DATA SCIENTIST MASTER'S PROGRAM

In collaboration with IBM

Contents

About the Course 03

Key Features of Data Scientist Master's Program 04

About IBM and Simplilearn collaboration 05

Learning Path Visualization 06

Program Outcomes 07

Who Should Enroll 08

Courses 09

- **Step 1:** Statistics Essential 9
- **Step 2:** R Programming for Data Science 11
- **Step 3:** Data Science Certification Training - R Programming 12
- **Step 4:** Python for Data Science 14
- **Step 5:** Data Science with Python 15
- **Step 6:** Machine Learning 17
- **Step 7:** Tableau 19
- **Step 8:** Big Data Hadoop and Spark Developer 21
- **Step 9:** Data Science Capstone 23

Electives 24

Certificates 25

Classroom-Level Immersion: Delivered Digitally 26

Customer Reviews 27

Advisory Board Members 28

Corporate training 29

About the Course

This Data Scientist Master's Program in collaboration with IBM accelerates your career in Data Science providing you with world-class training and skills required to become successful in this domain. The program offers extensive training on the most in-demand Data Science and Machine Learning skills with hands-on exposure to key tools and technologies including R, Python, Tableau, Hadoop, and Spark. Become an expert in Data Science by deep diving into the nuances of data interpretation, interworking technologies like Machine Learning, and mastering powerful programming skills to take your career in Data Science to the next level.

Key Features

Industry-recognized certificates from IBM (for IBM courses) and Simplilearn

Portfolio worthy capstone demonstrating mastered concepts

15+ Real-life projects providing hands-on industry training

30+ In-demand skills

Lifetime access to self-paced learning and class recordings

About IBM and Simplilearn collaboration

A joint partnership with Simplilearn and IBM introduces students to an integrated blended learning, making them an expert in Artificial Intelligence and Data Science. The program in collaboration with IBM will make students industry ready for Artificial Intelligence and Data Science job roles. IBM is a leading cognitive solution and cloud platform company, headquartered in Armonk, New York, offering a plethora of technology and consulting services. Each year, IBM invests $6 billion in research and development and has achieved five Nobel Laureates, nine US National Medals of Technology, five US National Medals of Science, six Turing Awards, and 10 Inductees in US Inventors Hall of Fame.

About Simplilearn

Simplilearn is the world's #1 online bootcamp provider that enables learners through rigorous and highly specialized training. We focus on emerging technologies and processes that are transforming the digital world, at a fraction of the cost and time as traditional approaches. Over one million professionals and 2000 corporate training organizations have harnessed our award-winning programs to achieve their career and business goals.

Learning Path - Data Scientist

1. **Statistics Essential**
   (4.3 hours)
2. **R Programming for Data Science**
   (1.7 hours)
3. **Data Science Certification Training - R Programming**
   (74.6 hours)
4. **Python for Data Science**
   (2.2 hours)
5. **Data Science with Python**
   (72 hours)
6. **Machine Learning**
   (72 hours)
7. **Tableau Training**
   (54.3 hours)
8. **Big Data Hadoop and Spark Developer**
   (84 hours)
9. **Data Science Capstone**
   (38 hours)
10. **Master's Certificate**

Electives

- SQL
- Industry Master Class – Data Science

Data Scientist Master's Program Outcomes

Gain an in-depth understanding of data structure and data manipulation

Understand and use linear and non-linear regression models and classification techniques for data analysis

Obtain an in-depth understanding of supervised and unsupervised learning models such as linear regression, logistic regression, clustering, dimensionality reduction, K-NN and pipeline

Perform scientific and technical computing using the SciPy package and its sub-packages such as Integrate, Optimize, Statistics, IO, and Weave

Gain expertise in mathematical computing using the NumPy and Scikit-Learn package

Understand the different components of the Hadoop ecosystem

Learn to work with HBase, its architecture and data storage, learning the difference between HBase and RDBMS, and use Hive and Impala for partitioning

Understand MapReduce and its characteristics, plus learn how to ingest data using Sqoop and Flume

Master the concepts recommendation engine, and time series modeling and gain practical mastery over principles, algorithms, and applications of Machine Learning

Learn to analyze data using Tableau and become proficient in building interactive dashboards

Who Should Enroll in this Program?

