Table of Contents for PG in Data Scientist

[COURSE 1 : Data Science with R 3](#_Toc45548919)

[Lesson 0 : Course Introduction 3](#_Toc45548920)

[Lesson 1 : Introduction to Business Analytics 3](#_Toc45548921)

[Lesson 2 : Introduction to R Programming 3](#_Toc45548922)

[Lesson 3 : Data Structures 3](#_Toc45548923)

[Lesson 4 : Data Visualization 3](#_Toc45548924)

[Lesson 5 : Statistics for Data Science I 3](#_Toc45548925)

[Lesson 6 : Statistics for Data Science II 4](#_Toc45548926)

[Lesson 7 : Regression Analysis 4](#_Toc45548927)

[Lesson 8 : Classification 4](#_Toc45548928)

[Lesson 9 : Clustering 4](#_Toc45548929)

[Lesson 10 : Association 5](#_Toc45548930)

[COURSE 2 : Data Science with SAS 6](#_Toc45548931)

[Lesson 0 : Course Introduction 6](#_Toc45548932)

[Lesson 1 : Analytics Overview 6](#_Toc45548933)

[Lesson 2 : Introduction to SAS 6](#_Toc45548934)

[Lesson 3 : Combining and Modifying Datasets 6](#_Toc45548935)

[Lesson 4 : PROC SQL 6](#_Toc45548936)

[Lesson 5 : SAS Macros 7](#_Toc45548937)

[Lesson 6 : Basics of Statistics 7](#_Toc45548938)

[Lesson 7 : Statistical Procedures 7](#_Toc45548939)

[Lesson 8 : Data Exploration 8](#_Toc45548940)

[Lesson 10 : Working with Time Series Data 8](#_Toc45548941)

[Lesson 11 : Designing Optimization Models 9](#_Toc45548942)

[COURSE 3 : Data Science with Python 10](#_Toc45548943)

[Lesson 0 : Course Overview 10](#_Toc45548944)

[Lesson 1 : Data Science Overview 10](#_Toc45548945)

[Lesson 2 : Data Analytics Overview 10](#_Toc45548946)

[Lesson 3 : Statistical Analysis and Business Applications 10](#_Toc45548947)

[Lesson 4 : Python Environment Setup and Essentials 10](#_Toc45548948)

[Lesson 5 : Mathematical Computing with Python (NumPy) 11](#_Toc45548949)

[Lesson 6 : Scientific computing with Python (Scipy) 11](#_Toc45548950)

[Lesson 7 : Data Manipulation with Pandas 11](#_Toc45548951)

[Lesson 8 : Machine Learning with Scikit–Learn 11](#_Toc45548952)

[Lesson 9 : Natural Language Processing with Scikit Learn 12](#_Toc45548953)

[Lesson 10 : Data Visualization in Python using matplotlib 12](#_Toc45548954)

[Lesson 11 : Web Scraping with BeautifulSoup 12](#_Toc45548955)

[Lesson 12 : Python integration with Hadoop MapReduce and Spark 13](#_Toc45548956)

[Lesson 13 : Practice Projects 13](#_Toc45548957)

[COURSE 4 : Machine Learning 14](#_Toc45548958)

[Lesson 1 : Course Introduction 14](#_Toc45548959)

[Lesson 2 : Introduction to AI and Machine Learning 14](#_Toc45548960)

[Lesson 3 : Data Preprocessing 14](#_Toc45548961)

[Lesson 4 : Supervised Learning 14](#_Toc45548962)

[Lesson 5 : Feature Engineering 15](#_Toc45548963)

[Lesson 6 : Supervised Learning Classification 15](#_Toc45548964)

[Lesson 7 : Unsupervised Learning 16](#_Toc45548965)

[Lesson 8 : Time Series Modeling 16](#_Toc45548966)

[Lesson 9 : Ensemble Learning 17](#_Toc45548967)

[Lesson 10 : Recommender Systems 17](#_Toc45548968)

[Lesson 11 : Text Mining 17](#_Toc45548969)

[Lesson 12 : Project Highlights 18](#_Toc45548970)

[Lesson 13 : Practice Projects 18](#_Toc45548971)

[COURSE 5 : Tableau 10 19](#_Toc45548972)

[Lesson 1 : Getting Started With Tableau 19](#_Toc45548973)

[Lesson 2 : Working With Tableau 19](#_Toc45548974)

[Lesson 3 : Deep diving with Data and Connections 19](#_Toc45548975)

[Lesson 4 : Creating Charts 19](#_Toc45548976)

[Lesson 5 : Adding calculations to your workbook 19](#_Toc45548977)

[Lesson 6 : Mapping data in Tableau 19](#_Toc45548978)

[Lesson 7 : Dashboards and Stories 19](#_Toc45548979)

[Lesson 8 : Visualizations For An Audience 20](#_Toc45548980)

[Lesson 9 :Practice Projects 20](#_Toc45548981)

[COURSE 6 : Big Data Hadoop and Spark Developer 21](#_Toc45548982)

[Lesson 1 : Course Introduction 21](#_Toc45548983)

[Lesson 2 : Introduction to Big Data and Hadoop 21](#_Toc45548984)

[Lesson 3 : Hadoop Architecture Distributed Storage (HDFS) and YARN 21](#_Toc45548985)

[Lesson 4 : Data Ingestion into Big Data Systems and ETL 22](#_Toc45548986)

[Lesson 5 : Distributed Processing MapReduce Framework and Pig 22](#_Toc45548987)

[Lesson 6 : Apache Hive 23](#_Toc45548988)

[Lesson 7 : NoSQL Databases HBase 23](#_Toc45548989)

[Lesson 8 : Basics of Functional Programming and Scala 23](#_Toc45548990)

[Lesson 9 : Apache Spark Next Generation Big Data Framework 24](#_Toc45548991)

[Lesson 10 : Spark Core Processing RDD 24](#_Toc45548992)

