



Post Graduate Program in Data Science

Co-Developed with IBM

simpli|learn

Table of Contents

03	About the Program
04	The Key Features of the Post Graduate Program in Data Science by Purdue University
05	About Post Graduate Program in Data Science in Partnership with Purdue University
05	About Simplilearn
06	Program Eligibility Criteria and Application Process
08	Learning Path Visualization
09	Program Outcomes
10	Who Should Enroll in this Program
11	Courses
11	● Step 1 - Data Science in Real life
12	● Step 2 - Statistics Essential for Data Science
14	● Step 3 - R Programming for Data Science
15	● Step 4 - Data Science with R
17	● Step 5 - Python for Data Science
18	● Step 6 - Data Science with Python
20	● Step 7 - Machine Learning
22	● Step 8 - Tableau
24	● Step 9 - Big Data for Data Engineering
25	● Step 10 - Big Data Hadoop and Spark Developer
27	● Step 11 - Data Science Capstone Project
28	Electives
30	Certificates and Badges
31	Advisory Board Members



About the Program

Accelerate your career with this acclaimed Post Graduate Program in Data Science, in partnership with Purdue University and co-developed with IBM, and featuring the perfect mix of theory, case studies, and extensive hands-on practicum. In partnership with Purdue University, this program is a comprehensive Data Science education – leveraging Purdue’s academic excellence in Data Science and Simplilearn’s partnership with IBM.

Designed to give recent graduates and experienced professionals an extensive Data Science education, this Post Graduate Program is a blend of online self-paced videos, live virtual classes, hands-on projects, and labs, with mentorship sessions to provide a high-engagement learning experience and real-world applications to help you master essential Data Science skills. This program offers hands-on exposure to technologies including R, SAS, Python, Big Data, Machine Learning, Tableau, and prepares you for an exciting career in Data Science.

The Key Features of the Post Graduate Program in Data Science by Purdue University



Purdue Post Graduate Program certificate



Purdue University alumni status



International recognition by Purdue University



Leading industry-recognized analytics course



Co-developed program with IBM



Industry-recognized IBM certificates



25+ hands-on projects and one capstone



Limited class size for an optimal experience



Enrollment in Simplilearn's JobAssist

About the Post Graduate Program in Data Science in Partnership with Purdue University

Purdue University, a top public research institution, offers higher education at its highest proven value. Committed to affordability, the University has frozen tuition and most fees at 2012-13 levels. Committed to student success, Purdue is changing the student experience with a greater focus on faculty-student interaction and creative use of technology. Committed to pursuing scientific discoveries and engineered solutions, Purdue has streamlined pathways for faculty and student innovators who have a vision for moving the world forward.

This Data Science Post Graduate Program in partnership with Purdue University will open pathways for your career in virtually every realm of business—from healthcare to education to manufacturing.

Upon successfully completing this program, you will:

- > Receive a joint Purdue-Simplilearn certificate of completion
- > Be eligible able to join the Purdue alumni association and participate in its various networking opportunities and career events



About Simplilearn

Simplilearn is a leader in digital skills training, focused on the emerging technologies that are transforming our world. Our unique blended learning approach drives learner engagement and backed by the industry's highest course completion rates. Partnering with professionals and companies, we identify their unique needs and provide outcome-centric solutions to help them achieve their professional goals.

Program Eligibility Criteria and Application Process

Those wishing to enroll in the Post Graduate Program in Data Science in partnership with Purdue University will be required to apply for admission to the program.

Eligibility Criteria

For admission to this Post Graduate Program in Data Science, candidates should have:

- ✓ A bachelor's degree with an average of 50% or higher marks
- ✓ Proficiency in a programming language, collegiate mathematics, and statistics
- ✓ Current university students in their final year with an average of 50% or higher marks are also eligible

Application Process

The application process consists of three simple steps. An offer of admission will be made to the selected candidates and accepted by the candidates by paying the admission fee.

STEP 1 SUBMIT AN APPLICATION

Complete the application and include a brief statement of purpose to telling our admissions counselors why you're interested and qualified to be part of the Post Graduate Program in Data Science.

STEP 2 APPLICATION REVIEW

After you submit your application, a panel of admissions counselors will review your application and statement of purpose to determine your qualifications and interest in the program.

