

DATA ANALYSIS AND ANALYTICS (1 HOUR)	Types of Data and their Significance Introduction to Data Analytics Difference between Data Analysis and Analytics
DATA REQUIREMENTS (1 HOUR)	Identifying Correct Data Required for Analysis Categorizing Data
DATA COLLECTION (1 HOUR)	Identifying Sources of Data Gathering Meaningful Data
DATA PROCESSING (1 HOUR)	Integrating Data Collected from Various Sources Organizing Data Processing data
DATA CLEANING (2 HOURS)	Finding Issues like Incompleteness, Duplication and Inconsistency in Data Removing and Correcting Issues
EXPLORATORY DATA ANALYSIS (2 HOURS)	Information Design Interactive Data Visualization Descriptive and Inferential Statistics Statistical Graphics
MODELING AND ALGORITHMS (4 HOURS)	Selecting Appropriate Model and Algorithm Based on Given Scenario Applying Algorithms to Identify Relationship Between Variables
TECHNIQUES FOR ANALYZING DATA (2 HOURS)	Hypothesis Testing Regression Analysis
INTRODUCTION TO DATA SETS AND DATA MODELS (2 HOURS)	Creating Datasets from Various Sources Identifying Relevant Variables
WORKING WITH UNSTRUCTURED AND LARGE DATA SETS (2 HOURS)	Handling Unstructured Data Creating Datasets using Unstructured Data Reading Large Datasets
DATA WAREHOUSING AND OLAP (2 HOURS)	Understanding the Concepts of Database, Data Mart and Data Warehouse Introduction to Online Analytical Processing (OLAP)
INTRODUCTION TO BIG DATA (1 HOURS)	Characteristics of Big Data Challenges for Big Data Popular Tools Used to Store, Process, Analyze & Visualize Big Data Use Cases for Big Data
HADOOP ECO-SYSTEM & ARCHITECTURE (2 HOURS)	What is Hadoop? Hadoop's Key Characteristics Hadoop Eco-system & Core Components Where Hadoop Fits? Traditional vs. Hadoop's Data Analytics Architecture When to Use & Not Use Hadoop? Apache Hadoop & Distributions Hadoop Job Trends
HDFS ARCHITECTURE (3 HOURS)	Introduction to Hadoop Distributed File System HDFS Architecture and Features Files and Data Blocks Anatomy of a File Read/ Write on HDFS Replication & Rack Awareness
YARN ARCHITECTURE (3 HOURS)	YARN Architecture Classic vs. YARN YARN Daemons Containers Speculative Execution HDFS Federation Authentication & High Availability
HADOOP SETUP (6 HOURS)	Hadoop Deployment Modes Setting up a Pseudo-distributed Cluster Hortonworks Sandbox Installation & Configuration Linux Terminal Commands Configuration Parameters and Values HDFS File System Operations Working with Hadoop Services using Ambari HDFS, MapReduce and YARN Parameters Hadoop Web UI Filesystem & Linux Commands
MAP REDUCE BASICS (3 HOURS)	What is MapReduce? MapReduce Framework, Architecture and Use Cases Input Splits Hands on with MapReduce Programming Packaging MapReduce Jobs in a JAR
MAP REDUCE ADVANCED (6 HOURS)	Setting Mapper & Reducer Counts Combiners Partitioners & Custom Partitioners Input & Output Formats Sequence Files & Compressions Distributed Cache Map Side Join & Reduce Side Join
HADOOP STREAMING USING PYTHON (4 HOURS)	Hadoop Streaming Concepts Hadoop Streaming using Python Demo: Writing Python Scripts for Streaming Testing Python Scripts Executing YARN Jar on Python Script
USING PIG (5 HOURS)	Background Pig Architecture Pig Latin Basics Pig Execution Modes Pig Processing – Loading and Transforming Data Pig Built-in Functions Filtering, Grouping, Sorting Data Relational Join Operators Pig User Defined Functions

