



CERTIFIED DATA SCIENTIST PROGRAM BROCHURE

YOUR GOAL IS OUR MISSION

Our aim is to equip learners with the skills necessary to pursue successful careers in Analytics, Data Science, and AI

TABLE OF CONTENTS

<u>DATAMITES® ACCOLADES</u>	... 2
<u>WHY DATAMITES®</u>	... 3
<u>PROGRAM STRUCTURE</u>	... 4
<u>REAL-TIME INTERNSHIP</u>	... 5
<u>JOB READY PROGRAM</u>	... 6
<u>PROGRAM CURRICULUM</u>	... 7
<u>ADMISSIONS AND CONTACTS</u>	... 18

DATAMITES® ACCOLADES

10+ Years of Excellence ★ **100K+** Learners ★ **20+** Accreditations



AWARDED GLOBAL TOP 5
DATA SCIENCE INSTITUTES BY IABAC



AWARDED TOP 10
BEST INSTITUTE BY SILICON INDIA

TECHGIG

India's largest tech community

1ST RANKED INSTITUTE

BASED ON THE RESEARCH STUDY BY
TECHGIG, DATAMITES IS RANKED AS THE
TOP INSTITUTE FOR DATA SCIENCE

GCREDO
Global Credentialing Office

1ST RANKED AI INSTITUTE

GLOBAL CREDENTIALLING OFFICE AWARDED
DATAMITES THE 1ST RANKED INSTITUTE FOR AI



Confederation of Indian Industry

CII PARTNER

DATAMITES® CHOSEN AS A PARTNER BY CII
FOR PROVIDING AI TRAINING C-LEVEL
EXECUTIVES MNCs IN INDIA

NASSCOM®

 MINISTRY OF
ELECTRONICS &
INFORMATION TECHNOLOGY
GOVERNMENT OF INDIA

NASSCOM PARTNER

ALIGNING CURRICULUM WITH INDUSTRY
REQUIREMENTS. ASSESS AND CERTIFY
LEARNERS BY GOVT OF INDIA

WHY DATAMITES?

TOP 4 REASONS

1

Curriculum aligned with Industry

Syllabus aligned with industry as per global accreditation bodies



NASSCOM®

2

Ashok Veda as Mentor

Highly respected Data Science coach and AI Expert as lead mentor ensuring quality mentorship



linkedin.com/in/ashokveda/

3

Realtime Internship

Every learner gets Internship in the selected industry with Analytics, Data Science and AI roles for real-time experience, which is valuable in their career progress

4

Top Placement records

A dedicated job assistance team helped thousands of learners in transitioning into their dream job.

[Check out Success Stories](#)



KEY HIGHLIGHTS

1. Flexible Learning

Learners can repeat sessions, change batches , change learning modes, ad-hoc doubts sessions anytime.

2. Job-oriented curriculum

The course curriculum is aligned with Industry requirement by expert content team, ensuring job-oriented curriculum

3. Elite instructors

Elite mentors and faculties members holding real-time experience from leading companies. and from league institutes such as IIMs

4. Exclusive Practice Lab

Learners get exclusive access to AI and Data Science online lab enabling learners to practice the concepts taught in class

5. Learning Community

Exclusive Online learning community with thousands of active learners, mentors and Alumni available for clarifying doubts and mentoring

6. Life-Time Access

Learners have life-time access to core materials supporting continuous learner beyond the course, ensuring continual learning

7. Unlimited Projects

Unlimited projects with flexibility to choose from various industries but a minimum of 5 projects are required to complete projects phase.

8. Placement Assistance

A dedicated placement assistance team will work with the learners to support in career transition. DataMites records highest placements in India.

PROGRAM STRUCTURE

STRUCTURED 3 PHASE LEARNING APPROACH

THE COURSE FOR BEGINNERS' AND INTERMEDIATE LEARNERS IN THE FIELD OF DATA SCIENCE . THIS IS A CAREER-ORIENTED PROGRAM, DESIGNED TO IMPART A STRONG FOUNDATION IN THE DATA SCIENCE CORE KNOWLEDGE AREAS, INCL. PYTHON, STATISTICS, MACHINE LEARNING, VISUAL ANALYTICS, ML MODELING AND AI BASICS.