STEP 3 ADMISSION

An offer of admission will be made to qualified candidates, and you can accept this offer by paying the program fee.

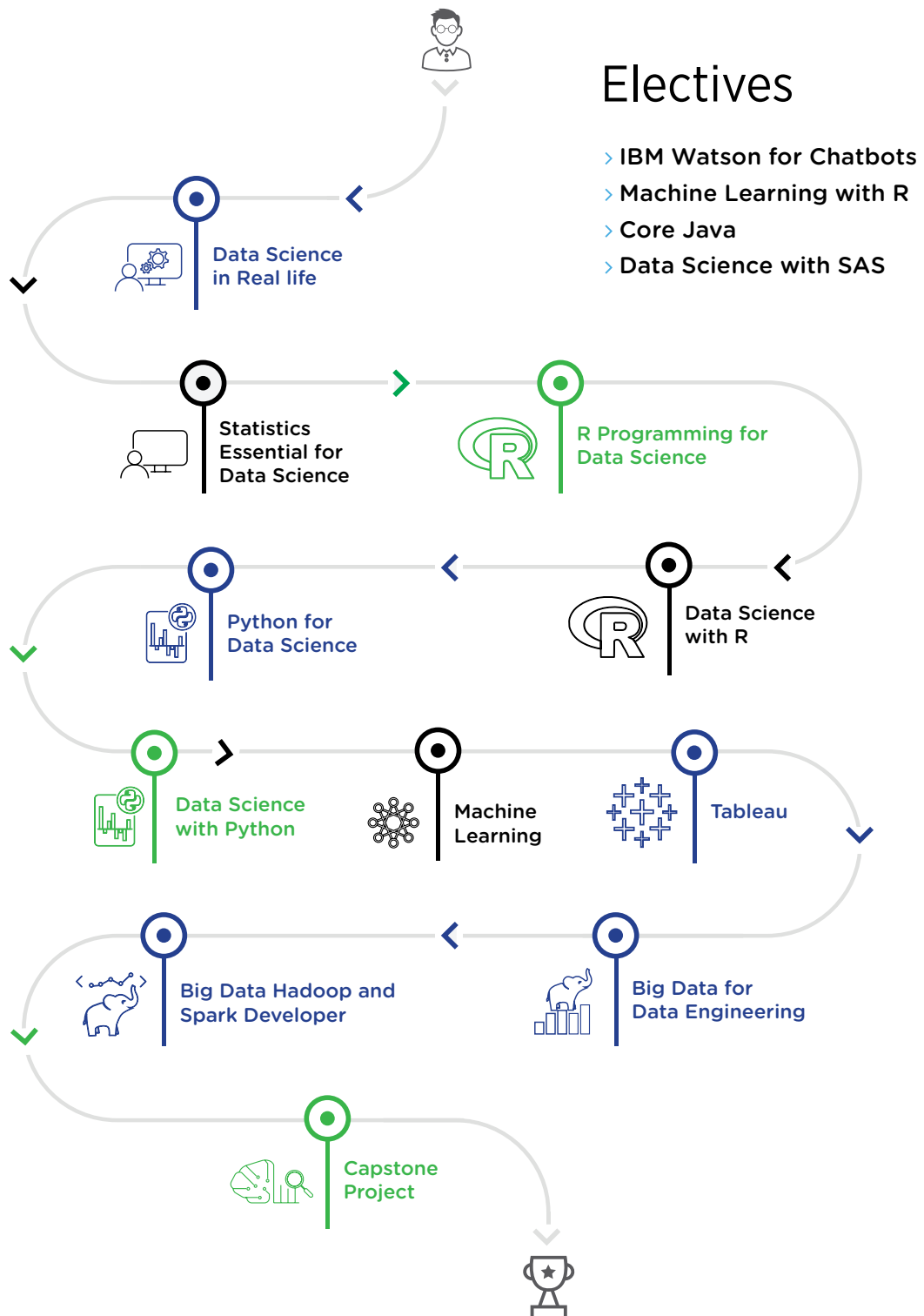


Talk to an Admissions Counselor

We have a team of dedicated admissions counselors who are here to help guide you in applying to the program. They are available to:

- ✓ Address questions related to the application
- ✓ Assist with financial aid (if required)
- ✓ Help you resolve your questions and understand the program

Learning Path



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 a comprehensive knowledge 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



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



Understand the different components of the Hadoop ecosystem and 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



Who Should Enroll in this Program?

