Data Science

1. Tools and Languages for Data Science: Python, R, Hadoop, Spark, SQL, Tableau
2. Total Duration: 200 Hours (approx. 5 months according to 2 hours daily (5 day in a week)

   Approx. 45 days according 5 hour in a day (6 day in a week)
3. Overview:-

   - Core Python
   - Algorithms & Data Structures
   - Advance Python
   - Web Application (Flask & Django)
   - Web Scrapping
   - R Language
   - DBMS (SQL & No-SQL)
   - Statistics
   - Basic Linux
   - Data Analysis
   - Data Science
   - AWS Integration
   - Big-Data Hadoop
   - Data Visualization
   - Machine Learning

Programming with Python

Core Python: Introduction to Python, Basic Syntax, Data Types, Variables, Operators, Input/output, Flow of Control (Modules, Branching), If, If- else, Nested if-else, Looping, For, While, Nested loops, Control Structure, Break, Continue, Pass, Strings and Tuples, Accessing Strings, Basic Operations, String slices, Working with Lists, Introduction, Accessing list, assign and retrieve

values from Lists, Introducing Tuples, Accessing tuples Operations, Function and Methods, Files, Modules, Dictionaries, Advance Dictionaries, Functions and Functional Programming, Declaring and calling Functions, Special Functions in python lambda, map and reduce. Advance functions in python var length arguments and Closures and Decorators. Namespace and Generator and Iterators.

Advanced Python: Object Oriented, OOPs concept, Class and object, Attributes, Inheritance, Overloading, Overriding, Data hiding, Meta Classes, Shared Memory concepts, Exception Handling, except clause, Try finally clause, User Defined Exceptions, Debugging modules pdb, doctest and loggers. Python Libraries- NUMPY, SCIPY, PANDAS, Scikit-Learn, matplotlib, bs4 etc.

Projects: Bank Application, Quiz Maker, Chat Application, Data Entry and Processing projects, Pattern Finder using regx, text editor.
_____

Statistical Analysis

Probability & Statistics: Introduction to Statistics- Descriptive Statistics, Summary Statistics Basic probability theory, Statistical Concepts (uni-variate and bi-variate sampling, distributions, re-sampling, statistical Inference, prediction error), Probability Distribution (Continuous and discrete- Normal, Bernoulli, Binomial, Negative Binomial, Geometric and Poisson distribution) , Bayes' Theorem, Central Limit theorem, Data Exploration & preparation, Concepts of Correlation, Regression, Covariance, Outliers etc.
_____

R Programming

R Programming: Introduction & Installation of R, R Basics, Finding Help, Code Editors for R, Command Packages, Manipulating and Processing Data in R, Reading and Getting Data into R, Exporting Data from R, Data Objects-Data Types & Data Structure. Viewing Named Objects, Structure of Data Items, Manipulating and Processing Data in R (Creating, Accessing , Sorting data frames, Extracting, Combining, Merging, reshaping data frames), Control Structures, Functions in R (numeric, character, statistical), working with objects, Viewing Objects within Objects, Constructing Data Objects, Building R Packages, Running and Manipulating Packages, Non parametric Tests- ANOVA, chi-Square, t-Test, U-Test, Introduction to Graphical Analysis, Using Plots(Box

Plots, Scatter plot, Pie Charts, Bar charts, Line Chart), Plotting variables, Designing Special Plots, Simple Liner Regression, Multiple Regression

Project: Market Basket Analysis, Housing Price Predictions, Student Evaluation

_____

Data Collection & DBMS (Principles, Tools& Platforms)

Database Concepts (File System and DBMS), Database Storage Structures (Tablespace, Control files, Data files), Structured and Unstructured data, SQL Commands (DDL, DML & DCL), Data ware Housing concept, No-SQL, Data Models - XML, working with MongoDB).

SQL Programming: Introduction to SQL, Download the dataset, Shortcut to upload the data, SELECT * Statement, Using the WHERE clause to filter data, How to use Wildcards / Regular Expressions in SQL (% and _), Comments in SQL, Order By, Data Types in SQL, implicit Data Conversion in SQL, Using Cast() vs Convert(), Working with NULLs, Understanding how LEFT, RIGHT, INNER, and OUTER joins work, Joins with duplicate values, Joining on multiple fields
Tools - OLTP and OLAP, data preparation and cleaning techniques

Project: Data Entry and Processing, Joins, Creating Views to Predict from data, Automation Scripts for data entry and data manipulation at large datasets.

_____

Linux and cloud computing

Basics of Linux – File Systems, Users, Groups, Permissions, Mounting, Shell Scripts, Ip & Networking, Firewall, Services and Daemons, Apache Server, MariaDB Server, CGI scripts

Cloud Computing – Cloud Computing Concepts, Public and Private Cloud, Introduction to AWS, Azure, Digital Ocean and Google Cloud Services.

_____

Big Data Technologies (Hadoop)

Introduction to Big Data- Big data definition, enterprise / structured data, social / unstructured data, unstructured data needs for analytics, What is Big Data, Big Deal about Big Data, Big Data Sources, Industries using Big Data, Big Data challenges.

