

PG Program In Data Science – UPES & Edvancer

The PG Program in Data Science – UPES & Edvancer course is a 150 hour program for working professionals & freshers alike looking to start or make a switch into a data science career. This program is one of the most comprehensive ones available in India and covers all aspects of data science. It will equip you with all the technologies, conceptual knowledge and skillsets that you need to crack any data science interview, transition into a career in this field and prosper in it. This is a course designed to get you a job in data science!

Program Highlights



150 HRS OF ONLINE CLASSES
ON WEEKENDS OR VIDEO
BASED LEARNING



HANDS-ON TRAINING WITH
70% PRACTICAL CONTENT



TOP CLASS FACULTY FROM
TOP COMPANIES



24X7 LIFETIME ACCESS TO
ONLINE LEARNING CONTENT
& VIDEOS



WORK ON 9 DIFFERENT
CAPSTONE PROJECTS FROM
VARIOUS DOMAINS



GET INDUSTRY RECOGNISED
PG CERTIFICATE FROM UPES
WITH FULL JOB ASSISTANCE

Why Should You Take This Course



Become a **full-stack data scientist** with core skills in Machine Learning and Data Science



Create a **job ready project portfolio** to establish your skills & credibility and attract recruiters



Most comprehensive curriculum covering everything from predictive analytics & machine learning



Get a **huge hike** in your salary on becoming a data scientist



We will work closely with you to help build your data science portfolio and **start a data science career**



Learn data science from one of **India's top university and data science training institute**

Technologies Covered



Our Students Are Placed In



Deloitte.



About UPES

UPES, UGC recognized and NAAC accredited, was established in 2003 and is India's first public-private university. With over 100 programs, 600 faculty and 12,000 students it is one of India's largest universities. UPES works with 3,500 recruiters and has placed over 13,000 students till date. Its School of Computer Science offers programs in association with IBM.



Global QS Ratings- 5 star rating for Employability and 4 star for Teaching



Best University for Placements and Employability', Times Now National Education Excellence Award 2018



Member of Association of Indian Universities & International Association of Universities

About Edvancer

Edvancer is **India's leading data science training institute** where we provide a range of courses on data science to all levels of learners. We have trained over 5,000 students and delivered over 10,00,000+ hours of learning. Our alumni work with some of India's top companies in data science and even globally. Our corporate clients include PwC, E&Y, L&T, HP, JP Morgan, Cognizant, Accenture, TCS, Microsoft etc.

How it works

There are 2 options to learn this course. You can choose the option based on your comfort:

1. Live Online Classes Option:

Attend 140 hours of live online classes on weekends + go through 10 hours of self-paced videos.

Duration: 20 weekends (Sat & Sun)

Ask your questions and doubts to the faculty in the class like a normal class. Online sessions are recorded for you to view and revise later whenever you want or if you miss a class. Get the benefits of learning from your home through fully interactive, online classes. SQL will be delivered through videos only.

2. Self-Paced + Faculty Support:

Learn through **150 hours of recorded class videos** at your time and pace. The curriculum, content, projects, assignments and everything else remains the same as the live online classes. There are no deadlines or timelines for you to worry about. Get all your doubts and queries cleared by faculty through forums & emails. Learn easily at your own time and pace from anywhere. Enrol and start learning immediately!

Fees

As mentioned on website (<https://www.edvancer.in/course/data-science-specialization>)

Example List of Industry Projects

1. **Pharma:** Predict the sales volume of counterfeit medicines in order to guide law enforcement agencies in cracking down on top counterfeiters.
2. **BFSI:** Identify fraudulent transactions for a credit card providing company to create an early warning system to prevent frauds in real time.
3. **BPO:** Create a machine learning system which will automatically pick out customer complaints most likely to be unresolved and escalate them.
4. **Real Estate:** Predict housing prices in a certain city to help a realty tech company close negotiations early
5. **Insurance:** Create a model to cross-sell an new car insurance policy to an existing policy holder
6. **HR:** Predict employee attrition for an IT company
7. **Retail:** Identify locations where a retail giant's new stores should be opened and will do well.
8. **Manufacturing:** Predict inventory requirement of spare-parts for a large engineering co.
9. **Data Visualization for Retail:** Create a dashboard for the management of a large retailer based on sales and inventory data

Full Curriculum

Module 1: Predictive Analytics in R

What is this module about?: Predictive Analytics is the scientific process of deriving insights from raw data to support decision making and is the core of data science. Through this module you will learn **how to use statistical techniques and R language to solve business problems**. This is a **comprehensive module** which will take you from the basics right up to **building predictive models**.