This program caters to working professionals from a variety of industries and backgrounds; the diversity of our students adds richness to class discussions and interactions.

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 this Post Graduate Program in Data Science, including:

- ✓ IT professionals
- ✓ Analytics managers
- ✓ Business analysts
- ✓ Software developers
- ✓ Beginners or recent graduates with bachelor's or master's degree

Data Science in Real life

Data Science is a highly desired skill in our technological age. Explore the truth about what Data Science is and hear from practitioners telling real stories about what it means to work in Data Science and use cases for the same.

Key Learning Objectives

- ✓ Gain fundamental knowledge of what is Data Science and what do Data Science people do
- ✓ Learn about Data Science in a business context and what is the future of Data Science
- ✓ Understand Data Science applications and discover some use cases for Data Science

Course curriculum

- ✓ Lesson 1 - Defining Data Science
- ✓ Lesson 2 - What does a Data Science Professional do?
- ✓ Lesson 3 - Data Science in Business
- ✓ Lesson 4 - Use Cases for Data Science
- ✓ Lesson 5 - Data Science People

Statistics Essential for Data Science

Statistics is the science of assigning a probability through the collection, classification, and analysis of data. A foundational part of Data Science, this course 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 1 - Introduction
- ✓ Lesson 2 - Sample or Population Data?
- ✓ Lesson 3 - The Fundamentals of Descriptive Statistics
- ✓ Lesson 4 - Measures of Central Tendency, Asymmetry, and Variability
- ✓ Lesson 5 - Practical Example: Descriptive Statistics
- ✓ Lesson 6 - Distributions
- ✓ Lesson 7 - Estimators and Estimates
- ✓ Lesson 8 - Confidence Intervals: Advanced Topics
- ✓ Lesson 9 - 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 1 - R Basics
- ✔ Lesson 2 - Data Structures in R
- ✔ Lesson 3 - R Programming Fundamentals
- ✔ Lesson 4 - Working with Data in R
- ✔ Lesson 5 - Stings and Dates in R

Data Science with R

The next step to becoming a data scientist is learning R — the most in-demand open source technology. R is a powerful Data Science and analytics language, which has a steep learning curve and a very vibrant community. This is why R is quickly becoming the technology of choice for organizations who are embracing the power of analytics for a 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 1 - Introduction to Business Analytics
- ✓ Lesson 2 - Introduction to R Programming
- ✓ Lesson 3 - Data Structures
- ✓ Lesson 4 - Data Visualization
- ✓ Lesson 5 - Statistics for Data Science I
- ✓ Lesson 6 - Statistics for Data Science II
- ✓ Lesson 7 - Regression Analysis
- ✓ Lesson 8 - Classification
- ✓ Lesson 9 - 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 1 - Python Basics
- ✔ Lesson 2 - Python Data Structures
- ✔ Lesson 3 - Python Programming Fundamentals
- ✔ Lesson 4 - Working with Data in Python
- ✔ Lesson 5 - 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. In this Python for Data Science course, you'll learn the essential concepts of Python programming and gain in-depth knowledge of 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
- ✓ 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 1 - Data Science Overview
- ✓ Lesson 2 - Data Analytics Overview
- ✓ Lesson 3 - Statistical Analysis and Business Applications
- ✓ Lesson 4 - Python Environment Setup and Essentials
- ✓ Lesson 5 - Mathematical Computing with Python (NumPy)
- ✓ Lesson 6 - Scientific Computing with Python (Scipy)
- ✓ Lesson 7 - Data Manipulation with Pandas
- ✓ Lesson 8 - Machine Learning with Scikit-Learn
- ✓ Lesson 9 - Natural Language Processing with Scikit Learn
- ✓ Lesson 10 - Data Visualization in Python using Matplotlib
- ✓ 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 and Bagging techniques
- ✓ Comprehend the theoretical concepts and how they relate to the practical aspects of Machine Learning

