

# CF – Data Science Internship + Training Curriculum

---

## MODULE 1 – Orientation & Introduction to Data Science

**Kartikay Dev Sharma, India**

- Orientation Video with Afshan Meher and Kaustabh Raikar
- What is Data Science? What does a data scientists do?
- The Need for data science / Problems solved with Data Science.
- Difference between AI, ML and DL
- Job Roles
- Salary / Scope

## MODULE 2 – Python Programming

**Kaustabh Raikar, India**

- The Companies using Python. Popular Applications where Python is used.
- Python Environment Setup (Jupyter & Anaconda).
- **Fundamentals:** Variables, Numbers and Boolean, Strings, Arithmetic Operators, The Double Equality Sign, Reassign Values, Add Comments, Line Continuation, Indexing Elements, Structure Your Code with Indentation, Comparison Operators, Logical and Identity Operators.
- **Data Types:** Immutable – [Numbers, Strings, Tuples]; Mutable – [Lists, Dictionaries, Sets] and related operations.
- **Loops:** For Loops (break, continue, pass) & While Loops
- **Functions:** creating and calling functions, lambda, filter, map, in-built functions, user defined functions
- **Condition:** if, if-else, nested if-else, else-if
- **Library:** What is library, what is package, how to create packages, Introduction to PIP, Namespace, Using Python Packages, Installing Packages via PIP
- An intro to Tkinter – Python GUI Library
- An overview of the Spyder IDE

## MODULE 3 – Computing with NumPy and Pandas

Jibril Isimbabi, Kenya

- Mathematical computing with Python (NumPy)
- **Array:** Data Types in an Array, Dimensions of an Array, Operations on Array: Indexing, Slicing, Splicing, Sub-setting
- **Pandas:** Pandas Setup & Intro, Loading Data into Pandas, Reading Data, Sorting Values, Adding or Removing Columns, Rearranging Columns, Saving Data, Filtering Data, Filtering Based on Textual Patterns, GroupBy, Chunksize, DataFrame, Panel, Reindexing, Iteration, Sorting, Working with Text Data, Merging/Joining, Concatenation, Date Functionality, Categorical Data

## MODULE 4 – R Programming for Statistics & Data Science

Harsh Kr. Choudhary, India

- **Introduction:** Introduction of R, R Installation, Basic Structure of R program with a simple program, Constants, variables and declarations, Simple Input-Output statements, Operators – Arithmetic, Relational and Logical
- **Data Types in R:** What are datatypes in R?, Types of data types in R, Numeric Datatypes, Integer Datatypes, Logical Datatypes, Complex Datatypes, Character Datatypes, Vectors Datatypes, List Datatypes, Matrices Datatypes, Arrays Datatypes, Factors Datatypes, Data Frames in R
- **Variables in R:** Variables Definition, Declaring and Initialising variables, Syntax of declaring variables, Important method for variables, Scope of variables, Dynamic scoping in R Variables, Lexical scoping in R variables, Data type of a variable, Finding variables, Deleting variables.
- **Control Flow:** Control statement in R program, What are control statements, If conditions, If-else conditions, For Loop, Nested Loop, While loop, Repeat loop and break statements, Return statements, Next statement, Decision making in R prog with flow chart, Switch case in R, For loop in R, While loop in R, Repeat loop in R, Go to statements in R, Break and next statements in R
- **Functions:** Function Definition, Function components, Built-In functions, User defined functions, Single input- single output, Multiple Input – Multiple Output, Inline function, Passing arguments of a function, Lazy evaluation of a function, Function argument in R programming, Adding an argument in R, Adding multiple arguments in R, Adding default arguments in r, Dots arguments, Function as an argument, Type of function in R programming, How to define a function?, Calling a function, Types of function, Primitive function,

Infix function, Replacement function, Recursive function, Application of recursive function in R programming, Conversion of function.

- **Data Structures:** What are data structures in R programming?, R-Strings, R-Vectors, R-List, R-arrays, R-Matrices, R-Factors, R-Data frames
- **R- Object Oriented Programming:** Classes and Objects, Creating S3 class, Generic function, Attributes, Classes in R programming, R-Objects, Encapsulation in R programming, Polymorphism in R programming, R-Inheritance, Abstraction in R programming
- **Error Handling:** Handling error in R Programming, Condition handling in R program, Debugging in R program
- **File Handling:** File handling in R Programming, Reading files in R Programming, Writing files in R Programming, Binary files in R Programming
- **Data Interfaces:** Data Handling in R Programming, Importing data in R Programming, Exporting data in R Programming, Working with CSV files in R Programming, Working with XML files in R Programming, Working with EXCEL files in R Programming, Working with JSON files in R Programming, Working with DATABASES in R Programming