The Data Science role requires an amalgam of experience, Data Science knowledge, and using the correct tools and technologies. It is a solid career choice for both new and experienced professionals. Aspiring professionals of any educational background with an analytical frame of mind are most suited to pursue the Data Scientist Master's Program, including:

- IT Professionals
- Analytics Managers
- Business Analysts
- Banking and Finance Professionals
- Marketing Managers
- Supply Chain Network Managers
- Beginners or Recent Graduates in Bachelors or Masters Degree

Statistics Essential

Statistics is the science of assigning a probability to an event based on experiments. It is the application of quantitative principles to the collection, analysis, and presentation of numerical data. Ace the fundamentals of Data Science, statistics, and Machine Learning with this course. It will enable you to define statistics and essential terms related to it, explain measures of central tendency and dispersion, and comprehend skewness, correlation, regression, distribution. You will be able to make data-driven predictions through statistical inference.

Key Learning Objectives

- Understand the fundamentals of statistics
- Work with different types of data
- How to plot different types of data
- Calculate the measures of central tendency, asymmetry, and variability
- Calculate correlation and covariance
- Distinguish and work with different types of distribution
- Estimate confidence intervals
- Perform hypothesis testing
- Make data-driven decisions
- Understand the mechanics of regression analysis
- Carry out regression analysis
- Use and understand dummy variables
- Understand the concepts needed for data science even with Python and R!

Course curriculum

- Lesson 01 - Introduction
- Lesson 02 - Sample or population data?
- Lesson 03 - The fundamentals of descriptive statistics
- Lesson 04 - Measures of central tendency, asymmetry, and variability
- Lesson 05 - Practical example: descriptive statistics
- Lesson 06 - Distributions
- Lesson 07 - Estimators and estimates
- Lesson 08 - Confidence intervals: advanced topics
- Lesson 09 - Practical example: inferential statistics
- Lesson 10 - Hypothesis testing: Introduction
- Lesson 11 - Hypothesis testing: Let's start testing!
- Lesson 12 - Practical example: hypothesis testing
- Lesson 13 - The fundamentals of regression analysis
- Lesson 14 - Subtleties of regression analysis
- Lesson 15 - Assumptions for linear regression analysis
- Lesson 16 - Dealing with categorical data
- Lesson 17 - Practical example: regression analysis

R Programming for Data Science

Gain insight into the R Programming language with this introductory course. An essential programming language for data analysis, R Programming is a fundamental key to becoming a successful Data Science professional. In this course you will learn how to write R code, learn about R's data structures, and create your own functions. After the completion of this course, you will be fully able to begin your first data analysis.

Key Learning Objectives

- Learn about math, variables, and strings, vectors, factors, and vector operations
- Gain fundamental knowledge on arrays and matrices, lists, and data frames
- Get understanding on conditions and loops, functions in R, objects, classes, and debugging
- Learn how to accurately read text, CSV and Excel files plus how to write and save data objects in R to a file
- Understand and work on strings and dates in R

Course curriculum

- Lesson 01 - R basics
- Lesson 02 - Data structures in R
- Lesson 03 - R Programming fundamentals
- Lesson 04 - Working with Data in R
- Lesson 05 - Stings and Dates in R

Data Science Certification Training - R Programming

The next step to a data scientist is learning R - the upcoming and most in-demand open source technology. R is an extremely powerful Data Science and analytics language which has a steep learning curve and a very vibrant community. This is why it is quickly becoming the technology of choice for organizations who are adopting the power of analytics for competitive advantage.

Key Learning Objectives

- Gain a foundational understanding of business analytics
- Install R, R-studio, and workspace setup, and learn about the various R packages.
- Master R programming and understand how various statements are executed in R.
- Gain an in-depth understanding of data structure used in R and learn to import/export data in R.
- Define, understand and use the various apply functions and DPYR functions.
- Understand and use the various graphics in R for data visualization.
- Gain a basic understanding of various statistical concepts.
- Understand and use hypothesis testing method to drive business decisions.
- Understand and use linear, non-linear regression models, and classification techniques for data analysis.

- Learn and use the various association rules and Apriori algorithm.
- Learn and use clustering methods including K-means, DBSCAN, and hierarchical clustering.

