



**millionlights<sup>®</sup>**

*Nayi Shiksha, Nayi Soch*

**COMPLETE DATA  
SCIENCE TRACK  
COURSE CONTENT**



# **[I] Get Started with Data Science**

## **Introduction to Data Science**

**Duration:** 24 hrs

**Course Content:**

❖ **Introduction**

- Before You Start
- Course Introduction

❖ **Module 1: An Introduction to Data**

- Exploring Data
- Lab: Exploring Data

❖ **Module 2: Data Analysis Fundamentals**

- Analyzing and Visualizing Data
- Lab: Analyzing and Visualizing Data

❖ **Module 3: Getting Started with Statistics**

- An Introduction to Statistics
- Lab: Basic Statistical Analysis

❖ **Module 4: Machine Learning Basics**

- Introduction to Machine Learning
- Lab: Machine Learning





## **[II] Analyse and Visualize Data** **2 TRAINING OPTIONS AVAILABLE**

### ➤ **Option 1: Analysing and Visualizing Data with Power BI**

**Duration:**24 hrs

#### **Course Content:**

##### ❖ **Start Here!**

- Welcome to the Course!
- Set up the Lab Environment
- DataSet / Examples Usage
- Pre-course survey

##### ❖ **Introduction**

- What is Power BI?

##### ❖ **Power BI Desktop Data Transformations**

- Transforming Data using Power BI Desktop
- Lab 1

##### ❖ **Power BI Desktop Modelling**

- Data Modelling in Power BI Desktop
- Lab 2

##### ❖ **Power BI Desktop Visualization**

- Visualizing Your Data
- Working with Multiple Visualizations
- Lab 3

##### ❖ **Power BI Service**

- Working with Power BI Service
- Viewing Power BI Dashboard
- Lab 4

##### ❖ **Working with Excel**

- Connecting and Collaborating with Excel
- Lab 5

##### ❖ **Organization Packs, Security and Groups**

- Working with Others
- Lab 6.





### ❖ **Direct Connectivity**

- Direct Connectivity
- Lab 7

### ❖ **Developer API**

- Developer API
- Lab 8

### ❖ **Mobile App**

- Mobile App

### ❖ **Course Wrap-up**

## ➤ **Option 2: Analyzing and Visualizing Data with Excel**

**Duration:**24 hrs

**Course Content:**

### ❖ **Start Here!**

- Welcome to the Course!
- Set up the Lab Environment
- DataSet / Examples Usage
- Pre-course Survey

### ❖ **Module 1**

- Data Analysis in Excel
- Lab: Explore and Extend a Classic Excel Dashboard

### ❖ **Module 2**

- The Excel Data Model
- Basic DAX
- Lab: Explore an Excel Data Model

### ❖ **Module 3**

- Importing Data from a CSV File
- Lab: Importing Data from a CSV File

### ❖ **Module 4**

- Importing Data from Databases
- Importing Data from Multiple Files
- Create a Date Table in Excel Data Model
- Lab: Creating Mash-ups of Data from Multiple Sources





### ❖ **Module 5**

- Creating and Formatting Measures
- Using Advanced DAX Functions
- Lab: Creating Measures using Advanced DAX Functions

### ❖ **Module 6**

- Importing Data from a Formatted Excel Report
- Lab: Advanced Text Query

### ❖ **Module 7**

- Visualizing Data in Excel
- Lab: Data Visualizations in Excel

### ❖ **Module 8**

- Using Excel with Power BI
- Power BI Mobile App





## **[III] Communicate Data Insights** **Analytics Storytelling for Impact**

**Duration:** 24 hrs

**Course content:**

❖ **Welcome**

- Welcome
- Introductions and Overview
- Setup
- Data Set
- Getting Started

❖ **The Power of Analytics Storytelling**

- The Power of Storytelling
- Analytics Storytelling
- Lab

❖ **Craft Your Analytics Story**

- Scope Your Analytics Story
- Craft Your Analytics Story
- Lab

❖ **Perfect Your Analytics Story**

- Leverage Visual Science
- Analytics Story Examples
- Lab

❖ **Land Your Analytics Story**

- Land Your Story
- Next Steps for Mastery
- Lab

❖ **Final Evaluation**

- Final Evaluation





## **[IV] Apply Ethics and Law in Analytics**

### **Ethics and Law in Data and Analytics**

**Duration:** 18 hrs

**Course content:**

❖ **Welcome**

- Welcome
- Introductions and Overview
- Setup
- Getting Started

❖ **Data, Ethics, and Law**

- Data's Ethical Foundations
- Data's Legal Foundations
- Ethical Data Practice
- Lab