- ✓ **8** MONTHS PROGRAM
- ✓ **20** HOURS LEARNING A WEEK
- ✓ **300+** LEARNING HOURS
- ✓ **10+** CAPSTONE PROJECTS
- ✓ **1** CLIENT/LIVE PROJECT
- ✓ **GLOBAL CERTIFICATIONS**
- ✓ **INTERNSHIP EXPERIENCE CERTIFICATE**
- ✓ **JOB READY PROGRAM**



- PRE COURSE SELF-STUDY
- HIGH QUALITY VIDEOS WITH EASY LEARNING APPROACH.
- 4-MONTH DURATION
- LIVE TRAINING
- 20 HOUR A WEEK
- COMPREHENSIVE SYLLABUS
- HANDS-ON PROJECTS
- EXPERT TRAINERS AND MENTORS

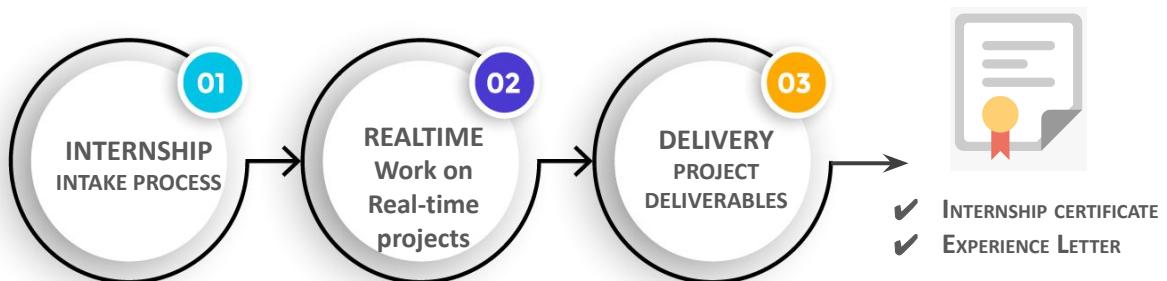
- 4-MONTH DURATION
- PROJECT MENTORING
- 10+ CAPSTONE PROJECTS
- REAL-TIME INTERNSHIP
- 1 CLIENT /LIVE PROJECT

REAL-TIME INTERNSHIP

REAL-WORLD EXPERIENCE IN IMPLEMENTING ML PROJECTS

DataMites has exclusive partnership with leading AI and Data Science companies providing internship for DataMites learners.

These internships provide a great opportunity for the learners to apply the knowledge gained in developing real-world data model that add value to the businesses with help of dedicated team of DataMites experts and Mentors.



internship@datamites.com

PROGRAM CURRICULUM

CERTIFIED DATA SCIENTIST - COURSE BUNDLE

- DataMites® Certified Data Scientist (CDS) – Course Bundle is one the **world's most popular, comprehensive, job-oriented** Data Science and Machine Learning course.
- **25,000+** Alumni Network : CDS course being one of the first courses (arguably) in the Data Science, launched in 2016, has the largest alumni network
- The course is **vigorously updated** as per the industry requirements and fine-tuned to make the learning process structured enabling lean learning.

BUNDLE CODE	CDM-CDS-BUN-021	LEARNING HOURS	300
ADD-ON	Internship, Placements	TOTAL DURATION	8 Months

ORDER	COURSE	CODE	ORDER
1	Data Science Foundation	CDM-DSF-112	20
2	Python Foundation	CDM-PYF-110	40
3	Statistics Essentials	CDM-STA-139	20
4	Machine Learning Associate	CDM-MLA-130	40
5	Machine Learning Expert	CDM-MLE-113	40
6	Advanced Data Science	CDM-ADS-114	40
7	Database: SQL and MongoDB	CDM-DBM-120	40
8	Version Control with Git	CDM-GIT-115	20
9	Big Data Foundation	CDM-BDF-117	20
10	Certified Business Intelligence(BI) Analyst	CDM-BIA-119	20

Important Note: The curriculum is subjected to change as required by the global accreditation bodies to align with industry requirements. Please check with your counsellor or drop email to care@datamites.com for updated curriculum

JOB READY PROGRAM

END TO END SUPPORT IN JOB ASSISTANCE

DEDICATED PLACEMENT ASSISTANCE TEAM (PAT) PROVIDES END TO END ASSISTANCE IN KEY AREAS TO PROVIDE SMOOTH TRANSITION TO ARTIFICIAL INTELLIGENCE CAREER.