<p>USING HIVE (6 HOURS)</p>	<p>Background of Hive Hive Architecture Warehouse Directory & Metastore Hive Query Language Managed & External Tables Data Processing – Loading Data into Tables Using Hive Built-in Functions Using Joins in Hive Partitioning Data using Hive - Static & Dynamic Bucketing in Hive</p>
<p>WORKING WITH HBASE (6 HOURS)</p>	<p>Hbase Overview HBase Data Model Row Oriented v/s Column Oriented Storage HBase Architecture HBase Shell Commands Bulk Load Data into HBase</p>
<p>SQOOP AND FLUME – DATA INGESTION (4 HOURS)</p>	<p>Why Flume? Setup MySQL RDBMS & Sqoop Sqoop Connectors, Commands Sqoop Options File Importing Data – to HDFS & Hive Exporting Data to MySQL Data Ingestion using Flume Flume Architecture Ingesting Twitter Data into HDFS using Flume</p>
<p>USING OOZIE (6 HOURS)</p>	<p>Overview, Features and Challenges of Oozie The Oozie DAG Architecture Setting up Database & Oozie Configuration Creating Workflows Submitting, Monitoring and Managing Oozie Jobs</p>
<p>USING OOZIE (6 HOURS)</p>	<p>Overview, Features and Challenges of Oozie The Oozie DAG Architecture Setting up Database & Oozie Configuration Creating Workflows Submitting, Monitoring and Managing Oozie Jobs</p>
<p>(PLUS: Project On Hadoop: Project 1 on Hadoop will be explained and assigned. Choose from various domains such as Retail, Telecom, Healthcare, Airlines, Banking and Finance.)</p>	
<p>DATA SCIENCE BASICS (1 HOUR)</p>	<p>Introduction to Data Science What makes a Data Scientist? Introduction to Statistics and Types: Descriptive and Inferential Intro to Research Methods: Pros and Cons of Each</p>
<p>DESCRIPTIVE STATISTICS (1.5 HOURS)</p>	<p>What is It and Methods of Descriptive Statistics Summarizing and Describing Data Sets using a Measure, Central Tendency and Variability</p>
<p>INFERENCE STATISTICS (1.5 HOURS)</p>	<p>What is it and Methods of Inferential Statistics Probability & Central Limit Theorem to Draw Inferences</p>
<p>DESIGN OF EXPERIMENT (2 HOURS)</p>	<p>Use Design of Experiment Methodology for Data Collection</p>
<p>HYPOTHESIS TESTING (2 HOURS)</p>	<p>How to Formulate & Test Hypotheses</p>
<p>INTRODUCTION TO MACHINE LEARNING (2 HOURS)</p>	<p>What is Machine Learning? Types of Problems and Tasks Features, Models and Design of ML Study</p>
<p>DISTANCE-BASED AND LINEAR MODELS (2 HOURS)</p>	<p>KNN K Means Support Vector Machine Perceptron</p>
<p>PROBABILISTIC MODELS (2 HOURS)</p>	<p>Events and their Probabilities Rules of Probability Conditional Probability and Independence Distribution of a Random Variable Expectation and Variance Algorithms like Logistic Regressions & Naïve Bayes</p>
<p>TREE AND BAYESIAN NETWORK MODELS (2 HOURS)</p>	<p>Decision Trees Random Forests Bayesian Models</p>
<p>ADVANCED REGRESSION AND TIME SERIES MODELING (2 HOURS)</p>	<p>Logistic Regression and Decision Trees Definitions, Applications and Techniques What are Moving Average or Smoothing Techniques? Univariate Time Series Models Multivariate Time Series Models</p>
<p>NEURAL NETWORKS (2 HOURS)</p>	<p>Recurrent and Gaussian Neural Networks</p>
<p>BUSINESS COMMUNICATION (5 HOURS)</p>	<p>Business Communication Email/Chat/Telephone Etiquette Listening/Questioning Skills Business Etiquette Interpersonal Skills</p>
<p>ASSESSMENT AND PROJECT EVALUATION (5 HOURS)</p>	<p>Written assessment for Semester 1 and Project 1 Submission & Evaluation</p>

