



**upGrad**



International  
Institute of Information  
Technology Bangalore

# Post Graduate Diploma in Data Science

12 MONTHS | ONLINE



**upGrad** is an online education platform to help individuals develop their professional potential in the most engaging learning environment. Online education is a fundamental disruption that will have a far-reaching impact. At upGrad, we are working towards transforming this online education wave into a tsunami! We are taking a full-stack approach of leveraging content, technology, marketing and services to offer quality education at scale in partnership with corporates and academics to offer a rigorous and industry-relevant program.

Based on our market research and conversation with the industry, we have identified Data Science as one of the sectors with critical supply demand imbalance. Our vision is to design and deliver a quality online Post Graduate Diploma in Data Science to drive the growth of the sector and make India a global hub for data science.

If you are reading this, you may wish to accelerate your career in Data Science. With upGrad, we promise to equip you with the perfect mix of business acumen and technical capabilities to help you achieve exactly the same.

## **Ronnie Screwvala**

Co-founder & Chairman  
upGrad



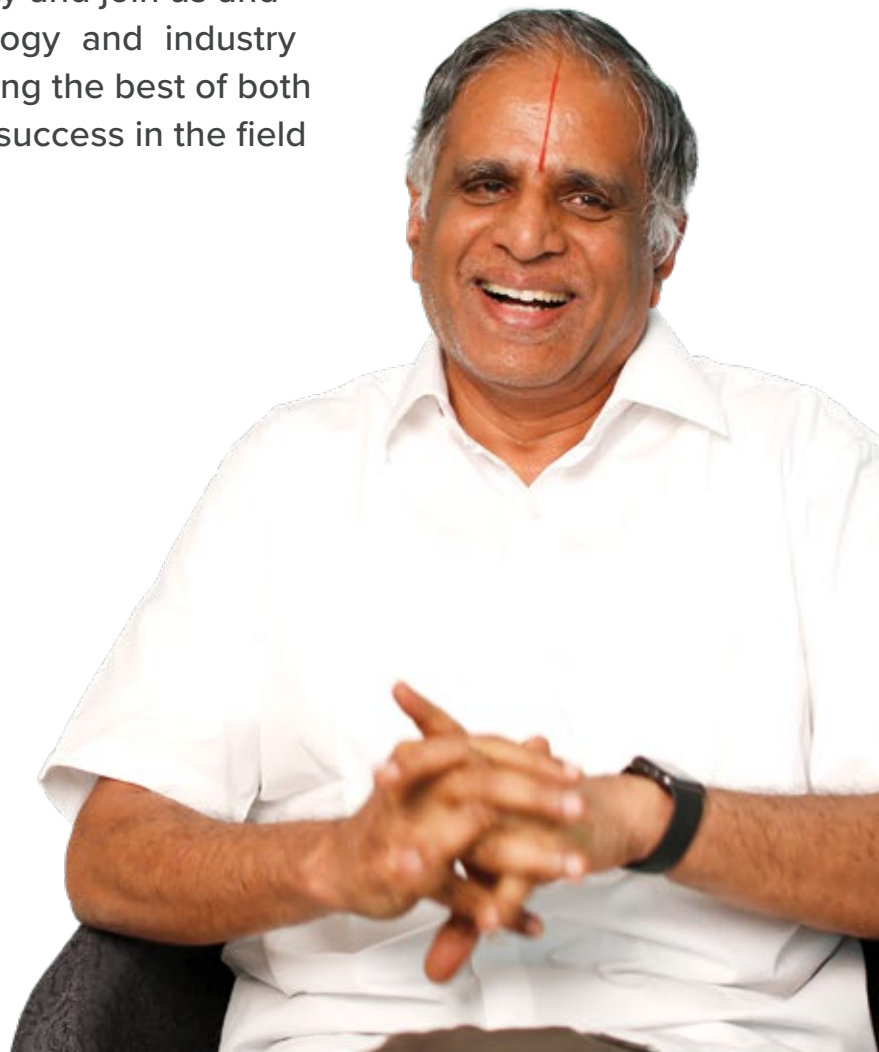
**IIIT-B** is well respected in the global academic circles as well as in the global analytics and IT industry. We have outstanding world-class faculty, with years of experience, in charge of running the Analytics Essentials Executive Education Course.