❖ **Data, Individuals, and Society**

- Data Bias and Identity
- Data Privacy and Power
- Lab

❖ **Data Ethics and Law in Business**

- Business and Ethical Data Use
- Business and Data Privacy
- Lab

❖ **Artificial Intelligence and Future Opportunities**

- AI and Design
- XAI
- Lab

❖ **Final Evaluation**

- Final Evaluation





# **[V] Query Relational Data**

## **Querying Data with Transact-SQL**

**Duration:**30 hrs

### **Course Content:**

#### **❖ Before You Start**

- Introduction

#### **❖ Section 1: Modules 1-2**

- Module 1: Introduction to Transact-SQL
- Lab 1: Introduction to Transact-SQL
- Module 2: Querying Tables with SELECT
- Lab 2: Querying Tables with SELECT

#### **❖ Section 2: Modules 3-5**

- Module 3: Querying Multiple Tables with Joins
- Lab 3: Querying Multiple Tables with Joins
- Module 4: Using Set Operators
- Lab 4: Using Set Operators
- Module 5: Using Functions and Aggregating Data
- Lab 5: Using Functions and Aggregating Data

#### **❖ Section 3: Modules 6-8**

- Module 6: Using Subqueries and APPLY
- Lab 6: Using Subqueries and APPLY
- Module 7: Using Table Expressions
- Lab 7: Using Table Expressions
- Module 8: Grouping Sets and Pivoting Data
- Lab 8: Grouping Sets and Pivoting Data

#### **❖ Section 4: Modules 9- 11**

- Module 9: Modifying Data
- Lab 9: Modifying Data
- Module 10: Programming with Transact-SQL
- Lab 10: Programming with Transact-SQL
- Module 11: Error Handling and Transactions
- Lab 11: Error Handling and Transactions

#### **❖ Final Assessment**

- Final Assessment







## **[VI] Explore Data with Code**

### **2 TRAINING OPTIONS AVAILABLE**

#### ➤ **Option 1: Introduction to R for Data Science**

**Duration:** 12 hrs

**Course Content:**

##### ❖ **Start Here**

- Welcome to the Course!
- Pre-course survey

##### ❖ **Basics**

- Lecture: R: The true basics
- Lab: R: The true basics
- Lecture: Basic data types
- Lab: Basic Data Types
- Further Readings

##### ❖ **Vectors**

- Lecture: Create and name vectors
- Lab: Create and name vectors
- Lecture: Vector Arithmetic
- Lab: Vector Arithmetic
- Lecture: Vector Subsetting
- Lab: Vector Subsetting
- Further Readings

##### ❖ **Matrices**

- Lecture: Create and name matrices
- Lab: Create and name matrices
- Lecture: Matrix subsetting
- Lab: Matrix subsetting
- Lecture: Matrix Arithmetic
- Lab: Matrix Arithmetic
- Further Readings

##### ❖ **Factors**

- Lecture: Factors
- Lab: Factors
- Further Readings





### ❖ Lists

- Lecture: Create and name lists
- Lab: Create and name lists
- Lecture: Subset and extend lists
- Lab: Subset and extend lists
- Further Readings

### ❖ Data frames

- Lecture: Explore the data frame
- Lab: Explore the data frame
- Lecture: Subset, extend and sort your data frame
- Lab: Subset, extend and sort your data frame
- Further Readings

### ❖ Graphics

- Lecture: Basic graphics
- Lab: Basic graphics
- Lecture: Customizing your plots
- Lab: Customizing your plots
- Lecture: Multiple plots
- Lab: Multiple plots