Tools to be learnt: R

Duration: 60 hours + videos

Topic	What does it mean?
<p>Introduction to business analytics</p> <ul style="list-style-type: none"> • What is analytics & why is it so important? • Applications of analytics • Different kinds of analytics • Various analytics tools • Analytics project methodology • Real world case study 	<p>In this section we shall provide you an overview into the world of analytics. You will learn about the various applications of analytics, how companies are using analytics to prosper and study the analytics project methodology through a real-world case study</p>
R Training	
<p>Fundamentals of R</p> <ul style="list-style-type: none"> • Installation of R & R Studio • Getting started with R • Basic & advanced data types in R • Variable operators in R • Working with R data frames • Reading and writing data files to R • R functions and loops • Special utility functions • Merging and sorting data • Case study on data management using R • Practice assignment 	<p>This part is all about learning how to manage and manipulate data and datasets, the very first step of analytics. We shall teach you how to use R to work with data using a case study.</p>
<p>Data visualization in R</p> <ul style="list-style-type: none"> • Need for data visualization • Components of data visualization • Utility and limitations • Introduction to grammar of graphics • Using the ggplot2 package in R to create visualizations 	<p>Data visualization is extremely important to understand what the data is saying and gain insights in just one glance. Visualization of data is a strong point of the R software and you will learn the same in this module.</p>
<p>Data preparation and cleaning using R</p> <ul style="list-style-type: none"> • Needs & methods of data preparation • Handling missing values • Outlier treatment • Transforming variables • Derived variables • Binning data • Modifying data with Base R • Data processing with dplyr package • Using SQL in R • Practice assignment 	<p>Real world data is rarely going to be given to you perfect on a platter. It will always be dirty with missing data points, incorrect data, variables needing to be changed or created in order to analyze etc. A typical analytics project will have 60% of its time spent on preparing data for analysis. This is a crucial process as properly cleaned data will result in more accurate and stable analysis. We shall teach you all the techniques required to be successful in this aspect.</p>
Setting the base of business analytics	
<p>Understanding the data using univariate statistics in R</p> <ul style="list-style-type: none"> • Summarizing data, measures of central tendency • Measures of variability, distributions • Using R to summarize data • Case study on univariate statistics using R • Practice assignment 	<p>This is where you shall learn how to start understanding the story your data is narrating by summarizing the data, checking its variability and shape by visualizing it. We shall take you through various ways of doing this using R and also solve a case study</p>