INTRODUCTION TO SAS AND SAS PROGRAMS
(4 HOURS)

What is SAS? | Submitting a SAS Program | SAS Program Syntax

ELECTIVE 1: SAS

ACCESSING DATA, REPORTING AND FORMATTING DATA VALUES
(6 HOURS)

Examining SAS Datasets | Accessing SAS Libraries | Sorting and Grouping and Reporting Data Using SAS Formats

READING AND MANIPULATING DATA
(12 HOURS)

Reading SAS Datasets | Reading Excel Data, Raw Files, Database Data | Creating Summary Reports Combining Datasets

SUMMARIZING DATA
(10 HOURS)

Writing Observations | Writing to Multiple Datasets | Creating Accumulating Total for a Group of Data

(PLUS: Project On SAS: Project on SAS will be explained and assigned. Choose from various domains such as Retail, Telecom, Healthcare, Airlines, Banking and Finance.)

SUMMARIZING DATA
(4 HOURS)

Functions in SAS | Transposing Data | Character and Numeric Functions | Converting Variable Type Reading Formatted Input

RESTRUCTURING DATASET
(4 HOURS)

Do Loop Processing | Conditional Do Loop Processing | Using SAS Arrays | SAS Array Processing

BASIC STATISTICS, DECISION TREES AND NEURAL NETWORKS
(4 HOURS)

Mean, Median, Standard Deviation | Outliers | Statistical Graphs | Variables and its Type Measure of Central Tendency and Dispersion | Normal Distribution Introduction to Predictive Modelling with Decision Trees | Assumptions | Formulate the Model | Estimate the Parameters | Check the Prediction Accuracy

CROSS TABULATION AND CORRELATION ANALYSIS
(4 HOURS)

Sample and Population | Formulate the Hypothesis | Select an Appropriate Test | Choose level of Significance | Calculate Test Statistics | Determine the Probability | Compare the Probability and Make Decision | One Sample T-Test | Two Independent Samples Tests | Paired T-test Proportional Test | Non Parametric One Sample Test | Case Study on Hypothesis and Parametric Testing

CROSS TABULATION AND CORRELATION ANALYSIS
(10 HOURS)

One Way Analysis of Variance | Assumption of ANOVA | Statistics Associated with One Way Analysis of Variance | Decomposing the Variations | Interpreting the ANOVA Results | Two Way Analysis of Variance | Interpreting the ANOVA Results | Analysis of Covariance | Regression Analysis Linear and Logistic Regression | Statistics Associated with Regression | Relative Importance of Predictors Partial Regression Coefficient | Examination of Residual | Multi-Collinearity Case Study on ANOVA/Regression Analysis

MULTIVARIATE ANALYSIS
(5 HOURS)

Hierarchical Clustering | K-Means Clustering | Factor Analysis | Principal Component Analysis Market Basket Analysis

BUSINESS FORECASTING
(5 HOURS)

Basic Time Series | ARIMA Model | Exponential Smoothing Models

INTRODUCTION TO R
(3 HOURS)

R-Studio | Packages in R | Installing Packages | Setting Directories

BASIC OPERATIONS
(4 HOURS)

Programming Language Basics | Scalars, Vectors | Simple Calculations | Data Structure Data Frames

DATA MANIPULATION
(8 HOURS)

Data Acquisition(Import & Export) | Sub-setting Observations and Variables | Transforming Variables Renaming and Recoding Variables | Conditional Processing | Missing Values | Merging and Concatenating Datasets | Using SQL in R

(PLUS: Project on R: Project on R will be explained and assigned. Choose from various domains such as Retail, Telecom, Healthcare, Airlines, Banking and Finance.)