### ❖ Final Lab

- Final Lab
- Lab

## ➤ Option 2: Introduction to Python for Data Science

**Duration:** 24 hrs

**Course content:**

### ❖ Start Here

- Welcome to the Course!
- Pre-course survey

### ❖ Python Basics

- Lecture: Hello Python!
- Lab: Hello Python!
- Lecture: Variables and Types
- Lab: Variables and Types
- Further Readings





## ❖ **List - A Data Structure**

- Lecture: Python Lists
- Lab: Python Lists
- Lecture: Subsetting Lists
- Lab: Subsetting Lists
- Lecture: Manipulating Lists
- Lab: Manipulating Lists
- Further Readings

## ❖ **Functions and Packages**

- Lecture: Functions
- Lab: Functions
- Lecture: Methods
- Lab: Methods
- Lecture: Packages
- Lab: Packages
- Further Readings

## ❖ **Numpy**

- Lecture: Numpy
- Lab: Numpy
- Lecture: 2D Numpy Arrays
- Lab: 2D Numpy Arrays
- Lecture: Basic Statistics with Numpy
- Lab: Basic Statistics with Numpy
- Further Readings

## ❖ **Plotting with Matplotlib**

- Lecture: Basic Plot with matplotlib
- Lab: Basic Plots with matplotlib
- Lecture: Histograms
- Lab: Histograms
- Lecture: Customization
- Lab: Customization
- Further Readings

## ❖ **Control Flow and Pandas**

- Lecture: Boolean Logic and Control Flow
- Lab: Boolean Logic and Control Flow
- Lecture: Pandas
- Lab: Pandas





- Further Readings

❖ **Final Lab**

- Final Lab





## **[VII] Apply Math and Statistics to Data Analysis**

### **3 TRAINING OPTIONS AVAILABLE**

#### **➤ Option 1: Essential Math for Machine Learning: R Edition**

**Duration:** 48hrs

**Course content:**

##### **❖ Before You Start**

- Course Introduction

##### **❖ Equations, Graphs, and Functions**

- Lesson 1: Algebra Fundamentals
- Lesson 2: Quadratic Equations and Functions
- Module Assessment
- Knowledge Check

##### **❖ Derivatives and Optimization**

- Lesson 1: Differential Calculus Foundations
- Lesson 2: Differentiation and Derivatives
- Module Assessment
- Knowledge Check

##### **❖ Vectors and Matrices**

- Lesson 1: Vectors
- Lesson 2: Matrices
- Module Assessment
- Knowledge Check

##### **❖ Statistics and Probability**

- Lesson 1: Statistics Fundamentals
- Lesson 2: Probability





- Module Assessment
- Knowledge Check

➤ **Option 2: Essential Statistics for Data Analysis using Excel**

**Duration:** 24 hrs

❖ **Introduction**

- Online Courses from Microsoft
- Welcome
- Course Syllabus
- Course Schedule and Grading
- Statistics and Probability Primer
- Set up Excel
- Download Course Files
- Pre-Course Survey

❖ **Module 1: Descriptive Statistics**

- Defining Data
- Histograms and Skewness
- Descriptive Statistics with Analysis ToolPak
- Boxplots
- Categorical Data, PivotTables, and PivotCharts
- Summarizing Hierarchical Data
- 80-20 Rule and Pareto Charts
- Module 1 Discussion
- Module 1 Quiz

❖ **Module 2: Basic Probability**

- Introduction to Probability
- Law of Complements
- Mutually Exclusive and Independent Events
- Conditional Probability
- Law of Total Probability and Bayes Rule
- Additional Reading and Review
- Module 2 Discussion
- Module 2 Quiz

❖ **Module 3: Random Variables**

- Random Variable Definitions





- Mean, Variance, and Standard Deviation of a Random Variable
- Mean, Variance, and Standard Deviation for Sum of Random Variables
- Binomial Random Variable
- Poisson Random Variable
- Normal Random Variable
- Central Limit Theorem
- Z Scores
- Module 3 Discussion
- Module 3 Quiz