[Lesson 11 : Spark SQL Processing DataFrames 25](#_Toc45548993)

[Lesson 12 : Spark MLLib Modelling BigData with Spark 25](#_Toc45548994)

[Lesson 13 : Stream Processing Frameworks and Spark Streaming 25](#_Toc45548995)

[Lesson 14 : Spark GraphX 26](#_Toc45548996)

[Lesson 15 : Practice Projects 26](#_Toc45548997)

[COURSE 7 : Data Science Capstone 27](#_Toc45548998)

[Lesson 0 : Day Problem and approach overview 27](#_Toc45548999)

[Lesson 1 : Day Data pre processing techniques application on data set 27](#_Toc45549000)

[Lesson 2 : Day Model Building and fine tuning leveraging various techniques 27](#_Toc45549001)

[Lesson 3 : Day Dashboard problem statement to meet the business objective 27](#_Toc45549002)

[Lesson 4 : Day Final evaluation 27](#_Toc45549003)

# COURSE 1 : Data Science with R

## Lesson 0 : Course Introduction

0.1 Course Introduction

## Lesson 1 : Introduction to Business Analytics

1.1 Overview

1.2 Business Decisions and Analytics

1.3 Types of Business Analytics

1.4 Applications of Business Analytics

1.5 Data Science Overview

1.6 Conclusion

1.7 Knowledge Check

## Lesson 2 : Introduction to R Programming

## 

2.1 Overview

2.2 Importance of R

2.3 Data Types and Variables in R

2.4 Operators in R

2.5 Conditional Statements in R

2.6 Loops in R

2.7 R Script

2.8 Functions in R

2.9 Conclusion

2.10 Knowledge Check

## Lesson 3 : Data Structures

3.1 Overview

3.2 Identifying Data Structures

3.3 Demo Identifying Data Structures

3.4 Assigning Values to Data Structures

3.5 Data Manipulation

3.6 Demo Assigning Values and Applying Functions

3.7 Conclusion

3.8 Knowledge Check

## Lesson 4 : Data Visualization

4.1 Overview

4.2 Introduction to Data Visualization

4.3 Data Visualization Using Graphics in R

4.4 Ggplot2

4.5 File Formats of Graphic Outputs

4.6 Conclusion

4.7 Knowledge Check

## Lesson 5 : Statistics for Data Science I

5.1 Overview

5.2 Introduction to Hypothesis

5.3 Types of Hypothesis

5.4 Data Sampling

5.5 Confidence and Significance Levels

5.6 Conclusion

5.7 Knowledge Check

## Lesson 6 : Statistics for Data Science II

6.1 Overview

6.2 Hypothesis Test

6.3 Parametric Test

6.4 Non-parametric Test

6.5 Hypothesis Tests About Population Means

6.6 Hypothesis Tests About Population Variance

6.7 Hypothesis Tests About Population Proportions

6.8 Conclusion

6.9 Knowledge Check

## Lesson 7 : Regression Analysis

7.1 Overview

7.2 Introduction to Regression Analysis

7.3 Types of Regression Analysis Models

7.4 Linear Regression

7.5 Demo Simple Linear Regression

7.6 Non-linear Regression

7.7 Demo Regression Analysis With Multiple Variables

7.8 Cross Validation

7.9 Non-linear to Linear Models

7.10 Principal Component Analysis

7.11 Factor Analysis

7.12 Conclusion

7.13 Knowledge Check

## Lesson 8 : Classification

8.1 Overview

8.2 Classification and Its Types

8.3 Logistic Regression

8.4 Support Vector Machines

8.5 Demo Support Vector Machines

8.6 K-nearest Neighbours

8.7 Naive Bayes Classifier

8.8 Demo Naive Bayes Classifier

8.9 Decision Tree Classification

8.10 Demo Decision Tree Classification

8.11 Random Forest Classification

8.12 Evaluating Classifier Models

8.13 Demo K-fold Cross Validation

8.14 Conclusion

8.15 Knowledge Check

## Lesson 9 : Clustering

9.1 Overview

9.2 Introduction to Clustering

9.3 Clustering Methods

9.4 Demo K-means Clustering

9.5 Demo Hierarchical Clustering

9.6 Conclusion

9.7 Knowledge Check

## Lesson 10 : Association

10.1 Overview

10.2 Association Rule

10.3 Apriori Algorithm

10.4 Demo Apriori Algorithm

10.5 Conclusion

10.6 Knowledge Check

# COURSE 2 : Data Science with SAS

## Lesson 0 : Course Introduction

* 1. Introduction

## Lesson 1 : Analytics Overview

1.1 Introduction

1.2 Introduction to Business Analytics

1.3 Types of Analytics

1.4 Areas of Analytics

1.5 Analytical Tools

1.6 Analytical Techniques

1.7 Quiz

1.8 Key Takeaways

## Lesson 2 : Introduction to SAS

2.1 Introduction

2.2 What is Sas

2.3 Navigating in the Sas Console

2.4 Sas Language Input Files

2.5 Data Step

2.6 Proc Step and Data Step - Example

2.7 Data Step Processing

2.8 Sas Libraries

2.9 Demo - Importing Data

2.10 Demo - Exporting Data

2.11 Knowledge Check

2.12 Quiz

2.13 Key Takeaways

## Lesson 3 : Combining and Modifying Datasets

3.1 Introduction

3.2 Why Combine or Modify Data

3.3 Concatenating Datasets

3.4 Interleaving Method

3.5 One - to - One Reading

3.6 One - to - One Merging

3.7 Knowledge Check

3.8 Data Manipulation

3.9 Modifying Variable Attributes

3.10 Assignment

3.11 Assignment Solution

3.12 Assignment

3.13 Assignment Solution

3.14 Quiz

3.15 Key Takeaways

## Lesson 4 : PROC SQL

4.1 Introduction

4.2 What is Proc Sql

4.3 Retrieving Data from a Table

4.4 Demo - Retrieve Data from a Table

4.5 Selecting Columns in a Table

4.6 Knowledge Check

4.7 Retrieving Data from Multiple Tables