### **MODULE 5 - Advanced Statistics and Probability with R** **Harsh Kr. Choudhary, India**

- R-Statistics
- Mean, median and mode in R programming
- Calculate the average, variance and standard deviation in R Programming
- Descriptive analysis in R Programming
- Normal distribution in R Programming
- Binomial distribution in R Programming
- ANOVA test in R Programming
- Co-Variance and Correlation in R Programming
- SKEWNESS and KURTOSIS in R Programming
- Hypothesis testing in R Programming
- Bootstrapping in R Programming
- Time series analysis in R Programming

## **MODULE 6 – Introduction to Machine Learning Algorithms**

**Shrijoy Chowdhury, India**

- **Intro:** What is Machine Learning? How Machine Learning Works? Phases of Machine Learning, Some Applications of Machine Learning, Different Types of Machine Learning, Getting acquainted with Jupyter notebook, Supervised Learning in Depth
- Different Algorithms of Regression & Classification
- Evaluation Metrics of Regression & Classification
- Model Flow in Machine Learning
- Creating a Machine Learning Model
- Classification Model Building
- Classifications Metrics: Confusion matrix, Precision, Recall, F1-Score, ROC & AUC Curve
- Statistics: What is Statistics? Types of Data, Types of Statistical Analysis, Measures of Descriptive Statistics, Measures of Dispersion, Skewness & Kurtosis, Covariance & Correlation, Hypothesis Testing.
- Different Tests: Z-test, T-Test, Chi-Square Test
- Exercise on Different Statistical Tests
- Hands-on exercise in statistics using Jupyter Notebook
- Machine Learning Algorithms: Decision Tree, Ensemble Models, K-Means Clustering
- Using Stack Overflow to resolve bugs and issues.
- Introduction to Artificial Neural Network
- Deep Learning
- Convolutional Neural Network (CNN)
- Getting acquainted with Google Colab
- Recurrent Neural Network (RNN)

# College Finder

## **MODULE 8 – A Precursory Glance Into the World of Machine Learning**

**Sanskriti Rai, India & Shrijoy Choudhary, India**

- What is Machine Learning?
- Types of Machine Learning
- Applications of Machine Learning
- How Machine Learning helps us in Day-to-Day Life

- Major Companies ML Applications as a story.
  - ✚ Google Various Services
  - ✚ NASA
  - ✚ Facebook (Recommendation Engine)
  - ✚ Amazon (Alexa)
  - ✚ IBM (Chess Contest)
  - ✚ Apple (Siri, R.E)
  - ✚ Microsoft (Cortana)
  - ✚ Intel (Open VINO, Hardware)
  - ✚ Tesla
  - ✚ Grammarly
  - ✚ Finance Companies (Fraud detection, Loan Eligibility)
  - ✚ Medical Domain
- Linear Regression
- Logistic Regression

## MODULE 8 - Data Visualization

Anis Suraiya, Malaysia & Harsh Kr. Choudhary, India

- **Data Visualization with R**
  - ✚ Data Visualization Meaning
  - ✚ R-Line Graphs
  - ✚ R-Bar graphs
  - ✚ Histogram
  - ✚ Scatter plots
  - ✚ R-Pie charts
  - ✚ Boxplots
- **Data Visualization with Python**
  - ✚ **Matplotlib:** Data Visualization on Matplotlib, Bar Plot, Histogram Plot, Box Plot, Area Plot, Scatter Plot, Pie Plot
  - ✚ **Seaborn:** Introduction to Seaborn, Matplotlib vs Seaborn, Distribution Plot, Joint Plot, Hexagon Distribution, KDE Plot, Pair Plot, Rug Plot, Styling, Bar Plot, Count Plot, Box Plot, Violin Plot, Strip Plot, Swarm Plot, Palettes, Heatmaps, Cluster Map, Pair Grid, Facet Grid, Regression Plots

