SEMESTER 1: DATA ANALYSIS, SCIENCE AND BIG DATA HADOOP — 110 HOURS

DATA ANALYSIS (6 HOURS)	Applications | Careers | Types of Data |Data Mining & Architecture |Data Warehousing | MDM Vs. DW | What makes a Data Scientist? | Statistics
PROGRAMMING (8 HOURS)	Basics of Java | Creating Java Application Programs | Implementing Loops, Arrays | Basic Commands of Linux for Better Understanding of Hadoop
BIG DATA (2 HOURS)	Characteristics of Big Data | Challenges | Popular Tools Used to Store, Process, Analyze & Visualize Big Data | Use Cases for Big Data
HADOOP ECO-SYSTEM (2 HOURS)	Characteristics | Eco-system & Core Components | Where Hadoop Fits? | When to Use & Not Use | Apache Hadoop Distributions | Job Trends |
HDFS & YARN (4 HOURS)	HDFS Architecture and Features | Files and Data Blocks | | Classic vs. YARN | Daemons | Containers | Speculative Execution | HDFS Federation
HADOOP SETUP (6 HOURS)	Deployment Modes | Setting up a Pseudo-distributed Cluster | Hortonworks Sandbox Configuration | HDFS File System Operations | Hadoop Services using Ambari | Web UI | Filesystem & Linux Commands
MAPREDUCE (9 HOURS)	Architecture and Use Cases | Programming | Packaging MapReduce Jobs in a JAR | Mapper & Reducer Counts | Partitioners & Custom Partitioners
HADOOP STREAMING (2 HOURS)	Hadoop Streaming using Python | Demo: Writing Python Scripts for Streaming | Testing Python Scripts | Executing YARN Jar on Python Script
PIG (6 HOURS)	Architecture | Pig Latin Basics | Execution Modes | Pig Processing | Built-in and User Defined Functions & Operators | Filtering, Grouping, Sorting Data
HIVE (6 HOURS)	Architecture | Warehouse Directory & Metastore | Query Language | Data Processing | Built-in Functions | Joins and Bucketing | Partitioning Data
HBASE (4 HOURS)	Hbase Data Model | Row Oriented v/s Column Oriented | Storage | Architecture | Shell Commands | Bulk Load Data into Hbase
SQOOP AND FLUME (4 HOURS)	Setup MySQL RDBMS & Sqoop | Sqoop Connectors, Commands | Importing Data to HDFS & Hive | Exporting Data to MySQL | Flume Data Ingestion
OOZIE (6 HOURS)	Features and Challenges | DAG Architecture | Setting up Database & Oozie Configuration | Creating Workflows | Submittingand Managing Oozie Jobs
PROBABILITY THEORY (7 HOURS)	Events, Probabilities, Rules | Conditional Probability | Distribution| Central Limit Theorem | Expectation & Variance | Naïve Bayes |Design of Experiments
BASIC STATISTICS (2 HOURS)	Events and their Probabilities | Rules of Probability | Distribution of a Random Variable | Central Limit Theorem | Naïve Bayes
HYPOTHESIS & OTHER TESTS (6 HOURS)	Hypothesis | Probability | One Sample / Two Samples T-Test | Paired T-test | Proportional, Non Parametric One Sample, Chi Square, Z, F Test
CORRELATION ANALYSIS (2 HOURS)	Pattern Discovery | Statistics Associated with Cross Tabulations | Chi Square, Phi Coefficient, Contingency Coefficient | Correlation Analysis
LINEAR REGRESSION (6 HOURS)	Assumptions | Hypothesis | Variable and Model Significance | Regression Table | Anova Table | Multicollinearity | Heteroscedasticity |
ANOVA (4 HOURS)	One Way Analysis of Variance | Assumptions | Statistics | Interpreting Results | Two Way Analysis of Variance | Analysis of Covariance
LOGISTIC REGRESSION (6 HOURS)	Assumptions | Reason for Logit Transform | Hypothesis | Variable Model Significance | Regression Table | Chi Squre Test | ROC Curve

MACHINE LEARNING (2 HOURS)	What is Machine Learning? | Types of Problems and Tasks | Features, Models and Design of ML Study
OTHER MODELS (6 HOURS)	Distance-based and Non Linear Models | KNN | K Means | SVM |Bayesian Network Models | Neural Networks | Perceptron, MLP, Back Propagation