4.8 Selecting Data from Multiple Tables

4.9 Concatenating Query Results

4.10 Assignment

4.11 Assignment Solution

4.12 Assignment

4.13 Assignment Solution

4.14 Quiz

4.15 Key Takeaways

## Lesson 5 : SAS Macros

5.1 Introduction

5.2 Need for Sas Macros

5.3 Macro Functions

5.4 Macro Functions Examples

5.5 Sql Clauses for Macros

5.6 Knowledge Check

5.7 The Macro Statement

5.8 The Conditional Statement

5.9 Assignment

5.10 Assignment Solution

5.11 Quiz

5.12 Key Takeaways

## Lesson 6 : Basics of Statistics

6.1 Introduction

6.2 Introduction to Statistics

6.3 Statistical Terms

6.4 Procedures in Sas for Descriptive Statistics

6.5 Demo - Descriptive Statistics

6.6 Hypothesis Testing

6.7 Variable Types

6.8 Hypothesis Testing - Process

6.9 Knowledge Check

6.10 Demo - Hypothesis Testing

6.11 Parametric and Non - Parametric Tests

6.12 Parametric Tests

6.13 Non - Parametric Tests

6.14 Parametric Tests - Advantages and Disadvantages

6.15 Quiz

6.16 Key Takeaways

## Lesson 7 : Statistical Procedures

7.1 Introduction

7.2 Statistical Procedures

7.3 Proc Means

7.4 Proc Means - Examples

7.5 Proc Freq

7.6 Demo - Proc Freq

7.7 Proc Univariate

7.8 Demo - Proc Univariate

7.9 Knowledge Check

7.10 Proc Corr

7.11 Proc Corr Options

7.12 Demo - Proc Corr

7.13 Proc Reg

7.14 Proc Reg Options

7.15 Demo - Proc Reg

7.16 Proc Anova

7.17 Demo - Proc Anova

7.18 Assignment

7.19 Assignment Solution

7.20 Assignment

7.21 Assignment Solution

7.22 Quiz

7.23 Key Takeaways

## Lesson 8 : Data Exploration

8.1 Introduction

8.2 Data Preparation

8.3 General Comments and Observations on Data Cleaning

8.4 Knowledge Check

8.5 Data Type Conversion

8.6 Character Functions

8.7 Scan Function

8.8 Datetime Functions

8.9 Missing Value Treatment

8.10 Various Functions to Handle Missing Value

8.11 Data Summarization

8.12 Assignment

8.13 Assignment Solution

8.14 Quiz

8.15 Key Takeaways

Lesson 9 : Advanced Statistics

9.1 Introduction

9.2 Introduction to Cluster

9.3 Clustering Methodologies

9.4 Demo - Clustering Method

9.5 K Means Clustering

9.6 Knowledge Check

9.7 Decision Tree

9.8 Regression

9.9 Logistic Regression

9.10 Assignment

9.11 Assignment Solution

9.12 Assignment

9.13 Assignment Solution

9.14 Quiz

9.15 Key Takeaways

## Lesson 10 : Working with Time Series Data

10.1 Introduction

10.2 Need for Time Series Analysis

10.3 Time Series Analysis - Options

10.4 Reading Date and Datetime Values

10.5 White Noise Process

10.6 Stationarity of a Time Series

10.7 Knowledge Check

10.8 Demo — Stages of Arima Modelling

10.9 Plot Transform Transpose and Interpolating Time Series Data

10.10 Assignment Solution

10.11 Quiz

10.12 Key Takeaways

10.13 Assignment

## Lesson 11 : Designing Optimization Models

11.1 Introduction

11.2 Need for Optimization

11.3 Optimization Problems

11.4 Proc Optmodel

11.5 Optimization - Example

11.6 Optimization - Example

11.7 Assignment

11.8 Assignment Solution

11.9 Quiz

11.10 Key Takeaways

# COURSE 3 : Data Science with Python

## Lesson 0 : Course Overview

0.1 Course Overview

## Lesson 1 : Data Science Overview

1.1 Introduction to Data Science

1.2 Different Sectors Using Data Science

1.3 Purpose and Components of Python

1.4 Quiz

1.5 Key Takeaways

## Lesson 2 : Data Analytics Overview

2.1 Data Analytics Process

2.2 Knowledge Check

2.3 Exploratory Data Analysis(eda)

2.4 Eda-quantitative Technique

2.5 Eda - Graphical Technique

2.6 Data Analytics Conclusion or Predictions

2.7 Data Analytics Communication

2.8 Data Types for Plotting

2.9 Data Types and Plotting

2.10 Quiz

2.11 Key Takeaways

2.12 Knowledge Check

## Lesson 3 : Statistical Analysis and Business Applications

3.1 Introduction to Statistics

3.2 Statistical and Non-statistical Analysis

3.3 Major Categories of Statistics

3.4 Statistical Analysis Considerations

3.5 Population and Sample

3.6 Statistical Analysis Process

3.7 Data Distribution

3.8 Dispersion

3.9 Knowledge Check

3.10 Histogram

3.11 Knowledge Check

3.12 Testing

3.13 Knowledge Check

3.14 Correlation and Inferential Statistics

3.15 Quiz

3.16 Key Takeaways

## Lesson 4 : Python Environment Setup and Essentials

4.1 Anaconda

4.2 Installation of Anaconda Python Distribution (contd )

4.3 Data Types With Python

4.4 Basic Operators and Functions

4.5 Quiz

4.6 Key Takeaways

## Lesson 5 : Mathematical Computing with Python (NumPy)

5.1 Introduction to Numpy

5.2 Activity-sequence It Right

5.3 Demo -creating and Printing an Ndarray

5.4 Knowledge Check

5.5 Class and Attributes of Ndarray

5.6 Basic Operations

5.7 Activity-slice It

5.8 Copy and Views

5.9 Mathematical Functions of Numpy

5.10 Practice Project: Analyse Gdp of Countries

5.11 Assignment Demo

5.12 Practice Project: Analyse London Olympics Dataset

5.13 Assignment Demo

5.14 Quiz

5.15 Key Takeaways

