DATA SCIENCE USING R-PROGRAMMING
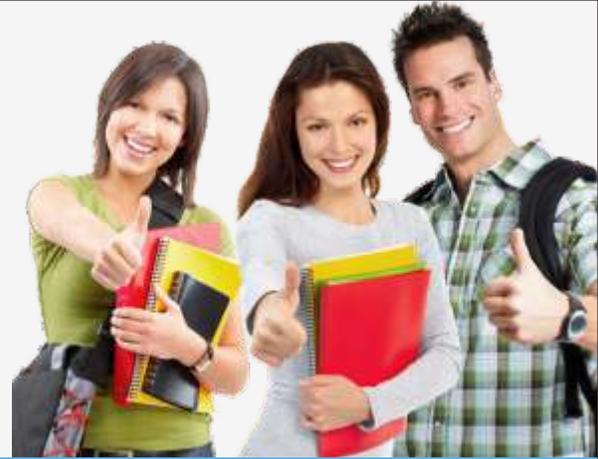
CURRICULUM

FUNDAMENTAL OF STATISTICS.

- Population and sample
- Descriptive and Inferential Statistics
- Statistical data analysis
- Variables
- Sample and Population Distributions
- Interquartile range
- Central Tendency
- Normal Distribution
- Skewness.
- Boxplot
- Five Number Summary
- Standard deviation
- Standard Error
- Emperical Formula
- central limit theorem
- Estimation
- Confidence interval
- Hypothesis testing
- p-value
- Scatterplot and correlation coefficient
- Standard Error
- Scales of Measurements and Data Types
- Data Summarization
- Visual Summarization
- Numerical Summarization
- Outliers & Summary

Module 1- Introduction to Data Analytics

Objectives:

- This module introduces you to some of the important keywords in R like Business Intelligence, Business
- Analytics, Data and Information. You can also learn how R can play an important role in solving complex analytical problems.
- This module tells you what is R and how it is used by the giants like Google, Facebook, etc.
- Also, you will learn use of 'R' in the industry, this module also helps you compare R with other software
- in analytics, install R and its packages.

Topics:

Business Analytics, Data, Information

- Understanding Business Analytics and R
- Compare R with other software in analytics
- Install R
- Perform basic operations in R using command line

Module 2- Introduction to R programming
Starting and quitting R

Recording your work
- Basic features of R.
- Calculating with R
- Named storage
- Functions
- R is case-sensitive
- Listing the objects in the workspace
- Vectors
- Extracting elements from vectors
- Vector arithmetic
- Simple patterned vectors
- Missing values and other special values
- Character vectors Factors
- More on extracting elements from vectors
- Matrices and arrays
- Data frames
- Dates and times

NOTE:-
- Assignments with Datasets

Import and Export data in R
- Importing data in to R
- CSV File
- Excel File
- Import data from text table

Topics
- Variables in R
- Scalars
- Vectors
- R Matrices
- List
- R – Data Frames
- Using c, Cbind, Rbind, attach and detach functions in R
- R – Factors
- R – CSV Files
- R – Excel File

NOTE-:
- Assignments
- Business Scenerio/Group Discussion.

R Nuts and Bolts-:
- Entering Input. – Evaluation- R Objects- Numbers- Attributes- Creating Vectors- Mixing Objects- Explicit Coercion- Summary- Names- Data Frames.

Module 3- Managing Data Frames with the dplyr package
- The dplyr Package
- Installing the dplyr package
- select()
- filter()
- arrange()
- rename()
- mutate()
- group_by()
- %>%

NOTE-:
- Assignments

- Business Scenerio/Group Discussion.

Module 4- Loop Functions
- Looping on the Command Line
- lapply()
- sapply()
- tapply()
- apply()

NOTE-:
- Assignments
- Business Scenerio/Group Discussion.

Module 5- Data Manipulation in R Objectives:
- In this module, we start with a sample of a dirty data set and perform Data Cleaning on it, resulting in a data set, which is ready for any analysis.
- Thus using and exploring the popular functions required to clean data in R.

