Name - Data Analytics using R

DAY 1:
DATA ANALYTICS - INTRODUCTION

· What is Data Science
· What is Machine Learning

· Machine Learning vs. Data Science vs. AI
· How leading companies are harnessing the power of Data Science with Python?
· Different phases of a typical Analytics/Data Science projects and role of python
· Anaconda vs. Python
· Machine Learning flow to code
· Regression vs. Classification
· Features, Labels, Classes
· Supervised Learning, Semi-Supervised and Unsupervised Learning
· Cost Function and Optimizers
Introduction to R Programming
· Installation and Setup
· Installing R
· Installing RStudio
· Installing Packages
· Working with Vectors
· Vectors
· Random Numbers, Rounding, and Binning
· Missing Values
· The which() Operator
· R Essentials
· Set Operations
· Sampling and Sorting
· Check Conditions
· For Loops
· Dataframes and Matrices
· Importing and Exporting Data

· Matrices and Frequency Tables
· Merging Dataframes
· Aggregation
· Melting and Cross Tabulations with dcast()
· Core Programming

· String Manipulation
· Functions
· Debugging and Error Handling
· Fast Loops with apply()
· Fast Loops with sapply(), lapply() and vapply()

Statistical Inference
· Normal Distribution, Central Limit Theorem, and Confidence Intervals
· Skewness in data
· Correlation and Covariance
· ANOVA
· Statistical Tests – F Test, T-Test
· DPlyR and Caret Packages.
· Aggregation and Special Functions
· Understanding Syntax, Creating and Updating Columns
· Chaining, Functions, and .SD
· Fast Loops with set(), Keys, and Joins
ACCESSING/IMPORTING AND EXPORTING DATA USING R PACKAGES
· Importing Data from various sources (CSV, txt, excel, access etc)
· Database Input (Connecting to database)
· Viewing Data objects - subsetting, methods

· Exporting Data to various formats
DATA MANIPULATION – CLEANSING
· Cleansing Data with R Programming
· Data Manipulation steps(Sorting, filtering, duplicates, merging, appending, subsetting, derived variables, sampling, Data type conversions, renaming, formatting etc)
· Data manipulation tools(Operators, Functions, Packages, control structures, Loops, arrays etc)
· Scaling and Normalizing data
· Pre-processing and Formatting data
· Feature selection – Correlation, P Values, Multi-Collinearity etc.
DATA ANALYSIS – VISUALIZATION USING R
· Introduction exploratory data analysis
· Basic Plots Vs. GGPLOT Library
· Making Plots with Base Graphics
· Drawing Plots with 2 Y Axes
· Multiplots and Custom Layouts
· Creating Basic Graph Types
· Creating graphs using GGPLOT.
· Descriptive statistics, Frequency Tables and summarization

· Univariate Analysis (Distribution of data & Graphical Analysis)
· Bivariate Analysis(Cross Tabs, Distributions & Relationships, Graphical Analysis)
· Creating Graphs- Bar/pie/line chart/histogram/ boxplot/ scatter/ densityplot etc)
DAY 2
BASIC STATISTICS & IMPLEMENT

REGRESSION ANALYSIS

· Overview
· Introduction to Regression Analysis
· Types of Regression Analysis Models
· Linear Regression
· Model
· Model statistics
· Gradient Descent Algorithm
· Demo: Simple Linear Regression
· Demo: Regression Analysis with Multiple Variables
· Cross Validation
· Factor Analysis
· Fitting model and Predictions
CLASSIFICATION ANALYSIS
· Decision Tree Classification
· Entropy & Gini Index
· Classification and Regression Trees
· Decision Tree Statistics
· Decision Tree
· Demo: Decision Tree Classification
· Random Forest Classification
· Evaluating Classifier Models
· K-Fold Cross Validation
CLUSTERING
· Overview
· Introduction to Clustering

· Clustering Example
· Clustering Methods: Prototype Based Clustering
· Centroids and Means
· Eucledian Distance Formula
· Elbow Method – Picking values of K
· Demo: K-means Clustering

Time Series
· Forecasting
· ARIMA
TABLEU
· Introduction
· Importing various files - Excel Data, CSV Files etc.
· Coverting data into Visualization
· Data insights / Study of Data using graphical representation.