## Lesson 6 : Scientific computing with Python (Scipy)

6.1 Introduction to Scipy

6.2 Scipy Sub Package - Integration and Optimization

6.3 Knowledge Check

6.4 Scipy Sub Package

6.5 Demo - Calculate Eigenvalues and Eigenvector

6.6 Knowledge Check

6.7 Scipy Sub Package - Statistics, Weave and Io

6.8 Practice Project: Solving Linear Algebra Problem Using Scipy

6.9 Assignment Demo

6.10 Practice Project: Perform Cdf and Pdf Using Scipy

6.11 Assignment Demo

6.12 Quiz

6.13 Key Takeaways

## Lesson 7 : Data Manipulation with Pandas

7.1 Introduction to Pandas

7.2 Knowledge Check

7.3 Understanding Dataframe

7.4 View and Select Data Demo

7.5 Missing Values

7.6 Data Operations

7.7 Knowledge Check

7.8 File Read and Write Support

7.9 Knowledge Check-sequence It Right

7.10 Pandas Sql Operation

7.11 Practice Project: Analyse the Federal Aviation Authority Dataset Using Pandas

7.12 Assignment Demo

7.13 Practice Project: Analyse Newyork City Fire Department Dataset

7.14 Assignment Demo

7.15 Quiz

7.16 Key Takeaways

## Lesson 8 : Machine Learning with Scikit–Learn

8.1 Machine Learning Approach

8.2 Steps One and Two

8.3 Steps Three and Four

8.4 How It Works

8.5 Steps Five and Six

8.6 Supervised Learning Model Considerations

8.7 Knowledge Check

8.8 Scikitlearn

8.9 Knowledge Check

8.10 Supervised Learning Models - Linear Regression

8.11 Supervised Learning Models - Logistic Regression

8.12 Unsupervised Learning Models

8.13 Pipeline

8.14 Model Persistence and Evaluation

8.15 Knowledge Check

8.16 Practice Project: Analysing Ad Budgets for Different Media Channels

8.17 Assignment One

8.18 Practice Project: Building a Model to Predict Diabetes

8.19 Assignment Two

8.20 Knowledge Check

8.21 Key Takeaways

## Lesson 9 : Natural Language Processing with Scikit Learn

9.1 Nlp Overview

9.2 Nlp Applications

9.3 Knowledge Check

9.4 Nlp Libraries-scikit

9.5 Extraction Considerations

9.6 Scikit Learn-model Training and Grid Search

9.7 Practice Project: Analysing Spam Collection Data

9.8 Demo Assignment

9.9 Practice Project: Sentiment Analysis Using Nlp

9.10 Demo Assignment

9.11 Quiz

9.12 Key Takeaway

## Lesson 10 : Data Visualization in Python using matplotlib

10.1 Introduction to Data Visualization

10.2 Knowledge Check

10.3 Line Properties

10.4 (x,y) Plot and Subplots

10.5 Knowledge Check

10.6 Types of Plots

10.7 Practice Project: Draw a Pair Plot Using Seaborn Library

10.8 Assignment Demo

10.9 Practice Project: Analysing Cause of Death

10.10 Assignment Demo

10.11 Quiz

10.12 Key Takeaways

## Lesson 11 : Web Scraping with BeautifulSoup

11.1 Web Scraping and Parsing

11.2 Knowledge Check

11.3 Understanding and Searching the Tree

11.4 Navigating Options

11.5 Demo3 Navigating a Tree

11.6 Knowledge Check

11.7 Modifying the Tree

11.8 Parsing and Printing the Document

11.9 Practice Project: Web Scraping of Simplilearn Website

11.10 Assignment Demo

11.11 Practice Project: Web Scraping of Simplilearn Website Resource Page

11.12 Assignment Demo

11.13 Quiz

11.14 Key Takeaways

## Lesson 12 : Python integration with Hadoop MapReduce and Spark

## Lesson 13 : Practice Projects

13.1 Practice Project: IBM Hr Analytics Employee Attrition Modelling

# COURSE 4 : Machine Learning

## Lesson 1 : Course Introduction

1.1 Course Introduction

1.2 Accessing Practice Lab

## Lesson 2 : Introduction to AI and Machine Learning

2.1 Learning Objectives

2.2 Emergence of Artificial Intelligence

2.3 Artificial Intelligence in Practice

2.4 Sci-fi Movies With the Concept of Ai

2.5 Recommender Systems

2.6 Relationship Between Artificial Intelligence, Machine Learning, and Data Science: Part a

2.7 Relationship Between Artificial Intelligence, Machine Learning, and Data Science: Part B

2.8 Definition and Features of Machine Learning

2.9 Machine Learning Approaches

2.10 Machine Learning Techniques

2.11 Applications of Machine Learning: Part a

2.12 Applications of Machine Learning: Part B

2.13 Key Takeaways

2.14 Knowledge Check

## Lesson 3 : Data Preprocessing

3.1 Learning Objectives

3.2 Data Exploration Loading Files: Part a

3.3 Data Exploration Loading Files: Part B

3.4 Demo: Importing and Storing Data

3.5 Practice Project: Practice: Automobile Data Exploration - a

3.6 Data Exploration Techniques: Part a

3.7 Data Exploration Techniques: Part B

3.8 Seaborn

3.9 Demo: Correlation Analysis

3.10 Practice Project: Practice: Automobile Data Exploration - B

3.11 Data Wrangling

3.12 Missing Values in a Dataset

3.13 Outlier Values in a Dataset

3.14 Demo: Outlier and Missing Value Treatment

3.15 Practice Project: Practice: Data Exploration - C

3.16 Data Manipulation

3.17 Functionalities of Data Object in Python: Part a

3.18 Functionalities of Data Object in Python: Part B

3.19 Different Types of Joins

3.20 Typecasting

3.21 Demo: Labor Hours Comparison

3.22 Practice Project: Practice: Data Manipulation