<p>Hypothesis testing and ANOVA in R to guide decision making</p> <ul style="list-style-type: none"> • Introducing statistical inference • Estimators and confidence intervals • Central Limit theorem • Parametric and non-parametric statistical tests • Analysis of variance (ANOVA) • Conducting statistical tests • Practice assignment 	<p>With 95% confidence we can say that there is an 85% chance, people visiting this site twice will enroll for the course 😊. Here, you learn how to create a hypothesis, test and validate it through data within a statistical framework and present it with clear and formal numbers to support decision making.</p>
<p>Predictive modelling in R</p>	
<p>1. Correlation and Linear regression</p> <ul style="list-style-type: none"> • Correlation • Simple linear regression • Multiple linear regression • Model diagnostics and validation • Case study 	<p>A statistical model is the core of predictive analytics and regression is one of the most powerful tools for making predictions by finding patterns in data. You shall learn the basic of regression modelling hands-on through real world cases</p>
<p>2. Logistic regression</p> <ul style="list-style-type: none"> • Moving from linear to logistic • Model assumptions and Odds ratio • Model assessment and gains table • ROC curve and KS statistic • Case Study 	<p>Logistic regression is the work-horse of the predictive analytics world. It is used to make predictions in cases where the outcomes are dual in nature i.e. an X or Y scenario where we need to predict if X will be the case or will Y, given some data. This is a must-know technique and we shall make you comfortable with it through real world problems.</p>
<p>3. Techniques of customer segmentation</p> <ul style="list-style-type: none"> • Need for segmentation • Criterion of segmentation • Types of distances • Hierarchical clustering • K-means clustering • Deciding number of clusters • Case study 	<p>Learn why and how to statistically divide a broad customer market into various segments of customers who are similar to each other so as to be able to better target and meet their needs in a cost effective manner. This is one of the most essential techniques in marketing analytics.</p>
<p>4. Time series forecasting techniques</p> <ul style="list-style-type: none"> • Need for forecasting • What are time series? • Smoothing techniques • Time series models • ARIMA 	<p>The ability to forecast into the future is very important for any business and it is necessary to have as accurate a forecasting as possible for financial and strategic planning. In this module learn the techniques of time series analysis without being misled by seasonal and cyclical impacts.</p>
<p>5. Decision trees & Random Forests</p> <ul style="list-style-type: none"> • What are decision trees • Entropy and Gini impurity index • Decision tree algorithms • CART • Random Forest • Case Study 	<p>Decision trees are predictive models which map observations about an item to conclusions about the item's target value. Learn the technique of decision trees, one of the most popular predictive analytics techniques</p>
<p>6. Boosting Machines</p> <ul style="list-style-type: none"> • Concept of weak learners • Introduction to boosting algorithms • Adaptive Boosting • Extreme Gradient Boosting (XGBoost) • Case study 	<p>Want to win a data science contest on Kaggle or data hackathons or be known as a top data scientist? Then learning boosting algorithms is a must as they provide a very powerful way of analysing data and solving hard to crack problems</p>

7. Cross Validation & Parameter Tuning

- Model performance measure with cross validation
- Parameter tuning with grid & randomised grid search

Learn how to make your model more accurate and perform the best on real -world data

Module 2: Machine Learning in Python

What is this module about?: Through this Machine Learning module, you will learn how to process, clean, visualize and automate decision making through data science by using Python, one of the most popular machine learning tools. You will learn cutting edge machine learning techniques in Python.

Tools to be learnt: Python (Libraries like pandas, numpy, scipy, scikit-learn, bokeh, beautifulsoup)

Duration: 63 hours + videos

Topic	What does it mean?
Introduction to Machine Learning in Python <ul style="list-style-type: none">• What is machine learning & why is it so important?• Applications of machine learning across industries• Machine Learning methodology• Machine Learning Toolbox• Tool of choice- Python: what & why?• Course Components	In this section we shall provide you an overview into the world of machine learning. You will learn about the various applications of machine learning, how companies from all sort of domains are solving their day to day to long term business problems. We'll learn about required skill sets of a machine learning expert which make them capable of filling up this vital role. Once the stage is set and we understand where we are heading we discuss why Python is the tool of choice in data science.
Python Training	
Introduction to Python <ul style="list-style-type: none">• Installation of Python framework and packages: Anaconda and pip• Writing/Running python programs using Spyder, Command Prompt• Working with Jupyter Notebooks• Creating Python variables: Numeric, string and logical operations• Basic Data containers: Lists, Dictionaries, Tuples & sets• Practice assignment	Python is one of the most popular & powerful languages for data science used by most top companies like Facebook, Amazon, Google, Yahoo etc. It is free and open source. This module is all about learning how to start working with Python. We shall teach you how to use the Python language to work with data.
Iterative Operations & Functions in Python <ul style="list-style-type: none">• Writing for loops in Python• List & Dictionary Comprehension• While loops and conditional blocks• List/Dictionary comprehensions with loops• Writing your own functions in Python• Writing your own classes and functions as class objects• Practice assignment	This is where we move beyond simple data containers and learn about amazing possibilities and functionalities hidden in various associated operators. We get introduced to wonderful world of loops, list and dictionary comprehensions. In addition to already existing functions and classes we learn to write our own custom functions and classes. This module sets the stage for handling data and ML algorithm implementation in python.
Data Summary; Numerical and Visual in Python <ul style="list-style-type: none">• Need for data summary• Summarising numeric data in pandas• Summarising categorical data• Group wise summary of mixed data• Need for visual summary• Introduction to ggplot & Seaborn• Visual summary of different data combinations• Practice Exercise	Data summary is extremely important to understand what the data is saying and gain insights in just one glance. Visualization of data is a strong point of the Python software using the latest ggplot package using much celebrated grammar of graphics. We also introduce you another powerful package seaborn in additional material section.

