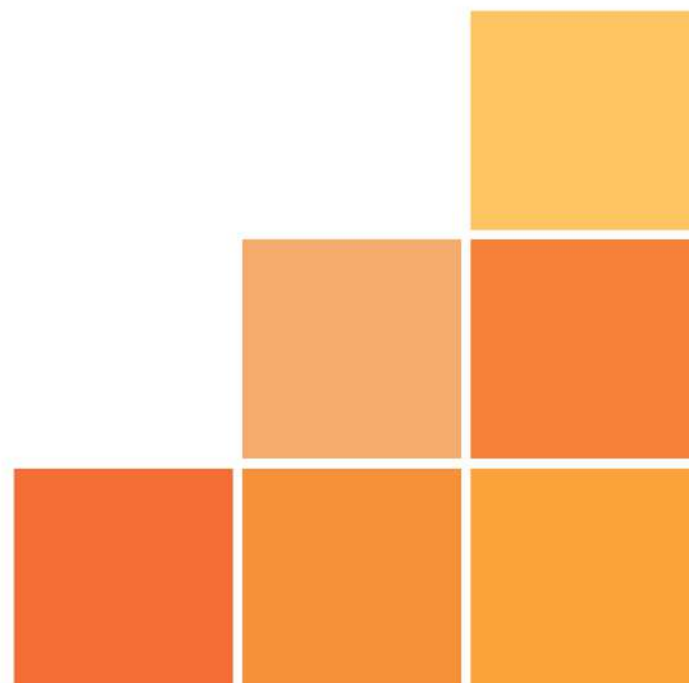
- Methods to Identify correct hypothesis test
- Interpreting statistical outputs into business solutions
- Ways to maximise insights gained from hypothesis tests
- 6 sigma rule
- Point Estimate v/s Confidence Interval

Chapter 6.14 : Combining Data Mining and Statistics

- Case Study to explore the approaches used for understanding, manipulating, mining and preparing data

Chapter 6.15 : Introduction to Predictive Modeling

- Converting Business Problems into Data Science based problems
- Evolution of Predictive Modeling Techniques
- Types of Predictive Models
- Types of Learning Algorithms



Python Silver Certification Certified Business Intelligence Specialist

Venture into the field Data Science with Machine Learning & learn to create predictive regression & classification models using Python.



No. of Hours

50 Hours



**Assignment &
Case Study**

6

Overview :

Python Silver Certification - Certified Business Intelligence Specialist :

We have designed this gold level training course for Data Scientists familiar with Python and interested in creating predictive models that apply Machine Learning to the data in their organization.

Throughout this course you will learn best practices, models, code, algorithms, and a framework for improving your projects. In this course you will learn how to combine state-of-the-art Machine Learning algorithms with Python tools and functionalities to take your Data Science projects to the next level after the completion of the course.

Chapter 7.1 : Encoding and Binning

- Curse of Dimensionality
- Problem of Multicollinearity
- Types of features
- Binning v/s Encoding
- Label Encoding v/s One-hot Encoding

Chapter 7.2 : Basics of Regression

- Understanding Regression Problems
- Basics of Linear Regression (theory)
- Creating preliminary model using Linear Regression

Chapter 7.3 : Regression Model Building : Pre-Modeling

- Data Quality Checks
- Data Preparation

Chapter 7.4 : Regression Model Building : Modeling

- Assumption Checks
- Feature Reduction
- Data split in development and validation sets

Chapter 7.5 : Regression Model Building - Post Modeling

- Model Evaluation
- Model Interpretation

Chapter 7.6 : Basics of Classification using Logistic Regression

- Understanding Classification Problems
- Basics of Logistic Regression (theory)
- Creating preliminary model using Logistic Regression

Chapter 7.7 : Classification Model Building : Pre-Modeling & Modeling

- Data Quality Checks
- Data Preparation
- Assumption Checks
- Feature Reduction
- Data split in development and validation sets