3.23 Key Takeaways

3.24 Knowledge Check

3.25 Practice Project: Storing Test Results

## Lesson 4 : Supervised Learning

4.1 Learning Objectives

4.2 Supervised Learning

4.3 Supervised Learning- Real-life Scenario

4.4 Understanding the Algorithm

4.5 Supervised Learning Flow

4.6 Types of Supervised Learning: Part a

4.7 Types of Supervised Learning: Part B

4.8 Types of Classification Algorithms

4.9 Types of Regression Algorithms: Part a

4.10 Regression Use Case

4.11 Accuracy Metrics

4.12 Cost Function

4.13 Evaluating Coefficients

4.14 Demo: Linear Regression

4.15 Practice Project: Practice: Boston Homes - a

4.16 Challenges in Prediction

4.17 Types of Regression Algorithms: Part B

4.18 Demo: Bigmart

4.19 Practice Project: Practice: Boston Homes - B

4.20 Logistic Regression: Part a

4.21 Logistic Regression: Part B

4.22 Sigmoid Probability

4.23 Accuracy Matrix

4.24 Demo: Survival of Titanic Passengers

4.25 Practice Project: Practice: Iris Species

4.26 Key Takeaways

4.27 Knowledge Check

4.28 Practice Project: Health Insurance Cost

## Lesson 5 : Feature Engineering

5.1 Learning Objectives

5.2 Feature Selection

5.3 Regression

5.4 Factor Analysis

5.5 Factor Analysis Process

5.6 Principal Component Analysis (pca)

5.7 First Principal Component

5.8 Eigenvalues and Pca

5.9 Demo: Feature Reduction

5.10 Practice Project: Practice: Pca Transformation

5.11 Linear Discriminant Analysis

5.12 Maximum Separable Line

5.13 Find Maximum Separable Line

5.14 Demo: Labeled Feature Reduction

5.15 Practice Project: Practice: Lda Transformation

5.16 Key Takeaways

5.17 Knowledge Check

5.18 Practice Project: Simplifying Cancer Treatment

## Lesson 6 : Supervised Learning Classification

6.1 Learning Objectives

6.2 Overview of Classification

6.3 Classification: a Supervised Learning Algorithm

6.4 Use Cases of Classification

6.5 Classification Algorithms

6.6 Decision Tree Classifier

6.7 Decision Tree Examples

6.8 Decision Tree Formation

6.9 Choosing the Classifier

6.10 Overfitting of Decision Trees

6.11 Random Forest Classifier- Bagging and Bootstrapping

6.12 Decision Tree and Random Forest Classifier

6.13 Performance Measures: Confusion Matrix

6.14 Performance Measures: Cost Matrix

6.15 Demo: Horse Survival

6.16 Practice Project: Practice: Loan Risk Analysis

6.17 Naive Bayes Classifier

6.18 Steps to Calculate Posterior Probability: Part a

6.19 Steps to Calculate Posterior Probability: Part B

6.20 Support Vector Machines : Linear Separability

6.21 Support Vector Machines : Classification Margin

6.22 Linear Svm : Mathematical Representation

6.23 Non-linear Svms

6.24 The Kernel Trick

6.25 Demo: Voice Classification

6.26 Practice Project: Practice: College Classification

6.27 Key Takeaways

6.28 Knowledge Check

6.29 Practice Project: Classify Kinematic Data

## Lesson 7 : Unsupervised Learning

7.1 Learning Objectives

7.2 Overview

7.3 Example and Applications of Unsupervised Learning

7.4 Clustering

7.5 Hierarchical Clustering

7.6 Hierarchical Clustering Example

7.7 Demo: Clustering Animals

7.8 Practice Project: Practice: Customer Segmentation

7.9 K-means Clustering

7.10 Optimal Number of Clusters

7.11 Demo: Cluster Based Incentivization

7.12 Practice Project: Practice: Image Segmentation

7.13 Key Takeaways

7.14 Knowledge Check

7.15 Practice Project: Clustering Image Data

## Lesson 8 : Time Series Modeling

8.1 Learning Objectives

8.2 Overview of Time Series Modeling

8.3 Time Series Pattern Types: Part a

8.4 Time Series Pattern Types: Part B

8.5 White Noise

8.6 Stationarity

8.7 Removal of Non-stationarity

8.8 Demo: Air Passengers - a

8.9 Practice Project: Practice: Beer Production - a

8.10 Time Series Models: Part a

8.11 Time Series Models: Part B

8.12 Time Series Models: Part C

8.13 Steps in Time Series Forecasting

8.14 Demo: Air Passengers - B

8.15 Practice Project: Practice: Beer Production - B

8.16 Key Takeaways

8.17 Knowledge Check

8.18 Practice Project: Imf Commodity Price Forecast

## Lesson 9 : Ensemble Learning

9.1 Ensemble Learning

9.2 Overview

9.3 Ensemble Learning Methods: Part a

9.4 Ensemble Learning Methods: Part B

9.5 Working of Adaboost

9.6 Adaboost Algorithm and Flowchart

9.7 Gradient Boosting

9.8 Xgboost

9.9 Xgboost Parameters: Part a

9.10 Xgboost Parameters: Part B

9.11 Demo: Pima Indians Diabetes

9.12 Practice Project: Practice: Linearly Separable Species

9.13 Model Selection

9.14 Common Splitting Strategies

9.15 Demo: Cross Validation

9.16 Practice Project: Practice: Model Selection

9.17 Key Takeaways

9.18 Knowledge Check

9.19 Practice Project: Tuning Classifier Model With Xgboost

## Lesson 10 : Recommender Systems

10.1 Learning Objectives

10.2 Introduction

10.3 Purposes of Recommender Systems

10.4 Paradigms of Recommender Systems

10.5 Collaborative Filtering: Part a