#### ❖ **Module 4: Sampling and Confidence Intervals**

- Populations and Samples
- Point Estimation of a Population Mean and Proportion
- The Standard Normal
- Confidence Interval Estimation
- Sample Size Determination
- The Finite Correction Factor
- Additional Reading
- Module 4 Discussion
- Module 4 Quiz

#### ❖ **Module 5: Hypothesis Testing**

- Defining Hypotheses
- Type I and Type II Error
- One Sample Z-Test
- One Sample T-Test
- Single Sample Test for Population Proportion
- Testing Equality of Variances
- Testing the Difference Between Two Population Means
- Chi-Squared Test for Independence
- Additional Reading
- Module 5 Discussion
- Module 5 Quiz

#### ❖ **Final**

- Final Exam





## **[VIII] Plan and Conduct Data Studies** **2 TRAINING OPTIONS AVAILABLE**

### ➤ **Option 1: Essential Math for Machine Learning: Python**

**Duration:** 18Hrs

**Course content:**

#### ❖ **Introduction**

- Welcome to the Course
- About this Course
- Setup
- Getting Started

#### ❖ **The Research Process**

- The Research Process
- The Psychology of Providing Data
- Knowledge Check

#### ❖ **Planning for Analysis**

- Planning for Analysis
- Power and Sample Size Planning
- Research Practices
- Knowledge Check
- Lab

#### ❖ **Research Claims**

- Frequency Claims
- Association Claims
- Causal Claims
- Knowledge Check
- Lab

#### ❖ **Measurement**

- Survey Design and Measurement
- Reliability and Validity







- Knowledge Check
- Lab

### ❖ **Correlational and Experimental Design**

- Bivariate and Multivariate Designs
- Between and Within Groups Experimental Designs
- Factorial Designs
- Knowledge Check
- Lab

### ❖ **Final Exam**

## ➤ **Option 2: Data Science Research Methods: Python Edition**

**Duration:** 18 hrs

**Course Content:**

### ❖ **Before You Start**

- Start here

### ❖ **Introduction to Machine Learning**

- High Level Data Science Process
- Overview of Machine Learning
- Lab

### ❖ **Exploring Data**

- Exploratory Data Analysis for Regression
- Exploratory Data Analysis for Classification
- Lab

### ❖ **Cleaning and Preparing Data**

- Data Preparation and Cleaning
- Feature Engineering
- Lab

### ❖ **Getting Started with Supervised Learning**

- Regression
- Classification
- Lab

### ❖ **Improving Model Performance**

- Principles of Model Improvement





- Techniques for Improving Models
- Dimensionality Reduction
- Lab

#### ❖ **Machine Learning Algorithms**

- Introduction to Decision Trees
- Ensemble Methods
- Neural Networks
- Support Vector Machines (SVMs)
- Bayes Theorem
  - Lab

#### ❖ **Unsupervised Learning**

- Clustering
- Lab





## **[IX] Build Machine Learning Models** **2 TRAINING OPTIONS AVAILABLE**

### ➤ **Option 1: Principles of Machine Learning: R Edition**

**Duration:**48 Hrs

#### ❖ **Course Content:**

- Before You Start
- Start Here

#### ❖ **Introduction to Machine Learning**

- High Level Data Science Process
- Overview of Machine Learning
- Lab

#### ❖ **Exploring Data**

- Exploratory Data Analysis for Regression
- Exploratory Data Analysis for Classification
- Lab

#### ❖ **Cleaning and Preparing Data**

- Data Preparation and Cleaning
- Feature Engineering
- Lab

#### ❖ **Getting Started with Supervised Learning**

- Regression
- Classification
- Lab





### ❖ **Improving Model Performance**

- Principles of Model Improvement
- Techniques for Improving Models
- Dimensionality Reduction
- Lab

### ❖ **Machine Learning Algorithms**

- Introduction to Decision Trees
- Ensemble Methods
- Neural Networks
- Support Vector Machines (SVMs)
- Bayes Theorem
- Lab