Course curriculum

- Lesson 01 - Introduction to Business Analytics
- Lesson 02 - Introduction to R Programming
- Lesson 03 - Data Structures
- Lesson 04 - Data Visualization
- Lesson 05 - Statistics for Data Science-I
- Lesson 06 - Statistics for Data Science-II
- Lesson 07 - Regression Analysis
- Lesson 08 - Classification
- Lesson 09 - Clustering
- Lesson 10 - Association

Python for Data Science

Kickstart your learning of Python for Data Science with this introductory course and familiarize yourself with programming. Carefully crafted by IBM, upon completion of this course you will be able to write your Python scripts, perform fundamental hands-on data analysis using the Jupyter-based lab environment, and create your own Data Science projects using IBM Watson.

Key Learning Objectives

- Write your first Python program by implementing concepts of variables, strings, functions, loops, conditions
- Understand the nuances of lists, sets, dictionaries, conditions and branching, objects and classes
- Work with data in Python such as reading and writing files, loading, working, and saving data with Pandas

Course curriculum

- Lesson 01 - Python Basics
- Lesson 02 - Python Data Structures
- Lesson 03 - Python Programming Fundamentals
- Lesson 04 - Working with Data in Python
- Lesson 05 - Working with NumPy arrays

Data Science with Python

This Data Science with Python course will establish your mastery of Data Science and analytics techniques using Python. With this Python for Data Science Course, you'll learn the essential concepts of Python programming and gain in-depth knowledge in data analytics, Machine Learning, data visualization, web scraping, and natural language processing. Python is a required skill for many Data Science positions, so jump start your career with this interactive, hands-on course.

Key Learning Objectives

- Gain an in-depth understanding of Data Science processes, data wrangling, data exploration, data visualization, hypothesis building, and testing. You will also learn the basics of statistics
- Install the required Python environment and other auxiliary tools and libraries
- Understand the essential concepts of Python programming such as data types, tuples, lists, dicts, basic operators and functions
- Perform high-level mathematical computing using the NumPy package and its vast library of mathematical functions
- Perform scientific and technical computing using the SciPy package and its sub-packages such as Integrate, Optimize, Statistics, IO, and Weave
- Perform data analysis and manipulation using data structures and tools provided in the Pandas package
- Gain expertise in Machine Learning using the Scikit-Learn package
- Gain an in-depth understanding of supervised learning and unsupervised learning models such as linear regression, logistic regression, clustering, dimensionality reduction, K-NN and pipeline

- Use the Scikit-Learn package for natural language processing
- Use the matplotlib library of Python for data visualization
- Extract useful data from websites by performing web scraping using Python
- Integrate Python with Hadoop, Spark, and MapReduce

Course curriculum

- Lesson 01 - Data Science Overview
- Lesson 02: Data Analytics Overview
- Lesson 03: Statistical Analysis and Business Applications
- Lesson 04: Python Environment Setup and Essentials
- Lesson 05: Mathematical Computing with Python (NumPy)
- Lesson 06 - Scientific computing with Python (Scipy)
- Lesson 07 - Data Manipulation with Pandas
- Lesson 08 - Machine Learning with Scikit–Learn
- Lesson 09 - Natural Language Processing with Scikit Learn
- Lesson 10 - Data Visualization in Python using matplotlib
- This lesson teaches you to visualize data in python using matplotlib and plot them.
- Lesson 11 - Web Scraping with BeautifulSoup
- Lesson 12 - Python integration with Hadoop MapReduce and Spark

Machine Learning

Simplilearn's Machine Learning course will make you an expert in Machine Learning, a form of Artificial Intelligence that automates data analysis to enable computers to learn and adapt through experience to do specific tasks without explicit programming. You will master Machine Learning concepts and techniques, including supervised and unsupervised learning, mathematical and heuristic aspects, and hands-on modeling to develop algorithms and prepare you for your role with advanced Machine Learning knowledge.

