YH PROVIDER
CONSULTING

Introduction to Data Analytics Using R. Two weeks

Course Description

Data Analytics is no longer a luxury; the understanding of it is necessity to improve business efficiency and understand client need and trends. Businesses today must collect, analyze, and distribute their data to keep competitive in the digital age. Data scientists know how to use their skills in statistics, math, programming, and other related subjects to transform data in valuable information for decision making. Then, they apply their knowledge to uncover solutions hidden in the data to take on business challenges and goals.

During this training participants will practice with real data using the programming language and free software R and the R Studio interface. This two-week training covers experimental design, an introduction to R programming, data acquisition and cleaning, exploratory data analysis, probability distributions, regression models, as well as how to build an application using R studio.

The course will be supplemented by practical class examples, group exercises and interactive group discussion designed to consolidate and reinforce learning, identify and offering solutions to specific problems associated with Data Analytics Using R.

Who Should Attend?

Data science teams have people from diverse backgrounds like chemical engineering, physics, economics, statistics, mathematics, operations research, computer science, etc. This course is designed for any professionals that want to have a better understanding of the applications of data science to improve the decision-making process and efficiency of an organization.

What You Will Gain:

- Understand the main terms and applications of data science
- Understand the data types and
- Learn the workflow and main steps to clean and prepare data for analysis
- Learn basic R programming commands
- Learn how to use the R Studio interface to run scripts, plot data and prepare reports
- Learn to perform exploratory data analysis
- Understand the types of probability distributions and its applications
- Learn how to perform linear regression with R
- Learn how to build an Application using R studio

Training Methodology

The training will combine lectures (30%) with interactive class exercises and case studies (50%), case studies and general discussions (20%).

Course Content

Introduction of Data Analytics and its Applications

- What is data science?
- What is data scientist?
- Definition of basic terms used in data science
- Data Science Applications
- Data analysis and decision making

R Studio Overview

- Introduction to R and R Studio software
- Installing R Studio
- RStudio's main interface
- Review of main R functionalities
- R packages
- R Scripts
- R Markdown overview
- R Markdown examples

Experimental Design

- Experimental design concepts
- Principles of experimental design
- Dependent, Independent, and confounding variables

Introduction to R programming

- Overview of f as a programming language
- Overview of writing codes in R
- R data types: attributes, lists, vectors, matrices, factors, data frames
- Reading data
- Review of most useful R commands
    - Input and display
    - Distributions
    - Data manipulation
    - Statistics and transformations
    - More statistics: Regression and Linear model

    - Graphics
    - Distributions
- Control structures; if, for, while
- Overview of Functions, and their use
- Overview of dates and time in R

Data Acquisition and Cleaning

- Objective of gathering and cleaning data for analysis
- Raw and processed data
- Types of data
- Structured and unstructured data
- Tidy data
- Summarizing data
- Removing NA values

Exploratory Data Analysis

- What is the objective of exploratory data analysis?
- Exploratory graphs
- Box plots
- Histogram
- Barplot
- Scatter plots
- Lattice Plot
- Simple base graphic system
- Exercise 1
- Exercise 2

Probability Distributions

- Type of Probability distributions and its applications
- Understanding probability distributions in R
- Generating random numbers
- Normal distribution properties
- Normal distribution functions
- Exercise 3
- Exercise 4

Regression models

- Introduction regression models
- Standard deviation and mean
- Normalization
- Least square estimation of regression lines

- Linear regression models
- Exercise 5
- Exercise 6

How to build an Application using R studio

- Overview of shiny web
- Shiny web generation workflow
- Shiny web examples
- Shiny web Project