10.6 Collaborative Filtering: Part B

10.7 Association Rule Mining

10.8 Association Rule Mining: Market Basket Analysis

10.9 Association Rule Generation: Apriori Algorithm

10.10 Apriori Algorithm Example: Part a

10.11 Apriori Algorithm Example: Part B

10.12 Apriori Algorithm: Rule Selection

10.13 Demo: User-movie Recommendation Model

10.14 Practice Project: Practice: Movie-movie Recommendation

10.15 Key Takeaways

10.16 Knowledge Check

10.17 Practice Project: Book Rental Recommendation

## Lesson 11 : Text Mining

11.1 Learning Objectives

11.2 Overview of Text Mining

11.3 Significance of Text Mining

11.4 Applications of Text Mining

11.5 Natural Language Toolkit Library

11.6 Text Extraction and Preprocessing: Tokenization

11.7 Text Extraction and Preprocessing: N-grams

11.8 Text Extraction and Preprocessing: Stop Word Removal

11.9 Text Extraction and Preprocessing: Stemming

11.10 Text Extraction and Preprocessing: Lemmatization

11.11 Text Extraction and Preprocessing: Pos Tagging

11.12 Text Extraction and Preprocessing: Named Entity Recognition

11.13 Nlp Process Workflow

11.14 Demo: Processing Brown Corpus

11.15 Practice Project: Practice: Wiki Corpus

11.16 Structuring Sentences: Syntax

11.17 Rendering Syntax Trees

11.18 Structuring Sentences: Chunking and Chunk Parsing

11.19 Np and Vp Chunk and Parser

11.20 Structuring Sentences: Chinking

11.21 Context-free Grammar (cfg)

11.22 Demo: Structuring Sentences

11.23 Practice Project: Practice: Airline Sentiment

11.24 Key Takeaways

11.25 Knowledge Check

11.26 Practice Project: Fifa World Cup

## Lesson 12 : Project Highlights

12.1 Project Highlights

12.2 Practice Project: Uber Fare Prediction

12.3 Practice Project: Amazon - Employee Access

## Lesson 13 : Practice Projects

13.1 Practice Project: California Housing Price Prediction

13.2 Practice Project: Phishing Detector With Lr

# COURSE 5 : Tableau 10

## Lesson 1 : Getting Started With Tableau

1.1 Introduction to Tableau and an Overview of the Different Versions

1.2 Installing Tableau Desktop

1.3 Tableau Help and Online Resources

## Lesson 2 : Working With Tableau

2.1 Understanding Tableau User Interface

2.2 Exploring Tableau File Types

2.3 Understanding Green and Blue Pills

2.4 Working With Available Data Sources

2.5 Working With Extracts Instead of Live Connections

## Lesson 3 : Deep diving with Data and Connections

3.1 Working With Excel Data Interpreter

3.2 Learning How to Split Fields

3.3 Pivoting Data

3.4 Understanding Metadata and Sharing Data Source Connections

3.5 Filtering Data from Your Data Source

## Lesson 4 : Creating Charts

4.1 The Show Me Feature

4.2 Crosstabs and Heat Maps

4.3 Using Bar, Stacked Bar and Side-by-side Bars

4.4 Pie Charts

4.5 Line and Area Charts

4.6 Working With Packed Bubble

4.7 Using Treemaps

4.8 Creating a Basic Scatter Plot

4.9 Creating a Basic Map

## Lesson 5 : Adding calculations to your workbook

5.1 Introduction to Calculations

5.2 Understanding Basic Calculations

5.3 Understanding String Calculations

5.4 Learning About Boolean, If-then Calculations, and Case Statements

5.5 Understanding the Basics of Date Calculations

5.6 Understanding Aggregation and Disaggregation

5.7 Using Calculations to Add Insight to Your Visualizations

## Lesson 6 : Mapping data in Tableau

6.1 Introduction to Mapping in Tableau

6.2 Create a Standard Map View

6.3 Dealing With Map Errors

6.4 Customizing a Standard Map View

6.5 Using Filters in Maps

## Lesson 7 : Dashboards and Stories

7.1 Introduction to Dashboards

7.2 Understanding Dashboard Actions

7.3 Understanding Dashboard Formatting Basics

7.4 Understanding Workbook Level Formatting

7.5 Assembling Your Dashboards Into a Story

## Lesson 8 : Visualizations For An Audience

8.1 Focusing on the Narrative

8.2 Using Color With Purpose

8.3 Understanding the Importance of Using Tooltips for Your Audience

8.4 Using a Parameter to Focus the Audience’s Attention

8.5 Removing Clutter in Your Dashboard

8.6 How to Export Your Dashboard to Another File Type

## Lesson 9 :Practice Projects

9.1 Practice Project: Customer Analysis

9.2 Practice Project: Product Analysis

9.3 Practice Project: Sales Dashboard

# COURSE 6 : Big Data Hadoop and Spark Developer

## Lesson 1 : Course Introduction

1.1 Course Introduction

1.2 Accessing Practice Lab

## Lesson 2 : Introduction to Big Data and Hadoop

2.1 Introduction to Big Data and Hadoop

2.2 Introduction to Big Data

2.3 Big Data Analytics

2.4 What is Big Data

2.5 Four vs of Big Data

2.6 Case Study Royal Bank of Scotland

2.7 Challenges of Traditional System

2.8 Distributed Systems

2.9 Introduction to Hadoop

2.10 Components of Hadoop Ecosystem Part One

2.11 Components of Hadoop Ecosystem Part Two

2.12 Components of Hadoop Ecosystem Part Three

2.13 Commercial Hadoop Distributions

2.14 Demo: Walkthrough of Simplilearn Cloudlab