Key Learning Objectives

- Master the concepts of supervised and unsupervised learning, recommendation engine, and time series modeling
- Gain practical mastery over principles, algorithms, and applications of Machine Learning through a hands-on approach that includes working on four major end-to-end projects and 25+ hands-on exercises
- Acquire thorough knowledge of the statistical and heuristic aspects of Machine Learning
- Implement models such as support vector machines, kernel SVM, naive Bayes, decision tree classifier, random forest classifier, logistic regression, K-means clustering and more in Python
- Validate Machine Learning models and decode various accuracy metrics. Improve the final models using another set of optimization algorithms, which include Boosting & Bagging techniques
- Comprehend the theoretical concepts and how they relate to the practical aspects of Machine Learning

Course curriculum

- Lesson 01 - Introduction to Artificial Intelligence and Machine Learning
- Lesson 02: Data Wrangling and Manipulation
- Lesson 03: Supervised Learning
- Lesson 04: Feature Engineering
- Lesson 05: Supervised Learning-Classification
- Lesson 06: Unsupervised learning
- Lesson 07: Time Series Modelling
- Lesson 08: Ensemble Learning
- Lesson 09: Recommender Systems
- Lesson 10: Text Mining

Tableau

This Tableau course helps you understand how to build visualizations, organize data, and design charts and dashboards to empower more meaningful business decisions. You'll be exposed to the concepts of Data Visualization, different combo charts, and stories, working with filters, parameters, and sets, and building interactive dashboards.

Key Learning Objectives

- Become an expert on visualization techniques such as heat map, treemap, waterfall, Pareto
- Understand metadata and its usage
- Work with Filter, Parameters, and Sets
- Master special field types and Tableau-generated fields and the process of creating and using parameters
- Learn how to build charts, interactive dashboards, story interfaces, and how to share your work
- Master the concepts of data blending, create data extracts and organize and format data
- Master arithmetic, logical, table, and LOD calculations

Course curriculum

- Lesson 01 - Getting Started with Tableau
- Lesson 02 - Core Tableau in Topics
- Lesson 03 - Creating Charts in Tableau
- Lesson 04 - Working with Metadata
- Lesson 05 - Filters in Tableau
- Lesson 06 - Applying Analytics to the worksheet
- Lesson 07 - Dashboard in Tableau
- Lesson 08 - Modifications to Data Connections
- Lesson 09 - Introduction to Level of Details in Tableau (LODS)

Big Data Hadoop and Spark Developer

Learn how to work Big Data and its components. Deep-dive into Hadoop and its ecosystem including MapReduce, HDFS, Yarn, HBase, Impala, Sqoop and Flume. This course provides an introduction to Apache Spark which is a next step after Hadoop. After completing this course, you will be able to successfully pass the Cloudera CCA175 certification but embrace this technology as part of your role as a Data Scientist.

Key Learning Objectives

- Master the concepts of the Hadoop framework and its deployment in a cluster environment
- Understand how the Hadoop ecosystem fits in with the data processing lifecycle
- Learn to write complex MapReduce programs
- Describe how to ingest data using Sqoop and Flume
- Get introduced to Apache Spark and its components
- List the best practices for data storage
- Explain how to model structured data as tables with Impala and Hive

Course curriculum

- Lesson 01 - Introduction to Big Data and Hadoop Ecosystem
- Lesson 02 - HDFS and Hadoop Architecture
- Lesson 03 - MapReduce and Sqoop
- Lesson 04 - Basics of Impala and Hive
- Lesson 05 - Working with Hive and Impala
- Lesson 06 - Type of Data Formats

- Lesson 07 - Advanced HIVE concept and Data File Partitioning
- Lesson 08 - Apache Flume and HBase
- Lesson 09 - Apache Pig
- Lesson 10 - Basics of Apache Spark
- Lesson 11 - RDDs in Spark
- Lesson 12 - Implementation of Spark Applications
- Lesson 13 - Spark Parallel Processing
- Lesson 14 - Spark RDD Optimization Techniques
- Lesson 15 - Spark Algorithm
- Lesson 16 - Spark SQL

Data Science Capstone

This Data Science Capstone project will give you an opportunity to implement the skills you learned throughout this Program. Through dedicated mentoring sessions, you'll learn how to solve a real-world, industry-aligned Data Science problem, from data processing and model building to reporting your business results and insights. The project is the final step in the learning path and will enable you to showcase your expertise in Data Science to future employers.