### ❖ **Unsupervised Learning**

- Clustering
- Lab

### ❖ **Final Exam and Survey**

- Final Challenge

## ➤ **Option 2: Principles of Machine Learning: Python Edition**

**Duration:**48Hrs

**Course Content:**

### ❖ **Before You Start**

- Start Here

### ❖ **Introduction to Machine Learning**

- High Level Data Science Process
- Overview of Machine Learning
- Lab

### ❖ **Exploring Data**

- Exploratory Data Analysis for Regression
- Exploratory Data Analysis for Classification
- Lab

### ❖ **Cleaning and Preparing Data**

- Data Preparation and Cleaning
- Feature Engineering
- Lab





- ❖ **Getting Started with Supervised Learning**
  - Regression
  - Classification
  - Lab
- ❖ **Improving Model Performance**
  - Principles of Model Improvement
  - Techniques for Improving Models
  - Dimensionality Reduction
  - Lab
- ❖ **Machine Learning Algorithms**
  - Introduction to Decision Trees
  - Ensemble Methods
  - Neural Networks
  - Support Vector Machines (SVMs)
  - Bayes Theorem
  - Lab
- ❖ **Unsupervised Learning**
  - Clustering
  - Lab
- ❖ **Final Exam and Survey**





## **[X] Build Predictive Solutions at Scale** **3 TRAINING OPTIONS AVAILABLE**

### ➤ **Option 1: Developing Big Data Solutions with Azure Machine Learning**

**Duration:** 16Hrs

**Course content:**

#### ❖ **Introduction**

- Before You Start

#### ❖ **Introduction to Azure Machine Learning**

- Getting Started with Azure Machine Learning Studio
- Working With Big Data Sources
- Lab: Getting Started with Azure Machine Learning
- Module Review

#### ❖ **Building Predictive Models with Azure Machine Learning**

- Introduction to Machine Learning
- Clustering and Recommenders
- Lab: Building Predictive Models
- Module Review

#### ❖ **Operationalizing Machine Learning Models**

- Predictive Experiments and Web Services
- Working with Web Services
- Lab: Publishing Predictive Web Services





- Module Review
- ❖ **Using Azure Machine Learning in Big Data Solutions**
  - Using Azure Machine Learning in Batch Processes
  - Using Azure Machine Learning in Streaming Processes
  - Lab: Building Predictive Big Data Solutions
  - Module Review
- ❖ **Final Exam**
  - Final Exam

➤ **Option 2: Analyzing Big Data with Microsoft R**

**Duration:** 16Hrs

**Course content:**

- ❖ **Getting Started**
- ❖ **Introduction**
  - Introduction
- ❖ **Reading and Preparing Data**
  - Reading the Data
  - Preparing the Data
  - LAB
- ❖ **Examining and Visualizing Data**
  - Examining the Data
  - Visualizing the Data
  - LAB
- ❖ **Clustering and Modeling**
  - Clustering
  - Predictive Modelling
  - LAB
- ❖ **Deploying and Scaling**
  - Deploying and Scaling
- ❖ **Final Exam and Wrap-up**
  - Final Exam





## ➤ **Option 3: Implementing Predictive Analytics with Spark in Azure HDInsight**

**Duration:**24Hrs

**Course Content:**

### ❖ **Course Introduction**

- Introduction
- Pre-Course Survey

### ❖ **Introduction to Data Science with Spark**

- Getting Started with Spark
- Exploring Data with Spark
- Further Reading
- Lab
- Review

### ❖ **Getting Started with Machine Learning**

- Introduction to Machine Learning in Spark
- Pipelines and Text Analysis
- Further Reading
- Lab
- Review

### ❖ **Evaluating and Optimizing Machine Learning Models**

- Evaluating Machine Learning Models
- Optimizing Model Parameters
- Further Reading
- Lab
- Review

### ❖ **Recommenders and Unsupervised Models**

- Recommenders
- Clustering
- Further Reading
- Lab
- Review

### ❖ **Final Exam and Post-Course Survey**

- Final Exam