Course curriculum

- ✓ Lesson 1 - Introduction to Artificial Intelligence and Machine Learning
- ✓ Lesson 2 - Data Wrangling and Manipulation
- ✓ Lesson 3 - Supervised Learning
- ✓ Lesson 4 - Feature Engineering
- ✓ Lesson 5 - Supervised Learning Classification
- ✓ Lesson 6 - Unsupervised Learning
- ✓ Lesson 7 - Time Series Modeling
- ✓ Lesson 8 - Ensemble Learning
- ✓ Lesson 9 - Recommender Systems
- ✓ Lesson 10 - Text Mining

Tableau

This Tableau Desktop 10 training will help you master the various aspects of the program and gain skills such as building visualization, organizing data, and designing dashboards. You will also learn concepts of statistics, mapping, and data connection. It is an essential asset to those wishing to succeed in Data Science.

Key Learning Objectives

- ✓ Grasp the concepts of Tableau Desktop 10, become proficient with statistics, and build interactive dashboards
- ✓ Master data sources and datable blending, create data extracts, and organize and format data
- ✓ Master arithmetic, logical, table and LOD calculations, and ad-hoc analytics
- ✓ Become an expert on visualization techniques such as heat map, tree map, waterfall, Pareto, Gantt chart, and market basket analysis
- ✓ Learn to analyze data using Tableau Desktop as well as clustering and forecasting techniques
- ✓ Gain command of mapping concepts such as custom geocoding and radial selections
- ✓ Master Special Field Types and Tableau Generated Fields and the process of creating and using parameters
- ✓ Learn how to build interactive dashboards, story interfaces, and how to share your work

Course curriculum

- ✓ Lesson 1 - Getting Started with Tableau
- ✓ Lesson 2 - Working with Tableau
- ✓ Lesson 3 - Deep Diving with Data and Connections
- ✓ Lesson 4 - Creating Charts
- ✓ Lesson 5 - Adding Calculations to your Workbook
- ✓ Lesson 6 - Mapping Data in Tableau
- ✓ Lesson 7 - Dashboards and Stories
- ✓ Lesson 8 - Visualizations for an Audience

Big Data for Data Engineering

This introductory course from IBM will teach you the basic concepts and terminology of Big Data and explore its applications across multiple industries. Learn how Big Data can improve business productivity through data processing and interpretation.

Key Learning Objectives

- ✓ Learn what Big Data is, its sources, and real-life examples
- ✓ Identify the differences between Big Data and Data Science
- ✓ Master how to use Big Data for operational analysis and improved customer service
- ✓ Understand the ecosystem of Big Data and Hadoop framework

Course curriculum

- ✓ Lesson 1 - What is Big Data?
- ✓ Lesson 2 - Big Data: Beyond the Hype
- ✓ Lesson 3 - Big Data and Data Science
- ✓ Lesson 4 - Use Cases
- ✓ Lesson 5 - Processing Big Data

Big Data Hadoop and Spark Developer

Learn how to work with 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 the next step after Hadoop. After completing this course, you will be able to successfully pass the Cloudera CCA175 certification and embrace this technology as part of your role in Data Science.

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 1 - Introduction to Big Data and Hadoop Ecosystem
- ✓ Lesson 2 - HDFS and Hadoop Architecture
- ✓ Lesson 3 - MapReduce and Sqoop
- ✓ Lesson 4 - Basics of Impala and Hive
- ✓ Lesson 5 - Working with Hive and Impala

- ✓ Lesson 6 - Type of Data Formats
- ✓ Lesson 7 - Advanced HIVE Concept and Data File Partitioning
- ✓ Lesson 8 - Apache Flume and HBase
- ✓ Lesson 9 - 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 Project

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

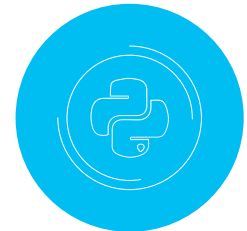
Simplilearn's online Data Science Capstone course will bring you through the Data Science decision cycle, including data processing, building a model and representing results. The project milestones are:

- ✓ **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, or SAS to develop your model. You will follow the complete model-building exercise from data split to test and validate 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 final step, you will be required to export your results into a dashboard with meaningful insights using Tableau.

Elective Course

IBM Watson for Chatbots

This course provides a practical introduction on how to build a chatbot with Watson Assistant without writing any code and then deploy your chatbot to a real website in less than five minutes. It will teach you to plan, build, test, analyze, and deploy your first chatbot.