SEMESTER 2: SPECIALIZATION IN EITHER R OR PYTHON — 85 HOURS

INTRODUCTION TO R SOFTWARE (10 HOURS)	Installation | Architecture | Installing Packages | Installing Packages | Setting Directories | Basic Operations | Scalars, Vectors
LINEAR REGRESSION (8 HOURS)	Basic Statistics Refresher | Covariance and Correlation | Multivariate Analysis | Assumptions of Linearity | Hypothesis Testing | Limitations of Regression |Case Study For Linear Regression: Case for Prediction Problem
LOGISTIC REGRESSION (8 HOURS)	The Logistic Transform | Logistic Regression Modelling | Model Optimisation | Understanding the ROC Curve | Case Study
REGRESSION (8 HOURS)	The Logistic Transform | Modelling | Model Optimisation | ROC Curve | Case Study For Logistic Regression: Case for Prediction Problem
DECISION TREE (4 HOURS)	Classification Trees | Regression Trees |Case Study for Decision Tree
SEGMENTATION (4 HOURS)	Clustering | Kmeans Algorithm | Cluster Size vs Definition Optimisation | K- medioid and Fuzzy K means |Case for Clustering on Bank Customer Data Set
ASSOCIATION RULE MINING (4 HOURS)	Supervised Vs Unsupervised Learning | Recommendation Engines | Association Rule Mining | Case Study For Market Basket Analytics
TIME SERIES (4 HOURS)	Time Series Decomposition | Moving Average & Exponential Smoothing Methods | AR, MA, ARIMA, SARIMA, RMSE and MAPE | Case Study
KNN ALGORITHM (6 HOURS)	K Nearest Neighbours Algorithm for Classification | Lazy Learning Notion | Data Transformations | Evaluation of Model | Pros and Cons | Case Study
NAÏVE BAYES ALGORITHM (6 HOURS)	Bayesian Theorem | Probabilities | Conditional and Joint Probabilities Notion | Traditional and Naive Approach | Model Building | Case Study
ANN & SVM (12 HOURS)	Neural Networks | Structure of Network | The ANN Model | Training the Model | Testing and Validation | SVM | Tuning the Model | Case Study
ENSEMBLE MODELS (6 HOURS)	Entropy | Information Value | Decision Tree Pruning | Model Validation & Performance | Bagging, Boosting Trees | Random Forests |Case Study
PYTHON BASICS (4 HOURS)	What is Python? | Installing Anaconda | Spyder Integrated Development Environment (IDE) | Python Basics and String Manipulation
DEALING WITH DATA (12 HOURS)	Data Management | Lists, Tuples, Dictionaries, Variables |Crud Operations | Pydoop, Pymongo, Pyspark |Data scraping and Collection | Data Structures in Python Used for Data Analysis: Numpy Arrays, Indexing | Pandas
DATA FRAME MANIPULATION (4 HOURS)	Data Management | Lists, Tuples, Dictionaries, Variables |Crud Operations | Pydoop, Pymongo, Pyspark |Data scraping and Collection | Data Structures in Python Used for Data Analysis: Numpy Arrays, Indexing | Pandas
NATURAL LANGUAGE PROCESSING (8 HOURS)	Text Preprocessing | Stemming | Bag of Words Approach and Naïve Bayes | Latent Semantic Analysis | Tagging, Categorization | Sentiment Analysis

VISUALIZATION (6 HOURS)	Image Processing | Extractors for Image Processing | Using Classifiers | Text Extraction from Image | Data Visualization in Python: Charts, Plots etc
OTHER TOOLS (4 HOURS)	Other Predictive Modelling Tools | Machine Learning | Sklearn Library and Statsmodels |Simple Regression Analysis | Multiple Regression
KNN AND NAÏVE BAYES ALGORITHMS (12 HOURS)	K Nearest Neighbours Algorithm for Classification | Naïve Bayes Algorithm for Multi Class Predictions |Model Building, Testing| Case Studies
ANN AND SVM (12 HOURS)	Artificial Neural Networks | Structure of Network | Support Vector Machines | Build, Test, Train, Validate Model | Case Studies
ENSEMBLE MODEL (6 HOURS)	Entropy | Information Value | Decision Tree Pruning | Model Validation & Performance | Bagging, Boosting Trees | Random Forests |Case Study

SEMESTER 3: SAS, TABLEAU AND INTERVIEW PREP — 70 HOURS

INTRODUCTION (12 HOURS)	What is SAS? | Submitting a SAS Program | SAS Program Syntax |Accessing Data, Reporting and Formatting Data Values |SAS Datasets |SAS Libraries
DATA MANIPULATION (12 HOURS)	Reading SAS Datasets, Excel Data, Raw Files, Database Data | Creating Summary Reports | Combining Datasets |Summarizing Data | Observations | Functions in SAS | Data Transformations |Reading Formatted Input | Restructuring Dataset | Do Loop Processing | SAS Array Processing
BASIC STATISTICS & REGRESSION (12 HOURS)	Measures of Central Tendency | Measures of Dispersion | Skewness and Kurtosis | Linear Regression & Modeling | Logistic Regression & ROC Curve | Model Parameter Significance Evaluation | Case Studies
TIME SERIES (6 HOURS)	Time Series Decomposition | Simple & Weighted Moving Average Method | Exponential Smoothing Methods | Stationarity of Data | ARIMA Models |Case Studies on Time Series
TABLEAU (2 HOURS)	Relevance of Visualization | What is Tableau? | Uses | Installation and Architecture | Working with Tableau | Exporting, Connecting and Loading | Sample Use Cases
DATA AND CHARTS (4 HOURS)	Different Types of Charts | Data Organization | Calculated Metrics | Sorting, Filtering | Totals & Sub Totals | Aggregated Measures
VISUALIZATION (4 HOURS)	Advanced Visualization | Combination Charts | Reference Lines, Reference Bands | Pareto Analysis | Market Basket Analysis | Mapping
DATA PRESENTATION (5 HOURS)	Dashboard Layouts and Formatting | Interactivity using Actions | Dashboard Best Practices | Data Summarization | Making Presentations Relevant | Publishing on Web