Our Top Hiring Partners

DataMites has a strong network of **1000+** hiring partners

KANTAR

IQVIA™

genpact

Capgemini

IBM

amazon

Deloitte.

Infosys

Mercedes-Benz

wipro

JPMorganChase

hp

Mindtree

cognizant

SymphonyAI

L&T Infotech

SAGASIT
ANALYTICS

tcs TATA
CONSULTANCY
SERVICES

DELL

HCL

EXAMROOM.AI

CISCO

Rubixe
Disruptive Technologies at Work

INTELIZION

EY

ZenSar
TECHNOLOGIES

Affine
Data + Insights + Transformation

LTIMindtree

Mphasis
The Next Applied

digitDefence
Cyber Security Made Simple.

TECH
mahindra

virtusa®
Accelerating Business Outcomes

accenture

swiggy

Mastek

happiest
minds
The Mindful IT Company

DATA SCIENCE FOUNDATION

COURSE CODE	CDM-DSF-112	LECTURE HOURS	24 hrs.
PREREQUISITES	Python Foundation	LEARNING HOURS	60 hrs.

MODULE 1

DATA SCIENCE ESSENTIALS

- Introduction to Data Science
- Evolution of Data Science
- Big Data Vs Data Science
- Data Science Terminologies
- Data Science vs AI/Machine Learning
- Data Science vs Analytics

MODULE 2

DATA SCIENCE DEMO

- Business Requirement: Use Case
- Data Preparation
- Machine learning Model building
- Prediction with ML model
- Delivering Business Value.

MODULE 3

ANALYTICS CLASSIFICATION

- Types of Analytics
- Descriptive Analytics
- Diagnostic Analytics
- Predictive Analytics
- Prescriptive Analytics
- EDA and insight gathering demo in Tableau

MODULE 4

DATA SCIENCE AND RELATED FIELDS

- Introduction to AI
- Introduction to Computer Vision
- Introduction to Natural Language Processing
- Introduction to Reinforcement Learning
- Introduction to GAN
- Introduction to Generative Passive Models

MODULE 5

DATA SCIENCE ROLES & WORKFLOW

- Data Science Project workflow
- Roles: Data Engineer, Data Scientist, ML Engineer and MLOps Engineer
- Data Science Project stages.

MODULE 6

MACHINE LEARNING INTRODUCTION

- What Is ML? ML Vs AI
- ML Workflow, Popular ML Algorithms
- Clustering, Classification And Regression
- Supervised Vs Unsupervised

MODULE 7

DATA SCIENCE INDUSTRY APPLICATIONS

- Data Science in Finance and Banking
- Data Science in Retail
- Data Science in Health Care
- Data Science in Logistics and Supply Chain
- Data Science in Technology Industry
- Data Science in Manufacturing
- Data Science in Agriculture

TOOLS/PLATFORMS COVERED



Pandas



matplotlib

NumPy



PYTHON FOUNDATION

COURSE CODE	CDM-PYF-110	LECTURE HOURS	16 hrs.
PREREQUISITES	None	LEARNING HOURS	40 hrs.

MODULE 1

PYTHON BASICS

- Introduction of python
- Installation of Python and IDE
- Python Variables
- Python basic data types
- Number & Booleans, strings
- Arithmetic Operators
- Comparison Operators
- Assignment Operators

MODULE 2

PYTHON CONTROL STATEMENTS

- IF Conditional statement
- IF-ELSE
- NESTED IF
- Python Loops basics
- WHILE Statement
- FOR statements
- BREAK and CONTINUE statements

MODULE 3

PYTHON DATA STRUCTURES

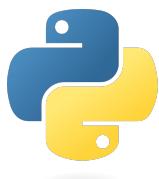
- Basic data structure in python
- Basics of List
- List: Object, methods
- Tuple: Object, methods
- Sets: Object, methods
- Dictionary: Object, methods

MODULE 4

PYTHON FUNCTIONS

- Functions basics
- Function Parameter passing
- Lambda functions
- Map, reduce, filter functions

TOOLS/PLATFORMS COVERED



ANACONDA®



Google
colab

NumPy Pandas



STATISTICS ESSENTIALS

COURSE CODE	CDM-STA-139	LECTURE HOURS	8 hrs.
PREREQUISITES	None	LEARNING HOURS	20 hrs.