2.15 Key Takeaways

2.16 Knowledge Check

## Lesson 3 : Hadoop Architecture Distributed Storage (HDFS) and YARN

3.1 Hadoop Architecture Distributed Storage (hdfs) and Yarn

3.2 What is Hdfs

3.3 Need for Hdfs

3.4 Regular File System vs Hdfs

3.5 Characteristics of Hdfs

3.6 Hdfs Architecture and Components

3.7 High Availability Cluster Implementations

3.8 Hdfs Component File System Namespace

3.9 Data Block Split

3.10 Data Replication Topology

3.11 Hdfs Command Line

3.12 Demo: Common Hdfs Commands

3.13 Practice Project: Hdfs Command Line

3.14 Yarn Introduction

3.15 Yarn Use Case

3.16 Yarn and Its Architecture

3.17 Resource Manager

3.18 How Resource Manager Operates

3.19 Application Master

3.20 How Yarn Runs an Application

3.21 Tools for Yarn Developers

3.22 Demo: Walkthrough of Cluster Part One

3.23 Demo: Walkthrough of Cluster Part Two

3.24 Key Takeaways

3.25 Knowledge Check

3.26 Practice Project: Hadoop Architecture,distributed Storage (hdfs) and Yarn

## Lesson 4 : Data Ingestion into Big Data Systems and ETL

4.1 Data Ingestion Into Big Data Systems and Etl

4.2 Data Ingestion Overview Part One

4.3 Data Ingestion Overview Part Two

4.4 Apache Sqoop

4.5 Sqoop and Its Uses

4.6 Sqoop Processing

4.7 Sqoop Import Process

4.8 Sqoop Connectors

4.9 Demo: Importing and Exporting Data from Mysql to Hdfs

4.10 Practice Project: Apache Sqoop

4.11 Apache Flume

4.12 Flume Model

4.13 Scalability in Flume

4.14 Components in Flume’s Architecture

4.15 Configuring Flume Components

4.16 Demo: Ingest Twitter Data

4.17 Apache Kafka

4.18 Aggregating User Activity Using Kafka

4.19 Kafka Data Model

4.20 Partitions

4.21 Apache Kafka Architecture

4.22 Demo: Setup Kafka Cluster

4.23 Producer Side Api Example

4.24 Consumer Side Api

4.25 Consumer Side Api Example

4.26 Kafka Connect

4.27 Demo: Creating Sample Kafka Data Pipeline Using Producer and Consumer

4.28 Key Takeaways

4.29 Knowledge Check

4.30 Practice Project: Data Ingestion Into Big Data Systems and Etl

## Lesson 5 : Distributed Processing MapReduce Framework and Pig

5.1 Distributed Processing Mapreduce Framework and Pig

5.2 Distributed Processing in Mapreduce

5.3 Word Count Example

5.4 Map Execution Phases

5.5 Map Execution Distributed Two Node Environment

5.6 Mapreduce Jobs

5.7 Hadoop Mapreduce Job Work Interaction

5.8 Setting Up the Environment for Mapreduce Development

5.9 Set of Classes

5.10 Creating a New Project

5.11 Advanced Mapreduce

5.12 Data Types in Hadoop

5.13 Outputformats in Mapreduce

5.14 Using Distributed Cache

5.15 Joins in Mapreduce

5.16 Replicated Join

5.17 Introduction to Pig

5.18 Components of Pig

5.19 Pig Data Model

5.20 Pig Interactive Modes

5.21 Pig Operations

5.22 Various Relations Performed by Developers

5.23 Demo: Analyzing Web Log Data Using Mapreduce

5.24 Demo: Analyzing Sales Data and Solving Kpis Using Pig

5.25 Practice Project: Apache Pig

5.26 Demo: Wordcount

5.27 Key Takeaways

5.28 Knowledge Check

5.29 Practice Project: Distributed Processing - Mapreduce Framework and Pig

## Lesson 6 : Apache Hive

6.1 Apache Hive

6.2 Hive Sql over Hadoop Mapreduce

6.3 Hive Architecture

6.4 Interfaces to Run Hive Queries

6.5 Running Beeline from Command Line

6.6 Hive Metastore

6.7 Hive Ddl and Dml

6.8 Creating New Table

6.9 Data Types

6.10 Validation of Data

6.11 File Format Types

6.12 Data Serialization

6.13 Hive Table and Avro Schema

6.14 Hive Optimization Partitioning Bucketing and Sampling

6.15 Non Partitioned Table

6.16 Data Insertion

6.17 Dynamic Partitioning in Hive

6.18 Bucketing

6.19 What Do Buckets Do

6.20 Hive Analytics Udf and Udaf

6.21 Other Functions of Hive

6.22 Demo: Real-time Analysis and Data Filteration

6.23 Demo: Real-world Problem

6.24 Demo: Data Representation and Import Using Hive

6.25 Key Takeaways

6.26 Knowledge Check

6.27 Practice Project: Apache Hive

## Lesson 7 : NoSQL Databases HBase

7.1 Nosql Databases Hbase

7.2 Nosql Introduction

7.3 Demo: Yarn Tuning

7.4 Hbase Overview

7.5 Hbase Architecture

7.6 Data Model

7.7 Connecting to Hbase

7.8 Practice Project: Hbase Shell

7.9 Key Takeaways

7.10 Knowledge Check

7.11 Practice Project: Nosql Databases - Hbase

## Lesson 8 : Basics of Functional Programming and Scala

8.1 Basics of Functional Programming and Scala

8.2 Introduction to Scala

8.3 Demo: Scala Installation

8.4 Functional Programming

8.5 Programming With Scala

8.6 Demo: Basic Literals and Arithmetic Operators

8.7 Demo: Logical Operators

8.8 Type Inference Classes Objects and Functions in Scala

8.9 Demo: Type Inference Functions Anonymous Function and Class

8.10 Collections

8.11 Types of Collections

8.12 Demo: Five Types of Collections

8.13 Demo: Operations on List

8.14 Scala Repl

8.15 Demo: Features of Scala Repl

8.16 Key Takeaways

8.17 Knowledge Check