Chapter 7.8 : Classification Model Building : Post Modeling

- Model Evaluation
- Model Interpretation

Chapter 7.9 : Introduction to Machine Learning

- Basics of Machine Learning
- Overfitting - Causes and Solutions
- Resampling
- Regularization

Chapter 7.10 : Decision Trees

- Decision Trees Regressor and Classifier Theory
- Implementing Decision Trees in Python

Chapter 7.11 : K Nearest Neighbour

- KNN Regressor and Classifier Theory
- Implementing KNN in Python



Python Gold Certification Certified Data Scientist

Complete your journey in Data science by resolving advanced business problems related to forecasting, clustering, natural language processing and implementation using Python.



No. of Hours

50 Hours



**Assignment &
Case Study**

6

Overview :

Python Gold Certified - Certified Data Scientist :

We have designed this gold level training course for Data Scientists familiar with Python and interested in creating predictive models that apply Machine Learning to the data in their organization.

Throughout this course you will learn best practices, models, code, algorithms, and a framework for improving your projects. In this course you will learn how to combine state-of-the-art Machine Learning algorithms with Python tools and functionalities to take your Data Science projects to the next level after the completion of the course..

Chapter 8.1 : Introduction to Unsupervised Models

- Understanding Segmentation
- Difference between Supervised and Unsupervised Learning Setup
- Difference between Heuristic and Scientific Segmentation techniques

Chapter 8.2 : Heuristic Segmentation

- Value Based Segmentation
- Loyalty Segmentation
- Life Stage Segmentation
- RFM Segmentation
- Application of Heuristic Segmentation in Python

Chapter 8.3 : Scientific Segmentation : Theory

- K-means
- DBSCAN
- Hierarchical Clustering

Chapter 8.4 : Scientific Segmentation : Application

- Applying Clustering Algorithms in Python
- Exploring methods of evaluation

Chapter 8.5 : Introduction to Forecasting

- Data Preparation for in Forecasting
- Types of Forecasting
- Time Series Components

Chapter 8.6 : Forecasting Algorithms

- Using Averages to Forecast
- Auto Regressive Models
- ETS Models
- Decomposition
- ARIMA/SARIMA
- ARIMAX/SARIMAX

Chapter 8.7 : Solving Univariate and Multivariate Forecasting Problems in Python

- Using forecasting algorithms to solve Univariate Forecasting Problems in Python
- Using forecasting algorithms to solve Bivariate Forecasting Problems in Python

Chapter 8.8 : Introduction to Text Mining

- Understanding the problems with text based data
- Converting unstructured text data into structured data
- Using sklearn to create basic Text Classification model

Chapter 8.9 : Exploring methods of Data Acquisition and Cleaning

- Web Scraping
- Learning Regex

Chapter 8.10 : Feature Reduction in Text Mining using NLP

- POS Tagging
- Stemming
- Lemmatization
- StopWords



Chapter 8.11 : Creating a Text Classification Model

- Performing Sentiment Analysis in Python

Chapter 8.12 : Creating a Text based Segmentation Model

- Performing Topic Mining using K-means
- Performing Topic Mining using LDA

Chapter 8.13 : Interactive Dashboard in Python

- Learning Basics of Streamlit
- Integration Visualization techniques with Streamlit

Chapter 8.14 : Automating Report Generation using Streamlit

- Learning Basics of Streamlit
- Integration Visualization techniques with Streamlit

Chapter 8.15 : Deploying a Prediction Model using Streamlit

- Creating a Loan Approval web-app



Python Platinum Certification Certified AI Engineer

Advance your journey with Data Science by learning about deep learning algorithms and induce artificial intelligence in your decision making



No. of Hours

50 Hours



**Assignment &
Case Study**

6

Overview :

Python Platinum Certification - Certified AI Engineer :

This course will help you to Master python, and learn how to build any python model using Artificial Intelligence. The student should already be familiar with Python basics, statistical modeling and machine learning. The primary libraries used in this module are Keras and TensorFlow.