MODULE 1

OVERVIEW OF STATISTICS

- Introduction to Statistics
- Descriptive And Inferential Statistics
- Basic Terms Of Statistics
- Types Of Data

MODULE 2

HARNESSING DATA

- Random Sampling
- Sampling With Replacement And Without Replacement
- Cochran's Minimum Sample Size
- Types of Sampling
- Simple Random Sampling
- Stratified Random Sampling
- Cluster Random Sampling
- Systematic Random Sampling
- Multi stage Sampling
- Sampling Error
- Methods Of Collecting Data

MODULE 3

EXPLORATORY DATA ANALYSIS

- Exploratory Data Analysis Introduction
- Measures Of Central Tendencies: Mean, Median And Mode
- Measures Of Central Tendencies: Range, Variance And Standard Deviation
- Data Distribution Plot: Histogram
- Normal Distribution & Properties
- Z Value / Standard Value
- Empirical Rule and Outliers
- Central Limit Theorem
- Normality Testing
- Skewness & Kurtosis
- Measures Of Distance: Euclidean, Manhattan And Minkowski Distance
- Covariance & Correlation

MODULE 4

HYPOTHESIS TESTING

- Hypothesis Testing Introduction
- P- Value, Critical Region
- Types of Hypothesis Testing
- Hypothesis Testing Errors : Type I And Type II
- Two Sample Independent T-test
- Two Sample Relation T-test
- One Way Anova Test
- Application of Hypothesis testing

MACHINE LEARNING ASSOCIATE

COURSE CODE	CDM-MLA-130	LECTURE HOURS	16 hrs.
PREREQUISITES	Python Foundation, DSF	LEARNING HOURS	40 hrs.

MODULE 1

MACHINE LEARNING INTRODUCTION

- What Is ML? ML Vs AI
- Clustering, Classification And Regression
- Supervised Vs Unsupervised

MODULE 2

PYTHON NUMPY PACKAGE

- Introduction to Numpy Package
- Array as Data Structure
- Core Numpy functions
- Matrix Operations, Broadcasting in Arrays

MODULE 3

PYTHON PANDAS PACKAGE

- Introduction to Pandas package
- Series in Pandas
- Data Frame in Pandas
- File Reading in Pandas
- Data munging with Pandas

MODULE 4

VISUALIZATION WITH PYTHON - Matplotlib

- Visualization Packages (Matplotlib)
- Components Of A Plot, Sub-Plots
- Basic Plots: Line, Bar, Pie, Scatter

MODULE 5

PYTHON VISUALIZATION PACKAGE - SEABORN

- Seaborn: Basic Plot
- Advanced Python Data Visualizations

MODULE 6

ML ALGO: LINEAR REGRESSION

- Introduction to Linear Regression
- How it works: Regression and Best Fit Line
- Modeling and Evaluation in Python

MODULE 7

ML ALGO: LOGISTIC REGRESSION

- Introduction to Logistic Regression
- How it works: Classification & Sigmoid Curve
- Modeling and Evaluation in Python

MODULE 8

ML ALGO: K MEANS CLUSTERING

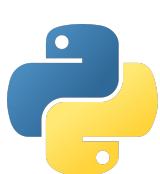
- Understanding Clustering (Unsupervised)
- K Means Algorithm
- How it works : K Means theory
- Modeling in Python

MODULE 9

ML ALGO: KNN

- Introduction to KNN
- How It Works: Nearest Neighbor Concept
- Modeling and Evaluation in Python

TOOLS/PLATFORMS COVERED



MACHINE LEARNING EXPERT

COURSE CODE	CDM-MLE-113	LECTURE HOURS	32 hrs.
PREREQUISITES	ML Associate	LEARNING HOURS	80 hrs.

MODULE 1

FEATURE ENGINEERING

- Introduction to Feature Engineering
- Feature Engineering Techniques: Encoding, Scaling, Data Transformation
- Handling Missing values, handling outliers
- Creation of Pipeline
- Use case for feature engineering

MODULE 2

ML ALGO: SUPPORT VECTOR MACHINE (SVM)

- Introduction to SVM
- How It Works: SVM Concept, Kernel Trick
- Modeling and Evaluation of SVM in Python

MODULE 3

PRINCIPAL COMPONENT ANALYSIS (PCA)

- Building Blocks Of PCA
- How it works: Finding Principal Components
- Modeling PCA in Python

MODULE 4

ML ALGO: DECISION TREE & RANDOM FOREST

- Introduction to Decision Tree & Random Forest
- How it works
- Modeling and Evaluation in Python