<p>Data Handling in Python using NumPy & Pandas</p> <ul style="list-style-type: none"> • Introduction to NumPy arrays, functions & properties • Introduction to pandas • Dataframe functions and properties • Reading and writing external data • Manipulating Data Columns 	<p>Python is a very versatile language and in this module we expand on its capabilities related to data handling. Focusing on packages numpy and pandas we learn how to manipulate data which will be eventually useful in converting raw data suitable for machine learning algorithms.</p>
<p>Machine Learning in Python</p>	
<p>Basics of Machine Learning</p> <ul style="list-style-type: none"> • Business Problems to Data Problems • Broad Categories of Business Problems • Supervised and Unsupervised Machine Learning Algorithm • Drivers of ML algorithms • Cost Functions • Brief introduction to Gradient Descent • Importance of Model Validation • Methods of Model Validation • Introduction to Cross Validation and Average Error 	<p>In this module we understand how we can transform our business problems to data problems so that we can use machine learning algorithms to solve them. We will further get into discovering what all categories of business problems and subsequently which machine learning algorithms are there.</p> <p>We'll learn what is the ultimate goal of any machine learning algorithm and go through a brief description of the mother of many modern optimisation methods- Gradient Descent. We'll wrap up this module with discussion on importance and methods of validation of our results.</p>
<p>Generalised Linear Models in Python</p> <ul style="list-style-type: none"> • Linear Regression • Limitation of simple linear models and need of regularisation • Ridge and Lasso Regression (L1 & L2 Penalties) • Introduction to Classification with Logistic Regression • Methods of threshold determination and performance measures for classification score models • Case Studies 	<p>We start with implementing machine learning algorithms in this module. We also get exposed to some important concepts related to regression and classification which we will be using in the later modules as well. Also this is where we get introduced to scikit-learn, the legendary python library famous for its machine learning prowess.</p>
<p>Tree Models using Python</p> <ul style="list-style-type: none"> • Introduction to decision trees • Tuning tree size with cross validation • Introduction to bagging algorithm • Random Forests • Grid search and randomized grid search • ExtraTrees (Extremely Randomised Trees) • Partial Dependence Plots • Case Studies • Home exercises 	<p>In this module we will learn a very popular class of machine learning models, rule based tree structures also known as Decision Trees. We'll examine their biased nature and learn how to use bagging methodologies to arrive at a new technique known as Random Forest to analyse data. We'll further extend the idea of randomness to decrease bias in ExtraTrees algorithm. In addition, we learn about powerful tools used with all kind of machine learning algorithms, gridSearchCV and RandomizedSearchCV.</p>
<p>Boosting Algorithms using Python</p> <ul style="list-style-type: none"> • Concept of weak learners • Introduction to boosting algorithms • Adaptive Boosting • Extreme Gradient Boosting (XGBoost) • Case study • Home exercise 	<p>Want to win a data science contest on Kaggle or data hackathons or be known as a top data scientist? Then learning boosting algorithms is a must as they provide a very powerful way of analysing data and solving hard to crack problems.</p>
<p>Support Vector Machines (SVM) and KNN in Python</p> <ul style="list-style-type: none"> • Introduction to idea of observation based learning • Distances and Similarities • K Nearest Neighbours (KNN) for classification • Introduction to SVM for classification • Regression with KNN and SVM • Case study • Home exercises 	<p>We step in a powerful world of "observation based algorithms" which can capture patterns in the data which otherwise go undetected. We start this discussion with KNN which is fairly simple. After that we move to SVM which is very powerful at capturing non-linear patterns in the data.</p>