Chapter 9.1 : Introduction to Artificial Intelligence

- Evolution of Neural Networks
- Parallels between Neurons And Artificial Neuron
- Single Layer Perceptron: the XOR problem
- Global Features
- Understanding Deep Learning Algorithms

Chapter 9.2 : Artificial Neural Networks - I

- Weights
- Bias
- Activation Functions
- Hidden Layers
- Number of Neurons
- Backpropagation

Chapter 9.3 : Artificial Neural Networks - II

- Understanding Keras and Tensorflow
- Applying ANN in Python for solving Regression Problem
- Applying ANN in Python for solving Classification Problem

Chapter 9.4 : Convolutional Neural Networks - I

- Architecture of CNN
- Layers used in CNN
- Bias
- Pooling and Flattening
- Full Connection

Chapter 9.5 : Convolutional Neural Networks - II

- Implementing CNN in python for Image Classification

Chapter 9.6 : Recurrent Neural Networks - I

- Architecture of RNN
- Layers used in CNN
- Types of LSTM

Chapter 9.7 : Recurrent Neural Networks - II

- Implementing RNN in python for Regression

Chapter 9.8 : Recurrent Neural Networks - III

- Implementing RNN in python for Classification

Chapter 9.9 : Self Organising Maps - I

- Introduction to use of Neural Networks in Unsupervised Problems
- Architecture of SOM
- Similarities to other ML algorithms

Chapter 9.10 : Self Organising Maps - II

- Implementing SOM in python for Customer Segmentation

Chapter 9.11 : Boltzmann Machine - I

- Introduction to Energy Based Models
- Architecture of Boltzmann Machine
- Types of Boltzmann Machine

Chapter 9.12 : Boltzmann Machine - II

- Creating Boltzmann Machine Algorithm using Tensorflow

Chapter 9.13 : Boltzmann Machine - III

- Applying Boltzmann Machine Algorithm to solve Predicative Problem

Chapter 9.14 : Auto Encoders - I

- Architecture of Auto Encoders
- Types of Auto Encoders

Chapter 9.15 : Auto Encoders - II

- Applying Auto Encoders in Python to solve unsupervised learning problem

Big Data Diamond Certification

Certified Big Data Architect

Due to Volume, variety, veracity & velocity in which data is being produced, there is requirement of certain tools that helps data scientist to fetch data from various sources and manipulate them to start with model building.



No. of Hours

60 Hours



**Assignment &
Case Study**

10

Overview :

Big Data Diamond Certification – Certified Big Data Architect :

Due to the Volume, Variety, Veracity and Velocity at which data is being produced, there is a requirement of certain tools that help Data Scientist to fetch data from various sources and manipulate them to start with model building.



Chapter 10.1 : Starting with Big Data and HDFS

- Introduction to Big Data
- Understanding the HDFS Architecture
- Installing Hadoop

Chapter 10.2 : Working with MapReduce

- Introduction to Map Reduce
- Jar Creation and Execution
- MapReduce Program Variations

Chapter 10.3 : Using Apache Sqoop

- Understanding Sqoop Fundamentals
- Sqoop Export and Incremental
- Sqoop Job

Chapter 10.4 : Apache Hive – Basic

- Introduction to Hive
- Applying Hive
- Subqueries and Index
- User Defined Functions
- Complex Datatypes

Chapter 10.5 : Apache Hive – Advanced

- Optimization
- Partioning
- Bucketing
- Window Functions
- Ranking & sorting
- Row & Column File Formats
- Specialized File Formats
- ORC and Parquet
- Compression Techniques
- Vectorization
- Altering Hive Engine
- Thrift Server
- MSCK Repair

Chapter 10.6 : Hbase

- Intro to Hbase
- Understanding Hbase Architecture
- CAP Theorem
- Understanding Cassandra
- Integration with Hive and Hadoop

Chapter 10.7 : Scala

- Intro to Scala
- Learning Functional Programming
- Object Oriented Operations

Big Data Sapphire Certificate Certified Big Data Engineer

The latest advancements in the fields of big data, one can combine their knowledge in Python with big data tools. Also, knowledge Cloud computing and work flow management tools is must with Big Data.