MODULE 5

ENSEMBLE TECHNIQUES - BAGGING

- Introduction to Ensemble technique
- Bagging and How it works
- Modeling and Evaluation in Python

MODULE 6

ML ALGO: NAÏVE BAYES

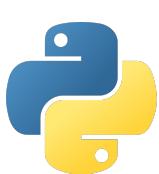
- Introduction to Naive Bayes
- How it works: Bayes' Theorem
- Naive Bayes For Text Classification
- Modeling and Evaluation in Python

MODULE 7

GRADIENT BOOSTING, XGBOOST

- Introduction to Boosting and XGBoost
- How it works?
- Modeling and Evaluation of in Python

TOOLS/PLATFORMS COVERED



ADVANCED DATA SCIENCE

COURSE CODE	CDM-ADS-114	LECTURE HOURS	16 hrs.
PREREQUISITES	Python Foundation, MLE, DSF	LEARNING HOURS	40 hrs.

MODULE 1

TIME SERIES FORECASTING - ARIMA

- What is Time Series?
- Trend, Seasonality, cyclical and random
- Stationarity of Time Series
- Autoregressive Model (AR)
- Moving Average Model (MA)
- ARIMA Model
- Autocorrelation and AIC
- Time Series Analysis in Python

MODULE 2

SENTIMENT ANALYSIS

- Introduction to Sentiment Analysis
- NLTK Package
- Case study: Sentiment Analysis on Movie Reviews

MODULE 3

REGULAR EXPRESSIONS WITH PYTHON

- Regex Introduction
- Regex codes
- Text extraction with Python Regex

MODULE 4

ML MODEL DEPLOYMENT WITH FLASK

- Introduction to Flask
- URL and App routing
- Flask application – ML Model deployment

MODULE 5

ADVANCED DATA ANALYSIS WITH MS EXCEL

- MS Excel core Functions
- Advanced Functions (VLOOKUP, INDIRECT..)
- Linear Regression with EXCEL
- Data Table
- Goal Seek Analysis
- Pivot Table
- Solving Data Equation with EXCEL

MODULE 6

AWS CLOUD FOR DATA SCIENCE

- Introduction of cloud
- Difference between GCC, Azure, AWS
- AWS Service (EC2 instance)

MODULE 7

AZURE FOR DATA SCIENCE

- Introduction to AZURE ML studio
- Data Pipeline
- ML modeling with Azure

MODULE 8

INTRODUCTION TO DEEP LEARNING

- Introduction to Artificial Neural Network, Architecture
- Artificial Neural Network in Python
- Introduction to Convolutional Neural Network, Architecture
- Convolutional Neural Network in Python

TOOLS/PLATFORMS COVERED

Natural
Language
ToolKit



Amazon
SageMaker



Azure
Machine
Learning

DATABASE: SQL AND MONGODB

COURSE CODE	CDM-DBM-120	LECTURE HOURS	6 hrs.
PREREQUISITES	None	LEARNING HOURS	15 hrs.

MODULE 1

DATABASE INTRODUCTION

- DATABASE Overview
- Key concepts of database management
- Relational Database Management System
- CRUD operations

MODULE 2

SQL BASICS

- Introduction to Databases
- Introduction to SQL
- SQL Commands
- MY SQL workbench installation

MODULE 3

DATA TYPES AND CONSTRAINTS

- Numeric, Character, date time data type
- Primary key, Foreign key, Not null
- Unique, Check, default, Auto increment

MODULE 4

DATABASES AND TABLES (MySQL)

- Create database
- Delete database
- Show and use databases
- Create table, Rename table
- Delete table, Delete table records
- Create new table from existing data types
- Insert into, Update records
- Alter table

MODULE 5

SQL JOINS

- Inner Join, Outer Join
- Left Join, Right Join
- Self Join, Cross join
- Windows function: Over, Partition, Rank

MODULE 6

SQL COMMANDS AND CLAUSES

- Select, Select distinct
- Aliases, Where clause
- Relational operators, Logical
- Between, Order by, In
- Like, Limit, null/not null, group by
- Having, Sub queries

MODULE 7

DOCUMENT DB/NO-SQL DB

- Introduction of Document DB
- Document DB vs SQL DB
- Popular Document DBs
- MongoDB basics
- Data format and Key methods

TOOLS/PLATFORMS COVERED



VERSION CONTROL WITH GIT

COURSE CODE	CDM-GIT-115	LECTURE HOURS	4 hrs.
PREREQUISITES	None	LEARNING HOURS	10 hrs.