## **MODULE 9 – SQL for Data Science**

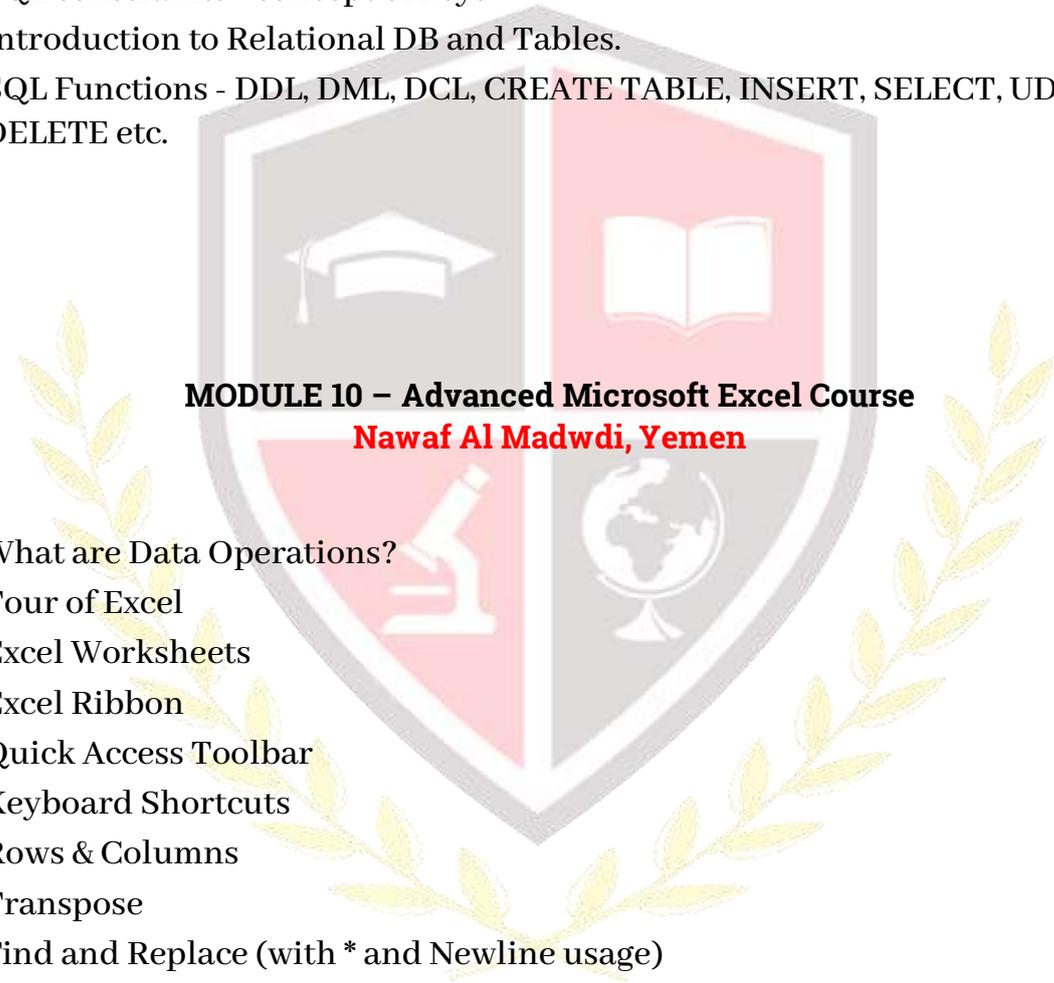
**Jeyalin Sheeju, India**

- Why SQL for Data Science?
- Getting Started with SQL.
- Data types in SQL
- Basic SQL queries + Joins
- SQL constraints - concept of keys
- Introduction to Relational DB and Tables.
- SQL Functions - DDL, DML, DCL, CREATE TABLE, INSERT, SELECT, UPDATE, DELETE etc.

## **MODULE 10 – Advanced Microsoft Excel Course**

**Nawaf Al Madwdi, Yemen**

- What are Data Operations?
- Tour of Excel
- Excel Worksheets
- Excel Ribbon
- Quick Access Toolbar
- Keyboard Shortcuts
- Rows & Columns
- Transpose
- Find and Replace (with \* and Newline usage)
- Formulas
- Excel Functions
- Reorder and Summarize Data
- Sorting & Filtering
- Pivot Tables & Charts
- Combine Data from Multiple Sources
- Sheet Protection
- Print Worksheets



College Finder

**Internship Projects**  
**Aadarsh Lalchandani & Other Course Trainers**

- Programming - Student Portfolio
- Programming - Rock Paper Scissor
- Programming - Multiplication Tables Generator
- Programming - Real Calculator
- Programming - Bulk File Rename
- Programming - News Application with Tkinter
- NumPy & Pandas - Handling Real World Financial Data with Pandas
- Machine Learning - Music Preference Predictor
- Machine Learning - Car Price Prediction using Regression Model
- Machine Learning - Credit Card Defaulters prediction using ANN
- Data Visualization - Covid Data Analysis with Matplotlib
- SQL – Five Real Life Corporate Assignments
- Python & SQL – Own Database Management System
- MS Excel – Colleges Data Cleaning & Entry using MS Excel

---

The curriculum is subject to minor changes.

**College Finder**