ANNEXURE -1

Detailed syllabus:

Name of the Group: Data Science

Course Name: Advanced Diploma in BigData Analytics

Course Code: I030

Module No	Topic	Details
1	Linux concepts, Java programming	Linux environment, commands, built-in tools for data analysis. Java programming in Linux-Basic ,Arrays,Class,Object,OOPS concepts ,Collections .
2	BigData concepts, Hadoop and MapReduce	Hadoop Architecture and HDFS, Mapreduce Architecture with examples, YARN Architecture, nosql databases with examples,
3	Hive, Pig, Sqoop, Flume, Hbase,Spark	Transferring data with Sqoop, data ingestion into Hadoop with Flume,Familiarization of Spark, Mllib and machine learning, Oozie, HBase, Hive and HiveQL, Pig, Distributed processing on a Cluster, Integrating R and Hadoop
4	R Programming	Setting up R environment, Variables, Data Types - Vectors, Factors, Lists, Matrices, Arrays, Data Frames, Subsetting. Control Structures, Functions, Debugging tools,. Reading data – Text, CSV, HTML, JSON, MySQL. Grouping functions-apply, lapply, sapply, mapply. Data visualization - barplot, pie, scatterplot, histogram, scatter matrix, ggplot. Statistical Analysis of data-Summary Statistics, Tabulation methods. Probability distributions in R- Normal distribution, Poisson distribution, Binomial distribution. Correlation and Regression, Hypothesis Testing, Graph visualization using igraph, Developing GUI with Shiny.

5	Python Programming	Python -features, program execution, data structures, List, Dictionary, Tuples, If statements, looping and loop control statements, Functions and Modules, Generators, import statement, namespaces-packages, Class concepts, Exception handling, Regular Expressions, Database access, XML parsing, Python for data analytics – using numpy, matplotlib and pandas,scipy, sci-kit learn .
6	Project	Students are required to develop software in any of the areas covered in the course under the guidance of the faculty. Students can also take up relevant projects from the industry.