We are excited to partner with upGrad to offer a one-of-its-kind, academically rigorous and industrially relevant PG Diploma in Data Science. IIIT-B's faculty will be discussing the conceptual depths of topics such as Data Science, Machine Learning and AI, Big Data Analytics and this will be complimented by industry relevant case studies from major industry verticals by industry leaders from upGrad's industry network. Further, our strong placement network, industry mentorship and the credibility of a PG Diploma will provide you with just the right push to accelerate your career in Data Science!

We invite you to take this opportunity and join us and make use of the excellent pedagogy and industry collaborations. You will truly be getting the best of both worlds, which will help you achieve success in the field of Data Science.

## **Prof. S. Sadagopan**

Director  
IIIT Bangalore



# PROGRAM HIGHLIGHTS



## PG DIPLOMA BY IIIT-B

Get certified by IIIT-B on successful completion of the PG Diploma in Data Science.



## FACULTY AND INDUSTRY EXPERTS

IIIT-B faculties with 15+ yrs of experience and industry experts with 8+ yrs of work experience



## BLENDED LEARNING

A perfect blend of recorded and live learning to suit every learning style



## PERSONAL MENTORSHIP

Access to personal student mentor and industry mentor for each learner



## SPECIALISED TRACKS

Learners can elect specialised tracks on the basis of their target role in the data domain



## INDUSTRY PROJECTS

Students will be solving 10+ real-life industry projects and assignments

# CAREER SUPPORT



## PERSONALISED INDUSTRY MENTORSHIP

Get mentored by an experienced data science industry expert and receive personalised feedback.



## MOCK INTERVIEWS

Get company and role-specific preparation with mock interviews before your actual interviews.



## RESUME REVIEW

Obtain specific, personalised inputs on your resume structure and content.



## OFFLINE HIRING EVENT

Dedicated offline hiring events in collaboration with reputed data science companies.



## DEDICATED CAREER MENTOR

Get a dedicated career mentor to help you alongside your career transition journey.



## JOB OPPORTUNITIES

Get an opportunity to be interviewed by upGrad's data science hiring partners.

# INSIGHTS FROM INDUSTRY EXPERTS



**TEJAS SANGHVI**  
Vice President  
Fractal Analytics



**ANSHUMAN GUPTA**  
Director - Data Science  
Pitney Bowes



**UJJYAINI MITRA**  
Head of Analytics  
Viacom 18



**S. ANAND**  
CEO  
Gramener



**SAMEER DHANRAJANI**  
CSO  
Fractal Analytics



**KALPANA SUBBARAMAPPA**  
Ex-Assis. VP, Decision Sciences  
GENPACT



**SAI ALLURI**  
PRO Analytics &  
Strategy Manager  
Uber



**HINDOL BASU**  
Partner  
Tata IQ

# CONCEPTS FROM TOP ACADEMICIANS



**PROF. S. SADAGOPAN**  
Director  
IIIT Bangalore



**TRICHA ANJALI**  
Associate Professor  
IIIT Bangalore



**CHANDRASHEKAR  
RAMANATHAN**  
Dean (Academics)  
IIIT Bangalore



**DINESH BABU JAYAGOPI**  
Assistant Professor  
IIIT Bangalore



**G SRINIVASARAGHAVAN**  
Professor  
IIIT Bangalore

# UPGRAD BASECAMP: OFFLINE MEET-UPS & LEARNING

Along with online learning, upGrad basecamp meetups act as a physical platform for extensive peer-to-peer learning, networking and idea exchanges. Basecamp brings together, a hybrid offline and online learning experience for all the learners and alumni of upGrad. Held across major cities in India, these fun, yet informative and career building events add on to the already great learning experience that upGrad provides.

## WHAT'S IN IT FOR OUR LEARNERS?



Networking with  
Faculty and  
Student Mentors



Career  
Building  
Sessions



Engaging  
Live Group  
Projects



Fun and  
Exciting  
Activities



Peer Interaction  
and Career  
Networking



# PROGRAM CURRICULUM

This curriculum is subject to change based on inputs from IIT-B and industry modules marked as (\*) are optional.

## CURRICULUM

### PREPARATORY COURSE

#### DATA ANALYTICS IN EXCEL

Taught by one of the most renowned data scientists in the country (S.Anand, CEO, Gramener), this module takes you from a beginner level Excel user to an almost professional user.

#### ANALYTICS PROBLEM SOLVING

This module covers concepts of the CRISP-DM framework for business problem-solving.