Topics
- Data sorting
- Find and remove duplicates record
- Cleaning data
- Merging data

Statistical Plotting-:
- Bar charts and dot charts
- Pie charts
- Histograms
- Box plots
- Scatterplots
- QQ plots

NOTE:-
- Assignments with Datasets

Objectives:
- Control Structure Programming with R
- The for() loop
- The if() statement
- The while() loop
- The repeat loop, and the break and next statements
- Apply
- Sapply
- Lapply

NOTE:-
- Assignments with Datasets

Factors
- Using Factors
- Manipulating Factors
- Numeric Factors
- Creating Factors from Continuous Variables
- Convert the variables in factors or in others.

Reshaping
- Data Modifying
- Data Frame Variables
- Recoding Variables
- The recode Function
- Reshaping Data Frames
- The reshape Package

NOTE:-
- Assignments with Datasets

Module 6- Statistical Learning-:
- What Is Statistical Learning?
- Why Estimate f?
- How Do We Estimate f?
- The Trade-Off Between Prediction Accuracy and Model Interpretability

- Supervised Versus Unsupervised Learning
- Regression Versus Classification Problems
- Assessing Model Accuracy

Module 7- Basics of Statistics & Linear & Multiple Regression
- This module touches the base of Descriptive and Inferential Statistics and Probabilities & 'Regression Techniques'.
- Linear and logistic regression is explained from the basics with the examples and it is implemented in R using two case studies dedicated to each type of Regression discussed.
- Assessing the Accuracy of the Coefficient Estimates.

- Assessing the Accuracy of the Model.
- Estimating the Regression Coefficients.

- Some Important Questions
- Lab: Linear Regression.
    - Libraries .
    - Simple Linear Regression
    - Multiple Linear Regression
    - Interaction Terms
    - Qualitative Predictors
    - Writing Functions

NOTE-:
- Assignments with Different Datasets.
- Business Scenerio/Group Discussion

Module 8- Classification-:
- An Overview of Classification.
- Why Not Linear Regression?
- Logistic Regression
- The Logistic Model
- Estimating the Regression Coefficients

- Making Predictions
- Logistic Regression for >2 Response Classes
- Lab: Logistic Regression.
- The Stock Market Data
- Logistic Regression

NOTE-:
- Assignments with Different Datasets.
- Business Scenerio/Group Discussion.

Module 9- Variance Inflation Factor-:
- Introduction
- Multicolinearity.
- How we can detect the multicolinearity.
- Effects of multicolinearity
- Lab: VIF
- Applications.
- Reduce the features.

NOTE-:
- Assignments with Different Datasets.
- Business Scenerio/Group Discussion.

Correlation
- Types of Correlation
- Properties of Correlation
- Methods of Calculating Correlation

Module 10- Best Model Selection-:
- Subset Selection
- Best Subset Selection
- Stepwise Selection
- Choosing the Optimal Model
- Lab 1: Subset Selection Methods
- Best Subset Selection
- Forward and Backward Stepwise Selection
- Choosing Among Models Using the Validation Set Approach and Cross-Validation

NOTE-:
- Assignments with Different Datasets.
- Business Scenerio/Group Discussion.

Module 11- Tree-Based Methods-:
- The Basics of Decision Trees
- Regression Trees
- Classification Trees
- Trees Versus Linear Models
- Advantages and Disadvantages of Trees
- Bagging, Random Forests, Boosting
- Bagging
- Random Forests
- Lab: Decision Trees
- Fitting Classification Trees
- Fitting Regression Trees .

NOTE-:
- Assignments with Different Datasets.
- Business Scenerio/Group Discussion.

Module 15- Time Series & Forcasting-:
- Time series
- Estimating and Eliminating the Deterministic Components if they are present in the Model.
- Estimating and Eliminating Seasonality if it is present in the Model
- Modeling the Remainder using Auto Regressive Moving Average (ARMA) Models
- Identify 'order' of the ARMA model
- 'Forecast' or Predict for Future Values
- Practise on R

NOTE-:
- Assignments with Different Datasets.
- Business Scenerio/Group Discussion.

Partners : ORACLE WORKFORCE DEVELOPMENT PROGRAM | redhat. CERTIFIED TRAINING PARTNER | UBI | edexcel | Java

---