BASIC STATISTICS IN R
(6 HOURS)

Descriptive Statistics | Cross-Tabs | T-Tests | ANOVA, n-way ANOVA | Correlations | Sampling Bootstrapping | Simulating and Random Number Generation

ELECTIVE 2: R

GRAPHICS IN R
(1.5 HOURS)

Line Plots | Bar Charts | Pie Charts | Histograms | Scatter Plots | 3-D, Parallel Coordinates

LINEAR MODELS
(2 HOURS)

Simple Regression Analysis | Multiple Regression

CURRICULUM DIPLOMA IN BIG DATA & ANALYTICS

LOGISTICS REGRESSION & MODEL SELECTION (6 HOURS)	Logit Function Maximum Likelihood Optimization Odds Ratio Selection Criterion Stepwise Methods
WRITING FUNCTIONS (5 HOURS)	Creating User Defined Functions
PREDICTIVE MODELING USING REGRESSION ANALYSIS (6 HOURS)	Linear Regression: Modelling Assumptions Logistic Regression: Modelling Assumptions Model Selection & Scoring Business Case Study on Regression Analysis
DECISION TREES ANALYSIS & RANDOM FOREST AND OTHER TOOLS (6 HOURS)	Recursive Partitioning CART Splitting Criterion Pruning Support Vector Machine Naïve Bayes K-Nearest Neighbor Algorithm Business Case Study on Predictive Analytics
TEXT MINING (3 HOURS)	Preparing and Loading a Corpus, pdf, Word, Text Files Performing Stemming Creating & Exploring Term-Document Matrix N-Gram Analysis
VISUALIZATIONS IN TEXT MINING (2 HOURS)	Creating and Modifying a WordCloud Reshaping, Resizing and Recoloring Word Association Analysis
MULTIVARIATE ANALYSIS (11 HOURS)	Clustering: Hierarchical Clustering & K-means Clustering Variable Reduction: Factor Analysis & Principal Component Analysis Association Mining: Market Basket Analysis Business Case Study on Multivariate Analysis
OPTIMIZATION (7.5 HOURS)	Linear Programming & Goal programming in R Objective Functions Constraints and Solving LP & GP Problems Integer Programming and Mixed Integer Linear Programming Objective Functions Constraints Solving ILP and MILP Problems Business Case Study on Optimization
BUSINESS FORECASTING (9 HOURS)	Basic Time Series Exponential Smoothing Models ARIMA Modelling in R Business Case Study on Forecasting
RESUME WRITING (5 HOURS)	The Why, The What and the How of Resumes Personal Branding Tips and Resources Interview Skills Using Social Media CV Discussion
ASSESSMENT AND PROJECT EVALUATION (5 HOURS)	Written assessment for Semester 2 and Project Submission & Evaluation

SEMESTER 3: ADVANCED

75 HOURS

PYTHON IDE (2 HOURS)	What is Python? Installing Anaconda Understanding the Spyder Integrated Development Environment (IDE)
DATA STRUCTURES IN PYTHON USED FOR DATA ANALYSIS (2 HOURS)	Intro to Numpy Arrays Creating ndarrays Indexing Data Processing using Arrays File Input and Output Getting Started with Pandas
DATA FRAME MANIPULATION (6 HOURS)	Data Acquisition(Import & Export) Indexing Selection and Filtering Sorting & Summarizing Descriptive Statistics Combining and Merging Data Frames Removing Duplicates Discretization and Binning String Manipulation
(PLUS: Project Work on Tableau)	
VISUALIZATION (3 HOURS)	Line Plots Bar Charts Pie Charts Histograms Scatter Plots Parallel Coordinates
OTHER PREDICTIVE MODELLING TOOLS (2 HOURS)	Intro to Machine Learning Intro to Sklearn Library and Statsmodels
ANALYTICS USING PYTHON (6 HOURS)	Simple Regression Analysis Multiple Regression Logit Function Maximum Likelihood Optimization Odds Ratio
ADVANCED ANALYTICS (3 HOURS)	Recursive Partitioning CART Splitting Criterion Pruning Random Forests Support Vector Machine K-Nearest Neighbor Algorithm Naive Bayes