MODULE 1

GIT INTRODUCTION

- Purpose of Version Control
- Popular Version control tools
- Git Distribution Version Control
- Terminologies
- Git Workflow
- Git Architecture

MODULE 2

GIT REPOSITORY and GitHub

- Git Repo Introduction
- Create New Repo with Init command
- Git Essentials: Copy & User Setup
- Mastering Git and GitHub

MODULE 3

COMMITTS, PULL, FETCH AND PUSH

- Code Commits
- Pull, Fetch and Conflicts resolution
- Pushing to Remote Repo

MODULE 4

TAGGING, BRANCHING AND MERGING

- Organize code with branches
- Checkout branch
- Merge branches
- Editing Commits
- Commit command Amend flag
- Git reset and revert

MODULE 5

GIT WITH GITHUB AND BITBUCKET

- Creating GitHub Account
- Local and Remote Repo
- Collaborating with other developers

TOOLS/PLATFORMS COVERED



BIG DATA FOUNDATION

COURSE CODE	CDM-BDF-117	LECTURE HOURS	4 hrs.
PREREQUISITES	Python Foundation	LEARNING HOURS	10 hrs.

MODULE 1

BIG DATA INTRODUCTION

- Big Data Overview
- Five Vs of Big Data
- What is Big Data and Hadoop
- Introduction to Hadoop
- Components of Hadoop Ecosystem
- Big Data Analytics Introduction

MODULE 2

HDFS AND MAP REDUCE

- HDFS – Big Data Storage
- Distributed Processing with Map Reduce
- Mapping and reducing stages concepts
- Key Terms: Output Format, Partitioners, Combiners, Shuffle, and Sort

MODULE 3

PYSPARK FOUNDATION

- PySpark Introduction
- Spark Configuration
- Resilient distributed datasets (RDD)
- Working with RDDs in PySpark
- Aggregating Data with Pair RDDs

MODULE 4

SPARK SQL and HADOOP HIVE

- Introducing Spark SQL
- Spark SQL vs Hadoop Hive

TOOLS/PLATFORMS COVERED



CERTIFIED BI ANALYST

COURSE CODE	CDM-BIA-119	LECTURE HOURS	6 hrs.
PREREQUISITES	None	LEARNING HOURS	15 hrs.

MODULE 1

TABLEAU FUNDAMENTALS

- Introduction to Business Intelligence & Introduction to Tableau
- Interface Tour, Data visualization: Pie chart, Column chart, Bar chart.
- Bar chart, Tree Map, Line Chart
- Area chart, Combination Charts, Map
- Dashboards creation, Quick Filters
- Create Table Calculations
- Create Calculated Fields
- Create Custom Hierarchies

MODULE 2

POWER-BI BASICS

- Power BI Introduction
- Basics Visualizations
- Dashboard Creation
- Basic Data Cleaning
- Basic DAX FUNCTION

MODULE 3

DATA TRANSFORMATION TECHNIQUES

- Exploring Query Editor
- Data Cleansing and Manipulation:
- Creating Our Initial Project File
- Connecting to Our Data Source
- Editing Rows
- Changing Data Types
- Replacing Values

MODULE 4

CONNECTING TO VARIOUS DATA SOURCES

- Connecting to a CSV File
- Connecting to a Webpage
- Extracting Characters
- Splitting and Merging Columns
- Creating Conditional Columns
- Creating Columns from Examples
- Create Data Model

TOOLS/PLATFORMS COVERED



CONTACTS & ADMISSION

CERTIFIED DATA SCIENTIST – PROGRAM ENQUIRY

DURATION : 8 MONTHS

LEARNING MODE : LIVE ONLINE / IN-PERSON CLASSROOM (SELECTED CITIES)

24x7 live chat @ www.datamites.com | admissions@datamites.com

INDIA :+91 1800-313-3434 | US: +1 628 228 6062 | UK: +44 752 066 5626



DATA SCIENCE RATED AS THE TOP 5 CAREER CHOICE
HIGHEST PAID – RECESSION PROOF – MILLIONS OF JOBS



TAKE YOUR FIRST STEP TOWARDS DATA SCIENCE CAREER

ENQUIRE NOW