
THE ELEMENTS OF DATA SCIENCE

(FOUNDATION STAGE)

Course Description

As we are aware that Data Science is a new form of Statistics associated with more alternatives for processing data. As we may know that data science deals with structured as well as unstructured data with a view to understanding or making predictions about facts, events, and other concerned projects. It is accepted as a multidisciplinary field comprising mathematics, computer science, information science, economics as well as statistics itself.

This course aims at providing an overview of application of R programming language in Data Science to cover a broad selection of key challenges in and methodologies for working with big data. Topics to be covered are extensively detailed for everyone to see. Real life applications with industrial illustrations are given to enhance participants' understanding of the course. This introductory course is integrative across the core disciplines of Data Science, including data visualization and graphics, data wrangling, data mining, statistics, high performance computing and business intelligence. Participants will surely acquire a working knowledge of data science through hands-on projects and case studies in a variety of businesses, science and engineering, social and management sciences.

Course Objectives

After taking this course, participants will be able to:

- (i) Write simple R programmes and perform basic statistical operations visa-a-viz data science
- (ii) Structure, transform, input (output), and visualizing simple data with R programming skills
- (iii) Collect and collate, manipulate and blend data from different data sources
- (iv) Develop various economic and business analytic transformations using R knowledge
- (v) Develop data products for business intelligence and applications
- (vi) Get acquitted with probability as well as applied general statistical terms.

Course Approach

Though this is an online class designed to learn the fundamental concepts of data science, lectures could still be held physically, depending on participant's request. The course runs for at least **8 weeks and at most 12 weeks**, depending on participant's level of assimilation. Each participant is allowed to **decide convenient time**. A laptop installed with R software package is required; we are ready to assist to get this done. Internet facility is a must, whether the lecture is online or physical. Each lecture is expected to take 90 – 120 minutes with at least 3 sessions per week. Participants should be reasonably proficient in English but not necessarily need to have a statistical background. Participants will be mentored to cope with the theoretical aspects required.

Target Audience

This course is targeted at program managers, monitoring and evaluation officers, policy makers, NGO staff, development partners, researchers, lecturers, statisticians, actuarial scientists, data scientists, data engineers, data analyst, economists, econometricians, bank staff, and administrative officers.

Course Fee

The fee is seventy thousand naira (N70,000) for this *Foundation Class* apart from the payment of five thousand naira (N5,000) for the application form. Participants may be, at the discretion of the management of Tim-R, allowed to make the payment available twice – two installments.

Course Certification

Upon successful completion of the short course, participants will be issued with a certificate of competency from our registered academic training institute: Tim-R Programming Consult, Nigeria.

Course Communication

We will be using zoom and/or WhatsApp platform to teach the participants. Our online platform will be 24-hour active. Participants may choose either WhatsApp channel – 2348068998580 or email channel – thompsondx@gmail.com or info@tim-rprogramming.com as means of communication. Kindly post your request via any of the convenient means as stated.

COURSE SYNOPSIS

Module	Major Outline	Description of the Contents	Remarks
1.	Introduction and Some Basics	Installation and Preamble, Some Basic Definitions, Some Mathematical Functions and Operations of data science concepts using R language, Setting Variables, Listing Variables, Deleting Variables, Creating a Vector, Computing some Basics, Creating Sequences, Comparing Vectors, Selecting Vector Elements, Performing Vector Arithmetic, Defining a Function, Preliminaries of Descriptive Statistics	E-Class Practical with R
2.	Navigating the Software with Industrial Applications	Getting and Setting the Working Directory, Saving Your Workspace, Viewing Your Command History, Saving the Result of the Previous Command, Displaying the Search Path, Accessing the Functions in a Package, Accessing Built-in Datasets, Viewing the List of Installed Packages, Installing Packages from CRAN, Setting a Default CRAN Mirror, Suppressing the Startup Message, Running a Script, Running a Batch Script, Getting and Setting Environment Variables, Locating the R Home Directory.	E-Class Practical with R
3.	Input and Output of Data with Applications in Data Science	Entering Data from the Keyboard, Printing Fewer Digits (or More Digits), Redirecting Output to a File, Listing Files, Dealing with 'Cannot Open File' in Windows, Reading Fixed-Width Records, Reading Tabular Data Files, Reading from CSV Files, Writing to CSV Files, Business Intelligence, Data Wrangling and Warehousing	E-Class Practical with R
4.	Data Structures	Appending Data to a Vector, Inserting Data into a Vector, Understanding the Recycling Rule, Creating a Factor (Categorical Variable), Combining Multiple Vectors into One Vector and a Factor, Creating a List, Selecting List Elements by Position, Selecting List Elements by Name, Building a Name/Value Association List, Removing an Element from a List, Flatten a List into a Vector, Removing NULL Elements from a List, Removing List Elements Using a Condition, Initializing a Matrix, Performing Matrix Operations, Giving Descriptive Names to the Rows and Columns of a Matrix, Selecting One Row or Column from a Matrix, Initializing a Data Frame from Column Data, Initializing a Data Frame from Row Data, Appending Rows to a Data Frame, Preallocating a Data Frame, Selecting Data Frame Columns by Position, Selecting Data Frame Columns by Name, Selecting Rows and Columns More Easily, Changing the Names of Data Frame Columns, Editing a Data Frame, Removing NAs from a Data Frame, Excluding Columns by Name, Combining Two Data Frames, Merging Data Frames by Common Column.	E-Class Practical with R

COURSE SYNOPSIS

Module	Major Outline	Description of the Contents	Remarks
5.	Data Transformations and Simple Descriptive Statistics	Splitting a Vector into Groups, Applying a Function to Each List Element, Applying a Function to Every Row, Applying a Function to Every Column, Applying a Function to Groups of Data, Applying a Function to Groups of Rows, Applying a Function to Parallel Vectors or Lists, Measure of Central Tendency, Measure of Variation.	E-Class Practical with R
6.	Strings and Dates	Getting the length of a String, Concatenating a String, Extracting Substrings, Splitting a String according to a Delimiter, Replacing Substrings, Seeing the special character in a String, Generating all pairwise combinations of Strings, Getting the current Dates, Converting a String into a Date, Converting a Date into a String, Converting year, month, and day into a date, Getting the Julian Date, Extracting the parts of a date, Getting a sequence of dates.	E-Class Practical with R
7.	Probability and its Concepts with Application to Data Science	Counting the Number of Combinations, Generating Combinations, Generating Random Numbers, Generating Reproducible Random Numbers, Generating a Random Sample, Generating Random Sequences, Randomly Permuting a Vector, Calculating Probabilities for Discrete Distributions, Calculating Probabilities for Continuous Distributions, Converting Probabilities to Quantiles, Plotting a Density Function, etcetera.	E-Class Practical with R
8.	Applied General Statistics	Summarizing Your Data, Calculating Relative Frequencies, Tabulating Factors and Creating Contingency Tables, Testing Categorical Variables for Independence, Calculating Quantiles (and Quartiles) of a Dataset, Inverting a Quantile, Converting Data to Z-Scores, Testing the Mean of a Sample (t Test), Forming a Confidence Interval for a Mean, Forming a Confidence Interval for a Median, Testing a Sample Proportion, Forming a Confidence Interval for a Proportion, Testing for Normality, Testing for Runs, Comparing the Means of Two Samples, Comparing the Locations of Two Samples Non-parametrically, Testing a Correlation for Significance, etc	E-Class Practical with R

Sincerely yours,



Timothy A. OGUNLEYE

Executive Director

Webpage: <https://www.tim-rprogramming.com>

GSM Number: +2348068998580

Reg. Number: 3286968