<p>Unsupervised learning in Python</p> <ul style="list-style-type: none"> • Need for dimensionality reduction • Introduction to Principal Component Analysis (PCA) • Difference between PCAs and Latent Factors • Introduction to Factor Analysis • Patterns in the data in absence of a target • Segmentation with Hierarchical Clustering and K-means • Measure of goodness of clusters • Limitations of K-means • Introduction to density based clustering (DBSCAN) 	<p>Many machine learning algorithms become difficult to work with when dealing with many variables in the data. In comes to rescue PCA which solves problems arising from data which has highly correlated variables. The same idea can be extended to find out hidden factors in our data with Factor Analysis which is used extensively in surveys and marketing analytics.</p> <p>We also learn about two very important segmentation algorithms; K-means and DBSCAN and understand their differences and strengths.</p>
<p>Neural Networks</p> <ul style="list-style-type: none"> • Introduction to Neural Networks • Single layer neural network • Multiple layer Neural network • Back propagation Algorithm • Moment up and decaying learning rate in context of gradient descent • Neural Networks implementation in Python 	<p>Artificial Neural Networks are the building blocks of artificial intelligence. Learn the techniques which replicate how the human brain works and create machines which can solve problems like humans.</p>
<p>Text Mining in Python</p> <ul style="list-style-type: none"> • Quick Recap of string data functions • Gathering text data using web scraping with urllib • Processing raw web data with BeautifulSoup • Interacting with Google search using urllib with custom user agent • Collecting twitter data with Twitter API • Introduction to Naive Bayes • Feature Engineering for text Data • Feature creation with TFIDF for text data • Case Studies 	<p>Unstructured text data accounts for more and more interaction records as most of our daily life moves online. In this module we start our discussion by looking at ways to collect all that data. In addition to scraping simple web data; we'll also learn to use data APIs with example of Twitter API, right from the point of creating a developer account on twitter. Further we discuss one of the very powerful algorithm when it comes to text data; Naive Bayes. Then we see how we can mine the text data.</p>
<p>Ensemble Methods in Machine Learning</p> <ul style="list-style-type: none"> • Making use of multiple ML models taken together • Simple Majority vote and weighted majority vote • Blending • Stacking • Case Study 	<p>Individual machine learning models extract pattern from the data in different ways , which at times results in them extracting different patterns from the data. Rather than sticking to just one algorithm and not making use of other's results is what we move past in this module. We learn to make use of multiple ML models taken together to make our predictive modelling solutions even more powerful.</p>
<p>Bokeh</p> <ul style="list-style-type: none"> • Introduction to Bokeh charts and plotting 	<p>For making quick prototypes of your solutions which can be scaled later as interactive visualisation in the form of standalone or hosted web pages, we introduce you to Bokeh, an evolving library in python which has all the tools that you'll need to do the same.</p>
<p>Version Control using Git and Interactive Data Products</p> <ul style="list-style-type: none"> • Need and Importance of Version Control • Setting up git and github accounts on local machine • Creating and uploading GitHub Repos • Push and pull requests with GitHub App • Merging and forking projects • Examples of static and interactive data products 	<p>We finish this module with a discussion on two very important aspects of a data scientist's work. First is version control which enables you to work on large projects with multiple team members scattered across the globe. We learn about git and most widely used public platform version control that is GitHub.</p>

Module 3: Domain Specialization

What is this module about?: Through this module you can specialize in data science for any 1 out of the below 5 domains. Here you will learn the nuances of data science for the specific domain and solve case studies and work on an industry project.