Machine Learning with R

In this course, you will learn how to write R code, learn about R's data structures, and create your own functions. With the knowledge gained, you will be ready to undertake your first very own data analysis. You'll further learn about Supervised versus Unsupervised Learning, look into how Statistical Modeling relates to Machine Learning, and do a comparison of each using R.



Core Java

If you're looking to master functions of Big Data and Hadoop, a core fundamental to your training will be understanding Core Java. Java, by Oracle, is used in a variety of platforms from gaming consoles, laptops, and mobile technology. Java is considered a central platform due to having its own runtime environment. After this course, you will understand the methods related to Big Data and Java as well as a basic understanding of Java 8 and appropriate use cases. Gain expertise in basic concepts of Core Java and acquire a complete understanding of JDBC architecture and JUnit Framework.

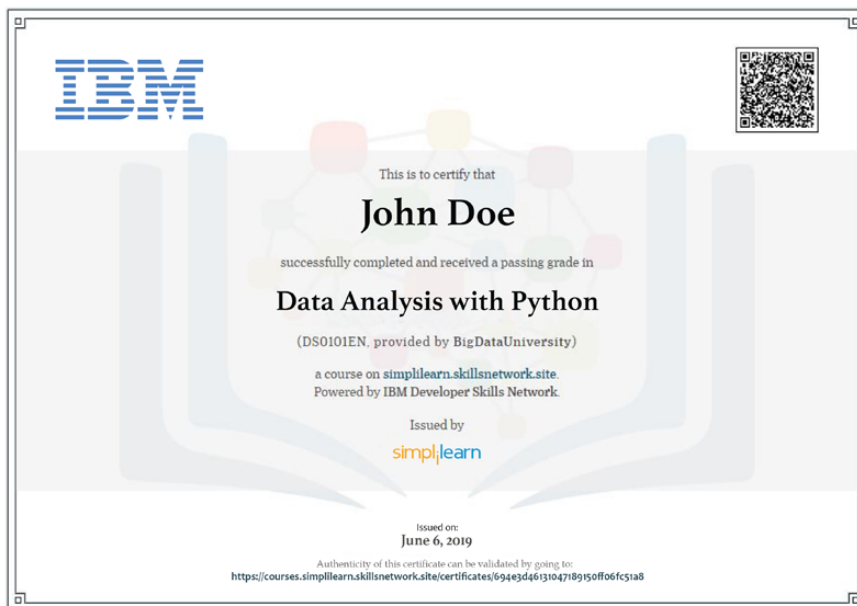


Data Science with SAS

Data Science with SAS training course is designed to enable learners to become adept in analytics techniques using SAS Data Science tools. This online course covers a holistic overview of analytics and graphic user interface (GUI). You will learn how to combine dataset methods, understand select statements and joins in SQL, and comprehend the need for macro variables. This online training course will also teach you how to apply data manipulation and optimization techniques; advanced statistical concepts like clustering, linear regression and decision trees, data analysis methods to solve real-world business problems, and predictive modeling techniques.



Certificates



Upon completion of this Post Graduate Program in Data Science by Purdue University, you will receive the Post Graduate certificate from Purdue University and IBM. You will also receive certificates from Simplilearn for the Data Science courses in the learning path. These certificates will testify to your skills as an expert in Data Science.

Advisory Board Member



Gerry McCartney

Executive Vice President for Purdue Online
and Oesterle Professor of Information
Technology at Purdue University

Gerry McCartney is the executive vice president for Purdue Online and Oesterle Professor of Information Technology at Purdue University. McCartney spearheads the online education initiative, Purdue Online, adopted in June 2018 by Purdue University President Mitch Daniels and the Board of Trustees. Under McCartney, Purdue's online offerings are rooted in market research and online analytics to provide the best experience and most value for students seeking the same kind of world-class education traditionally available from Purdue. McCartney previously held executive management positions at Purdue and the Wharton School of the University of Pennsylvania.



Ronald van Loon

Big Data Expert, Director - Advertisement

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



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, Harlunkunte
2nd Sector, HSR Layout
Bangalore - 560102
Call us at: 1800-212-7688

www.simplilearn.com