Key Learning Objectives

Simplilearn's online Data Science Capstone course will bring you through the Data Science decision cycle, including data processing, building a model and representing results. The project milestones are as follows:

- Data Processing - In this step, you will apply various data processing techniques to make raw data meaningful.
- Model Building - You will leverage techniques such as regression and decision trees to build Machine Learning models that enable accurate and intelligent predictions. You may explore Python, R to build your model. You will follow the complete model-building exercise from data split to test and training and validating data using the k-fold cross-validation process.
- Model Fine-tuning - You will apply various techniques to improve the accuracy of your model and select the champion model that provides the best accuracy.
- Dashboarding and Representing Results - As the last step, you will be required to export your results into a dashboard with meaningful insights using Tableau

Elective Course

SQL

This course gives you all the information you need to successfully start working with SQL databases and make use of the database in your applications. Learn to correctly structure your database, author efficient SQL statements, and clauses, and manage your SQL database for scalable growth.

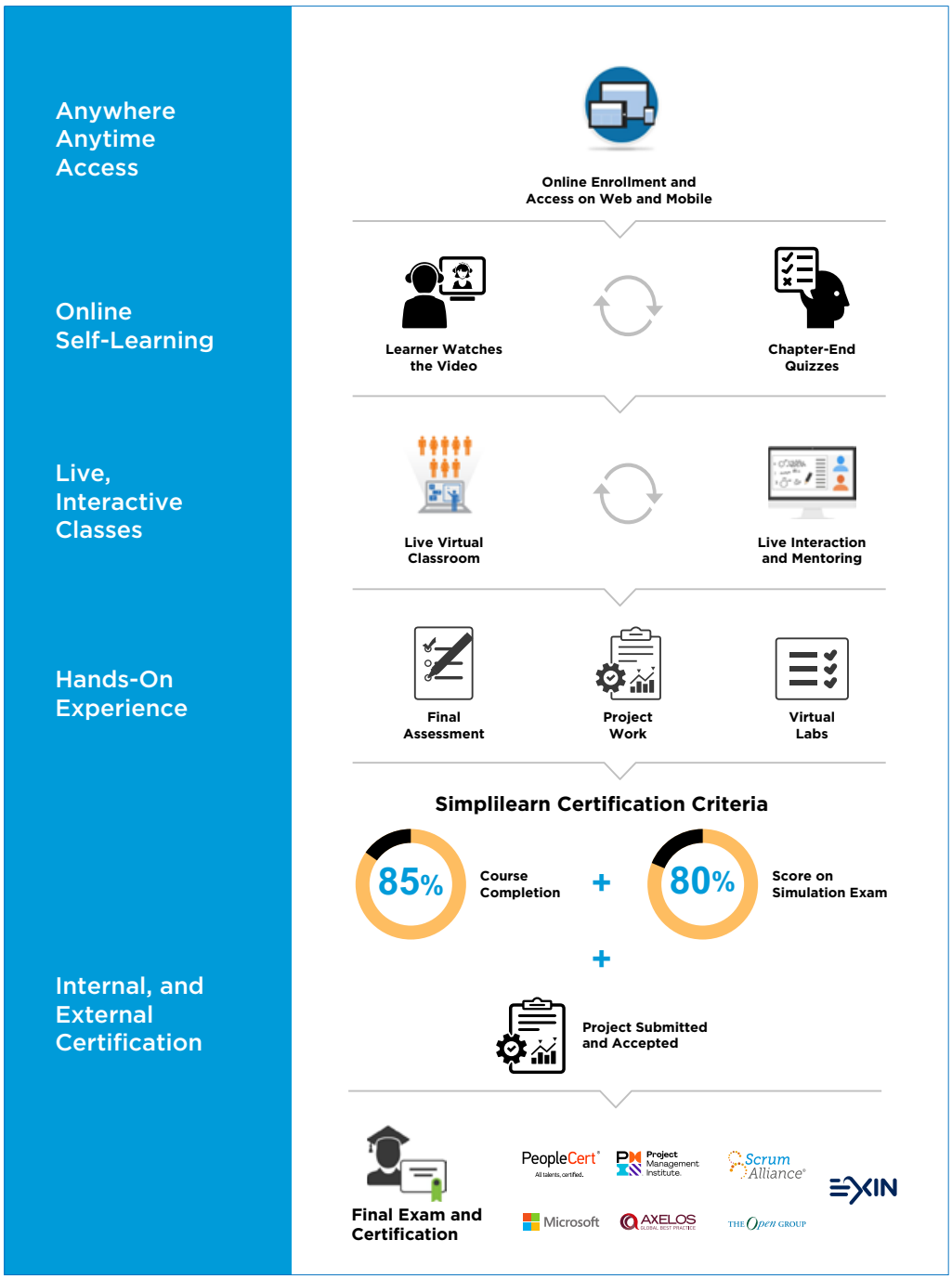
Industry Master Class – Data Science

Attend this online interactive industry master class to gain insights about Data Science advancements and AI techniques.