No. of Hours

60 Hours



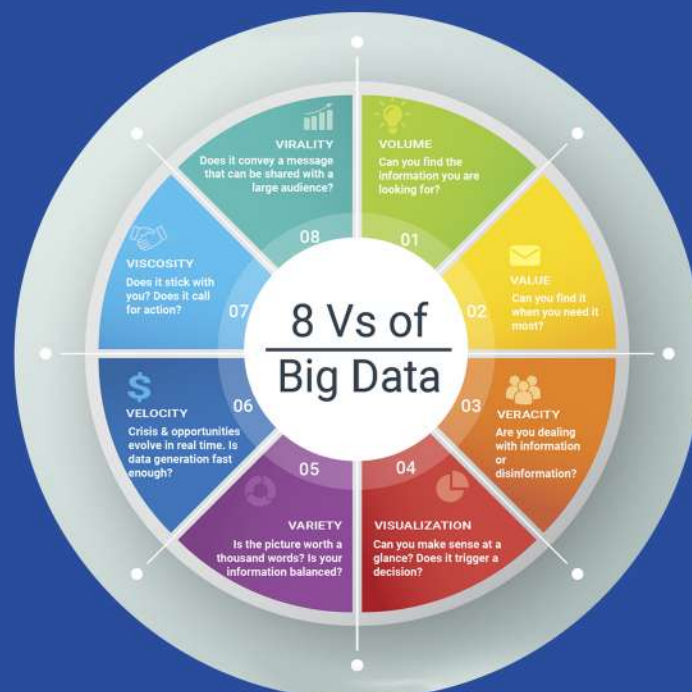
**Assignment &
Case Study**

10

Overview :

Big Data Sapphire Certification - Certified Big Data Engineer :

The latest advancements in the fields of big data, one can combine their knowledge in python with big data tools. Also, knowledge Cloud Computing and Work Flow management tools is a must with Big Data.





Chapter 11.1 : Spark – Basics

- Intro to Spark Basics
- Learning Spark Theory
- Applying Fundamentals of Spark

Chapter 11.2 : Spark – Advanced

- Broadcast Variables
- Accumulator
- Intro to YARN
- Advanced Application of Spark

Chapter 11.3 : Spark – DataFrame

- Learning advanced Spark theory
- Basic Operations of Spark with DataFrame
- Advanced Operations of Spark with DataFrame

Chapter 11.4 : Spark – Optimization

- Basics of Spark Optimization and Performance Tuning
- Broadcast Join with Structured API – RDD, DataFrames
- Client and Cluster Mode
- Sort and Hash Aggregate
- Catalyst Optimizer
- Connection to External Source

Chapter 11.5 : Spark – Streaming

- Introduction to Real Time Processing
- Dstream
- Stateless and Stateful Transformations
- Sliding Window
- Structured Streaming
- Advanced Application of Spark Streaming

Chapter 11.6 : Apache Kafka

- Introduction to Kafka
- Cluster Nodes, Brokers
- One Node vs Multi Node
- Scaling up Kafka Cluster

Chapter 11.7 : Cloud Computing with AWS – Athena

- Introduction to Cloud Computing
- Running Spark Code on AWS EMR
- Fundamentals of AWS
- Exploring Storage, Networking & CLI
- AWS EMR: Launch a EMR Cluster Using Advanced Options
- Introduction to AWS Athena
- AWS Athena Session
- AWS Athena with Glue

Chapter 11.8 : Cloud Computing with AWS – RedShift

- Understanding difference between Database, Datawarehouse and Data lake
- Introduction to Redshift
- Understanding Job Creation and Other Features
- Advantages of Redshift
- Redshift Spectrum
- SQL Workbench
- AWS Glue

Chapter 11.9 : AirFlow

- Introduction to Airflow
- Understanding Architecture
- Cran Job
- Creating Dags
- Tasks and Operators