Hadoop: Introduction of Big data programming-Hadoop, History of Hadoop, The ecosystem and stack, The Hadoop Distributed File System (HDFS), Components of Hadoop, Design of HDFS, Java interfaces to HDFS, Architecture overview, Development Environment, Hadoop distribution and basic commands, Eclipse development, The HDFS command line and web interfaces, The HDFS Java API (lab), Analysing the Data with Hadoop, Scaling Out, Hadoop event stream processing, complex event processing, MapReduce Introduction, Developing a Map Reduce Application, How Map Reduce Works, The MapReduce Anatomy of a Map Reduce Job run, Failures, Job Scheduling, Shuffle and Sort, Task execution, Map Reduce Types and Formats, Map Reduce Features, Real-World MapReduce.

Hadoop ETL: Hadoop ETL Development, ETL Process in Hadoop, Discussion of ETL functions, Data Extractions, Need of ETL tools, Advantages of ETL tools.

Pig and HIVE- Programming Pig: Engine for executing data flows in parallel on Hadoop, Programming with Hive: Data warehouse system for Hadoop, Optimizing with Combiners and Practitioners (lab), More common algorithms: sorting, indexing and searching (lab), Relational manipulation: map-side and reduce-side joins (lab), evolution, purpose and use, HDFS – Overview and concepts, data flow (read and write), interface to HDFS (HTTP, CLI and Java API), high availability and Name Node federation, Map Reduce developing and deploying programs, optimization techniques, Map Reduce Anatomy, Data flow framework programming Map Reduce best practices and debugging, Introduction to Hadoop ecosystem, integration R with Hadoop.

Hadoop Environment: Setting up a Hadoop Cluster, Cluster specification, Cluster Setup and Installation, Hadoop Configuration, Security in Hadoop, Administering Hadoop, HDFS – Monitoring & Maintenance, Hadoop benchmarks, Hadoop in the cloud.

Introduction to Apache Spark and Use Cases
Apache Spark APIs for large-scale data processing: Overview, Linking with Spark, Initializing Spark, Resilient Distributed Datasets (RDDs), External Datasets, RDD Operations, Passing Functions to Spark, Working with Key-Value Pairs, Shuffle operations, RDD Persistence, Removing Data, Shared Variables, Deploying to a Cluster

Project: Case Studies on Network Intrusion Detection, Sentiment Analysis, Fraud Detection, Twitter Data Analysis
_____

Data Visualization - Analysis and Reporting

Information Visualization, Data analytics Life Cycle, Analytic Processes and Tools, Analysis vs. Reporting, Modern Data Analytic Tools, Visualization Techniques, Visual Encodings, Visualization algorithms, Data collection and binding, Cognitive issues, Interactive visualization, Visualizing big data – structured vs unstructured, Visual Analytics, Geo-mapping, Dashboard Design
Tableau: Introduction, Installing Tableau Desktop and Tableau Public (FREE), Challenge description + view data in file, Connecting Tableau to a Data file - CSV file, Navigating Tableau - Measures and Dimensions, Creating a calculated field, Adding colours, Adding labels and formatting, Exporting your worksheet.
Project : Plotting Graphs for data processing, Comparing two different Datasets Graphically, Plotting Scatter Graphs to find Relations in datasets for predictions.
Data Science Project: Mail Spam Detection, Movie Recommendation System, H1B Visa Analysis, Fraud Detection, Earthquake Prediction, Network Intrusion Detection, Sentimental Analysis.

Machine Learning

Supervise Machine Learning: Introduction to machine learning, Linear Regression, Logistic Regression, Decision Trees, Support Vector Machine, and Recommender Systems.
Un-Supervise Machine Learning: Introduction to Un-Supervise machine Learning, Clustering, K-Means, and Principle Component Analysis (PCA).
_____

GRRAS SOLUTIONS PVT. LTD.

Grras UsP's

When you have so many options to consider, why Grras should be your preferred Training Partner?

Hence, we give you some Neutral advice:

- Certificates from GRRAS Solutions Pvt. Ltd.
- Internship Letter by GRRAS.
- Lifetime membership. (You can join same course again for purpose of revision and update at free of cost at our any centre) in India or you can solve your query by online help.)
- 24x7 our labs Infrastructure and availability.
- 100% Job Assistance for both fresher and Experienced.
- We are authorized & certified partner of Red hat & our entire team & Trainer expertise & are same.
- Exam Scenario based Preparation Included in training.
- From basic to practical based training on system.
- We are so flexible in time adjusting according to candidates.
- Boot camp. For exam preparation along with training.
- We have some unique features with our training programs.
- Three times Award by Red hat for best training partner.

"Be a part of GRRAS family and see the transformation"
THANK YOU FOR MORE INFORMATION VISIT OUR WEBSITE
WWW.GRRAS.COM

Head Office:- 219, Himmat Nagar, Behind Kiran Sweets, Gopalpura Turn, Jaipur (Raj).
Contact No. 9001997178, 8448448523