Certificates

Upon completion of this Master's Program, you will receive the certificates from IBM and Simplilearn in the Data Science courses in the learning path. These certificates will testify to your skills as an expert in Data Science. Upon program completion, you will also receive an industry recognized Master's Certificate from Simplilearn.

Classroom-Level Immersion: Delivered Digitally



Customer Reviews

Tajuddin Shaik

I enrolled in the Simplilearn Masters Program for Data Science to enhance my career and it was a great experience. The courses are delivered by very qualified and experienced trainers, who provided an excellent learning experience. The courses are accessible through the Simplilearn App for any time access. Yes, and 5 Star customer service.

Pramod Bhargav

Sr. Business Data Analyst and Lead

The trainer was entirely professional, knowledgeable, and helpful while clearing any doubts. Worth the money and time spent to learn from Simplilearn.

Deepika Vashishtha

The study material provided is perfect. And you get lab access on top of that which is very important when trying to learn big data technologies. I also found the live classes very beneficial. The projects and assignments provided help us

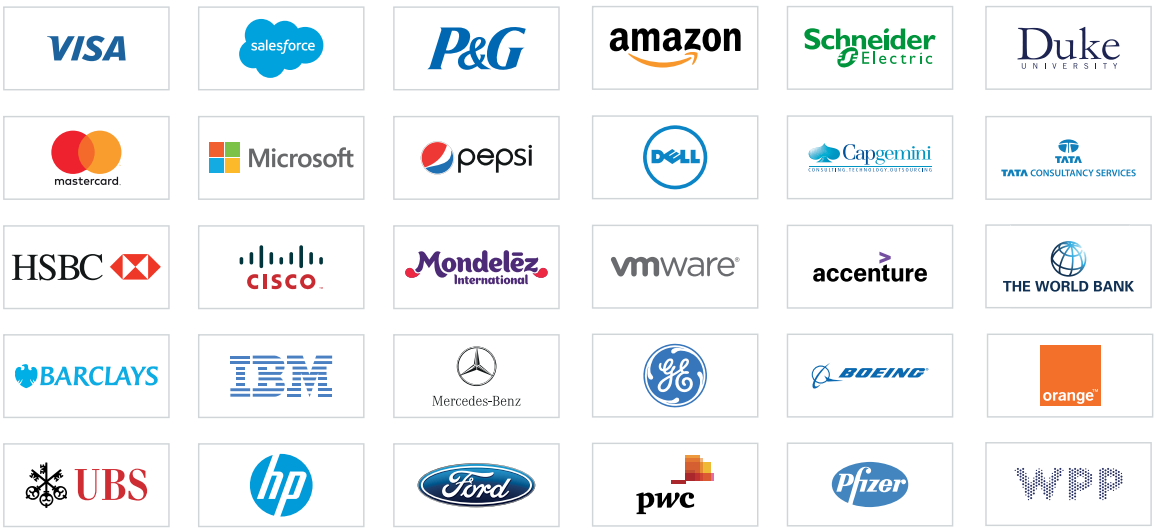
Advisory board member

Ronald Van Loon

Big Data Expert, Director Adversitement

Named by Onalytica as one of the 3 most influential people in Big Data, Ronald is an author for a number of leading Big Data and Data Science websites, including Datafloq, Data Science Central, and The Guardian. He is also a renowned speaker at industry events.

Top clients we work with:



Features of Corporate Training:

- Tailored learning solutions
- Flexible pricing options
- Enterprise-grade learning management system (LMS)
- Enterprise dashboards for individuals and teams
- 24X7 learner assistance and support

simplilearn

USA

Simplilearn Americas, Inc.
201 Spear Street, Suite 1100, San Francisco, CA 94105
United States
Phone No: +1-844-532-7688

INDIA

Simplilearn Solutions Pvt Ltd.
53/1 C, Manoj Arcade, 24th Main, Harlkunte
2nd Sector, HSR Layout
Bangalore - 560102
Call us at: 1800-212-7688

www.simplilearn.com