### COURSE 1 - DATA TOOLKIT

#### BASICS OF SQL

Data in companies is definitely not stored in excel sheets! Learn the fundamentals of database and extract information from RDBMS using the structured query language.

#### INTRODUCTION TO PYTHON

Build a foundation for the most in-demand programming language of the 21st century.

#### PROGRAMMING IN PYTHON

Learn how to approach and solve logical problems using programming.

#### PYTHON FOR DATA SCIENCE

Learn how to manipulate datasets in Python using Pandas which is the most powerful library for data preparation and analysis.

#### VISUALIZATION IN PYTHON

Humans are visual learners and hence no task related to data is complete without visualisation. Learn to plot and interpret various graphs in Python and observe how they make data analysis and drawing insights easier.

#### IMDB MOVIE ASSIGNMENT

Reinforce the concepts learnt in data science through this rigorous assignment involving the past hundred years of movie data.

#### EXPLORATORY DATA ANALYSIS

Learn how to find and analyse the patterns in the data to draw actionable insights.

#### INFERENCE STATISTICS

Build a strong statistical foundation and learn how to 'infer' insights from a huge population using a small sample.

#### HYPOTHESIS TESTING

Understand how to formulate and validate hypotheses for a population to solve real-life business problems.

#### EDA CASE STUDY

Solve a real industry problem through the concepts learnt in exploratory data analysis.

### COURSE 2 - MACHINE LEARNING

#### INTRODUCTION TO MACHINE LEARNING AND LINEAR REGRESSION

Venture into the machine learning community by learning how one variable can be predicted using several other variables through a housing dataset where you will predict the prices of houses based on various factors.

#### LINEAR REGRESSION ASSIGNMENT

Build a model to understand the factors car prices vary on and help a Chinese company enter the US car market.

#### LOGISTIC REGRESSION

Learn your first binary classification technique by determining which customers of a telecom operator are likely to churn versus who are not to help the business retain customers.

#### UNSUPERVISED LEARNING: CLUSTERING

Learn how to group elements into different clusters when you don't have any pre-defined labels to segregate them through K-means clustering, hierarchical clustering, and more.

#### BUSINESS PROBLEM SOLVING

Learn how to approach open ended real world problems using data as a lever to draw actionable insights.

#### ASSIGNMENT: UNSUPERVISED + SUPERVISED

Apply the machine learning concepts learnt to solve a real-life predictive analytics problem.

#### CASE STUDY: LEAD SCORING

Help the Sales team of your company identify which leads are worth pursuing through this classification case study.

## SPECIALISATION 1: NLP

### COURSE 3 - MACHINE LEARNING II

#### TREE MODELS

Learn how the human decision making process can be replicated using a decision tree and other powerful ensemble algorithms.

#### MODEL SELECTION & GENERAL ML TECHNIQUES

Learn the pros and cons of simple and complex models and the different methods for quantifying model complexity, alongwith general machine learning techniques like feature engineering, model evaluation, and many more.

#### BAGGING AND BOOSTING

Learn about ensemble modelling through bagging and boosting and understand how weak algorithms can be transformed into stronger ones.

#### ADVANCED REGRESSION

In this module, take a more advanced look at regression models and learn the concepts related to regularization.

#### ADVANCED REGRESSION ASSIGNMENT

Build a regularized regression model to understand the most important variables to predict the house prices in Australia.

#### PRINCIPAL COMPONENT ANALYSIS

Understand important concepts related to dimensionality reduction, the basic idea and the learning algorithm of PCA, and its practical applications on supervised and unsupervised problems.

#### TIME SERIES ANALYSIS

In this module, you will learn how to analyse and forecast a series that varies with time.

#### TELECOM CHURN CASE STUDY

Solve the most crucial business problem for a leading telecom operator in India and southeast Asia - predicting customer churn.

### COURSE 4 - NATURAL LANGUAGE PROCESSING

#### LEXICAL PROCESSING

Do you get annoyed by the constant spams in your mail box? Wouldn't it be nice if we had a program to check your spellings? In this module learn how to build a spell checker & spam detector using techniques like phonetic hashing, bag-of-words, TF-IDF, etc.

#### SYNTACTIC PROCESSING

"Learn how to analyse the syntax or the grammatical structure of sentences with the help of algorithms & techniques like HMMs, Viterbi Algorithm, Named Entity Recognition (NER), etc."

#### SYNTACTIC PROCESSING ASSIGNMENT

Build a POS tagger for tagging unknown words using HMM's & modified Viterbi algorithm.