<p>CLUSTERING (2 HOURS)</p>	<p>Hierarchical Clustering K-means Clustering</p>
<p>BASICS OF TEXT MINING (4 HOURS)</p>	<p>Creating and Modifying a WordCloud Reshaping, Resizing and Recoloring Word Association Analysis Text Clustering & Categorization Sentiment Analysis Text Clustering & Categorization</p>
<p>INTRODUCTION TO VISUALIZATION (2 HOURS)</p>	<p>Relevance of Visualization Comparison of Visualization What is Tableau Uses of Tableau</p>
<p>WORKING WITH TABLEAU (3 HOURS)</p>	<p>Installation and Architecture Working with Tableau Exporting, Connecting and Loading Sample Use Cases</p>
<p>VISUALIZATION IN DEPTH (3 HOURS)</p>	<p>Highlight Tables Heat Maps Circle Plots Side by Side Bars Continues Lines Scatter Plots Dual Charts Area Charts Tree Maps</p>
<p>(PLUS: Project Work on Tableau)</p>	
<p>DATA ORGANIZATION (3 HOURS)</p>	<p>Calculated Metrics Sorting, Filtering Totals and Sub Totals Aggregated Measures Percentages Data Spotlighting Summary Card Date and Time Functions String Functions and Logical Functions Data Items, Hierarchies, Sets</p>
<p>ADVANCED VISUALIZATION (3 HOURS)</p>	<p>Dual Charts, Dual Axis Combination Charts, Gantt Charts for Activity Tracking Motion Charts, Box and Whisker Plots Reference Lines, Reference Bands Pareto Analysis Water Fall Charts Market Basket Analysis</p>
<p>TIME DIMENSION (3 HOURS)</p>	<p>Quick Table Calculations Custom Table Calculations YTD Parallel Period Moving Averages Window Averages Predictive Models Forecasting under Trend Lines Use Cases, Tips-Tricks Recommendations for Forecasting</p>
<p>MAPPING (3 HOURS)</p>	<p>Filled Maps Symbol Maps Geo Coding Online Maps Sorting in Maps Background Maps</p>
<p>ENTERPRISE DASHBOARDS (3 HOURS)</p>	<p>Building a Dashboard Dashboard Layouts and Formatting Dashboard Interactivity using Actions Dashboard Best Practices</p>
<p>DATA PRESENTATION (3 HOURS)</p>	<p>Data Summarization Interacting With Data Making Presentations Relevant Publishing on the Web Visual Best Practices</p>
<p>WORKING WITH MULTIPLE SOURCES (2 HOURS)</p>	<p>Data Blending from Multiple Source Systems Custom SQL Queries Creating Incremental Loads Creating File Extractions, Parameters</p>
<p>APIs (1 HOURS)</p>	<p>Extract API Introduction Extract API Connecting to Data REST API Javascript API</p>
<p>INTERVIEW PREPARATION - HR (4 HOURS)</p>	<p>Interview Preparation Stage 1 - Preparation Stage 2 - First Impressions Count Stage 3 - The Interview Stage 4 - The Final Stage Questions You May Wish To Ask How To Ace The 50 Most Common Interview Questions</p>
<p>INTERVIEW PREPARATION - DOMAIN (4 HOURS)</p>	<p>Analytics Experts Guide You on How to Prepare for Technical Interview Round with Tips, Tricks and a Quick Refresher on Concepts Learnt</p>
<p>MOCK INTERVIEWS - HR AND DOMAIN (4 HOURS)</p>	<p>1:1 or Panel Mock Interviews with Industry Veterans to Clear the HR and Technical Round of Interviews to Give You Confidence to Face Real World Scenarios</p>
<p>FINAL PROJECT PRESENTATION (4 HOURS)</p>	<p>Final Project Submission in Front of Your Peers While Industry Experts Evaluate Your Project Work with Constructive Criticism</p>