8.18 Practice Project: Basics of Functional Programming and Scala

## Lesson 9 : Apache Spark Next Generation Big Data Framework

9.1 Apache Spark Next Generation Big Data Framework

9.2 History of Spark

9.3 Limitations of Mapreduce in Hadoop

9.4 Introduction to Apache Spark

9.5 Components of Spark

9.6 Application of In-memory Processing

9.7 Hadoop Ecosystem vs Spark

9.8 Advantages of Spark

9.9 Spark Architecture

9.10 Spark Cluster in Real World

9.11 Demo: Running a Scala Programs in Spark Shell

9.12 Demo: Setting Up Execution Environment in Ide

9.13 Demo: Spark Web Ui

9.14 Key Takeaways

9.15 Knowledge Check

9.16 Practice Project: Apache Spark Next Generation Big Data Framework

## Lesson 10 : Spark Core Processing RDD

10.1 Processing Rdd

10.2 Introduction to Spark Rdd

10.3 Rdd in Spark

10.4 Creating Spark Rdd

10.5 Pair Rdd

10.6 Rdd Operations

10.7 Demo: Spark Transformation Detailed Exploration Using Scala Examples

10.8 Demo: Spark Action Detailed Exploration Using Scala

10.9 Caching and Persistence

10.10 Storage Levels

10.11 Lineage and Dag

10.12 Need for Dag

10.13 Debugging in Spark

10.14 Partitioning in Spark

10.15 Scheduling in Spark

10.16 Shuffling in Spark

10.17 Sort Shuffle

10.18 Aggregating Data With Pair Rdd

10.19 Demo: Spark Application With Data Written Back to Hdfs and Spark Ui

10.20 Demo: Changing Spark Application Parameters

10.21 Demo: Handling Different File Formats

10.22 Demo: Spark Rdd With Real-world Application

10.23 Demo: Optimizing Spark Jobs

10.24 Key Takeaways

10.25 Knowledge Check

10.26 Practice Project: Spark Core Processing Rdd

## Lesson 11 : Spark SQL Processing DataFrames

11.1 Spark Sql Processing Dataframes

11.2 Spark Sql Introduction

11.3 Spark Sql Architecture

11.4 Dataframes

11.5 Demo: Handling Various Data Formats

11.6 Demo: Implement Various Dataframe Operations

11.7 Demo: Udf and Udaf

11.8 Interoperating With Rdds

11.9 Demo: Process Dataframe Using Sql Query

11.10 Rdd vs Dataframe vs Dataset

11.11 Practice Project: Processing Dataframes

11.12 Key Takeaways

11.13 Knowledge Check

11.14 Practice Project: Spark Sql - Processing Dataframes

## Lesson 12 : Spark MLLib Modelling BigData with Spark

12.1 Spark Mllib Modeling Big Data With Spark

12.2 Role of Data Scientist and Data Analyst in Big Data

12.3 Analytics in Spark

12.4 Machine Learning

12.5 Supervised Learning

12.6 Demo: Classification of Linear Svm

12.7 Demo: Linear Regression With Real World Case Studies

12.8 Unsupervised Learning

12.9 Demo: Unsupervised Clustering K-means

12.10 Reinforcement Learning

12.11 Semi-supervised Learning

12.12 Overview of Mllib

12.13 Mllib Pipelines

12.14 Key Takeaways

12.15 Knowledge Check

12.16 Practice Project: Spark Mllib - Modeling Bigdata With Spark

## Lesson 13 : Stream Processing Frameworks and Spark Streaming

13.1 Stream Processing Frameworks and Spark Streaming

13.2 Streaming Overview

13.3 Real-time Processing of Big Data

13.4 Data Processing Architectures

13.5 Demo: Real-time Data Processing

13.6 Spark Streaming

13.7 Demo: Writing Spark Streaming Application

13.8 Introduction to Dstreams

13.9 Transformations on Dstreams

13.10 Design Patterns for Using Foreachrdd

13.11 State Operations

13.12 Windowing Operations

13.13 Join Operations Stream-dataset Join

13.14 Demo: Windowing of Real-time Data Processing

13.15 Streaming Sources

13.16 Demo: Processing Twitter Streaming Data

13.17 Structured Spark Streaming

13.18 Use Case Banking Transactions

13.19 Structured Streaming Architecture Model and Its Components

13.20 Output Sinks

13.21 Structured Streaming Apis13.22 Constructing Columns in Structured Streaming

13.23 Windowed Operations on Event-time

13.24 Use Cases

13.25 Demo: Streaming Pipeline

13.26 Practice Project: Spark Streaming

13.27 Key Takeaways

13.28 Knowledge Check

13.29 Practice Project: Stream Processing Frameworks and Spark Streaming

## Lesson 14 : Spark GraphX

14.1 Spark Graphx

14.2 Introduction to Graph

14.3 Graphx in Spark

14.4 Graph Operators

14.5 Join Operators

14.6 Graph Parallel System

14.7 Algorithms in Spark

14.8 Pregel Api

14.9 Use Case of Graphx

14.10 Demo: Graphx Vertex Predicate

14.11 Demo: Page Rank Algorithm

14.12 Key Takeaways

14.13 Knowledge Check

14.14 Practice Project: Spark Graphx

14.15 Project Assistance

## Lesson 15 : Practice Projects

15.1 Practice Project: Car Insurance Analysis

15.2 Practice Project: Transactional Data Analysis

15.3 Practice Project: K-means Clustering for Telecommunication Domain

# COURSE 7 : Data Science Capstone

## Lesson 0 : Day Problem and approach overview

## Lesson 1 : Day Data pre processing techniques application on data set

## 

## Lesson 2 : Day Model Building and fine tuning leveraging various techniques

## 

## Lesson 3 : Day Dashboard problem statement to meet the business objective

## 

## Lesson 4 : Day Final evaluation