Class Duration: 15 hours

Domains:

1. Banking – Finance
2. E-commerce
3. Power
4. Logistics and Supply Chain
5. Aviation

Module 4: Data Analysis in SQL (Videos Only)

What is this module about?: This Data Analyst using SQL video tutorial teaches you how to use the ever popular SQL language to analyse data stored in databases. SQL is a requirement in almost all analytics roles and this module will make you eligible to work as a data analyst. In this SQL tutorial you will learn how to communicate with databases, extract data from them, manipulate and analyse it & create reports.

Tools to be learnt: MS SQL

Class Duration: 6 hours of pre-recorded videos

Topic	What does it mean?
Introduction To SQL <ul style="list-style-type: none">• What is SQL?• Why SQL?• What are relational databases?• SQL command group• MS SQL Server installation• Exercises	Structured Query Language (SQL) is a standard language for storing, manipulating and retrieving data in databases. It is a heavily used language and a must know for every data scientist. Here we will introduce you to SQL using MS SQL.
SQL Data Types & Operators <ul style="list-style-type: none">• SQL Data Types• Filtering Data• Arithmetic Operators• Comparison operators• Logical Operators• Exercises	Learn about various types of data and how to filter and conduct basic operations on data in databases using SQL.
Useful Operations in SQL <ul style="list-style-type: none">• Distinct Operation• Top N Operation• Sorting results• Combine results using Union• Null comparison• Alias	Learn more advanced operations on data.
Aggregating Data in SQL <ul style="list-style-type: none">• Aggregate functions• Group By clause• Having clause• Over clause• Exercises	Aggregate data using various conditions and clauses in SQL to gain the answers you are looking for.

<p>Writing Sub-Queries in SQL</p> <ul style="list-style-type: none"> • What are sub-queries? • Sub-query rules • Writing sub-queries • Exercises 	<p>A subquery is a SQL query within a query. Subqueries are nested queries that provide data to the enclosing query. In this module you will learn how to write various sub-queries.</p>
<p>Common function in SQL</p> <ul style="list-style-type: none"> • Ranking functions • Date & time functions • Logical functions • String functions • Conversion functions • Mathematical functions • Exercises 	<p>Learn some of the common functions available in SQL to transform the data into more meaningful data.</p>
<p>Analytic Functions in SQL</p> <ul style="list-style-type: none"> • What are analytic functions? • Various analytic functions • SQL syntax for analytic functions • Exercises 	<p>Here you will learn various analytics function in SQL to undertake data analysis in SQL.</p>
<p>Writing DML Statements</p> <ul style="list-style-type: none"> • What are DML Statements? • Insert statement • Update statement • Delete statement 	<p>DML is abbreviation of Data Manipulation Language in SQL. It is used to retrieve, store, modify, delete, insert and update data in databases.</p>
<p>Writing DDL Statements</p> <ul style="list-style-type: none"> • What are DDL Statements? • Create statement • Alter statement • Drop statement • Exercises 	<p>DDL refers to "Data Definition Language", a subset of SQL statements that change the structure of the database schema in some way, typically by creating, deleting, or modifying schema objects such as databases, tables, and views.</p>
<p>Using Constraints in SQL</p> <ul style="list-style-type: none"> • What are constraints? • Not Null Constraint • Unique constraint • Primary key constraint • Foreign key constraint • Check constraint • Default Constraint • Exercises 	<p>Constraints provide a standard mechanism to maintain the accuracy and integrity of the data inside table. There are several different types of constraints in SQL which you will learn here.</p>
<p>SQL Joins</p> <ul style="list-style-type: none"> • What are joins? • Cartesian Join • Inner Join • Left & Right Join • Full Join • Self Join 	<p>A SQL Join statement is used to combine data or rows from two or more tables. Learn the various joins in SQL in this module.</p>
<p>Views in SQL</p> <ul style="list-style-type: none"> • What are views? • Create View • Drop view • Update view 	<p>A view is a virtual table that consists of columns from one or more tables. Though it is similar to a table, it is not stored in the database. It is a query stored as an object. Hence, a view is an object that derives its data from one or more tables. Learn how to create these views in this module.</p>