#### SEMANTIC PROCESSING

Learn the most interesting area in the field of NLP and understand different techniques like word-embeddings, LSA, topic modelling to build an application that extracts opinions about socially relevant issues (such as demonetisation) on social media platforms

#### CHATBOT CASE STUDY

Imagine if you could make restaurant booking without opening Zomato. Build your own restaurant-search chatbot with the help of RASA - an open source framework and deploy it on Slack.

### COURSE 5 - CAPSTONE PROJECT

#### CAPSTONE PROJECT

Choose from a range of real-world industry woven projects on advanced topics like Recommendation Systems, Fraud Detection, Emotion Detection from faces, Social Media Listening, Speech Recognition among many others.

## SPECIALISATION 2: DEEP LEARNING

### COURSE 3 - MACHINE LEARNING II

#### TREE MODELS

Learn how the human decision making process can be replicated using a decision tree and other powerful ensemble algorithms.

#### MODEL SELECTION & GENERAL ML TECHNIQUES

Learn the pros and cons of simple and complex models and the different methods for quantifying model complexity, alongwith general machine learning techniques like feature engineering, model evaluation, and many more.

#### BAGGING AND BOOSTING

Learn about ensemble modelling through bagging and boosting and understand how weak algorithms can be transformed into stronger ones.

#### ADVANCED REGRESSION

In this module, take a more advanced look at regression models and learn the concepts related to regularization.

#### ADVANCED REGRESSION ASSIGNMENT

Build a regularized regression model to understand the most important variables to predict the house prices in Australia.

#### PRINCIPAL COMPONENT ANALYSIS

Understand important concepts related to dimensionality reduction, the basic idea and the learning algorithm of PCA, and its practical applications on supervised and unsupervised problems.

#### TIME SERIES ANALYSIS

In this module, you will learn how to analyse and forecast a series that varies with time.

#### TELECOM CHURN CASE STUDY

Solve the most crucial business problem for a leading telecom operator in India and southeast Asia - predicting customer churn.

### COURSE 4 - DEEP LEARNING

#### INTRODUCTION TO NEURAL NETWORKS

Learn the most sophisticated and cutting-edge technique in machine learning - Artificial Neural Networks or ANNs

#### NEURAL NETWORKS ASSIGNMENT

Build a neural network from scratch in Numpy to identify handwritten digits.

#### SYNTACTIC PROCESSING ASSIGNMENT

Build a POS tagger for tagging unknown words using HMM's & modified Viterbi algorithm.

#### CONVOLUTIONAL NEURAL NETWORKS - INTRODUCTION AND INDUSTRY APPLICATIONS

Learn the basics of CNN and OpenCV and apply it to Computer Vision tasks like detecting anomalies in chest X-Ray scans, vehicle detection to count & categorise them to help the government ascertain the width and strength of the road.

#### RECURRENT NEURAL NETWORKS

Ever wondered what goes behind machine translation, sentiment analysis, speech recognition? Learn how RNN helps in these areas having sequential data like text, speech, videos, and a lot more.

#### GESTURE RECOGNITION

Make a Smart TV system which can control the TV with user's hand gestures as the remote control

### COURSE 5 - CAPSTONE PROJECT

#### CAPSTONE PROJECT

Choose from a range of real-world industry woven projects on advanced topics like Recommendation Systems, Fraud Detection, Emotion Detection from faces, Social Media Listening, Speech Recognition among many others.

## SPECIALISATION 3: BUSINESS ANALYTICS

### COURSE 3 - MACHINE LEARNING II

#### TREE MODELS

Learn how the human decision making process can be replicated using a decision tree and other powerful ensemble algorithms.

#### TIME SERIES ANALYSIS

In this module, you will learn how to analyse and forecast a series that varies with time.

#### RETAIL-GIANT SALES FORECASTING ASSIGNMENT

Apply the concepts learnt in time series to solve a forecasting problem for a retail giant.

#### MODEL SELECTION & GENERAL ML TECHNIQUES

Learn the pros and cons of simple and complex models and the different methods for quantifying model complexity, alongwith general machine learning techniques like feature engineering, model evaluation, and many more.

#### TELECOM CHURN CASE STUDY

Solve the most crucial business problem for a leading telecom operator in India and southeast Asia - predicting customer churn.

### COURSE 4 - BUSINESS REQUIREMENTS

#### ADVANCED SQL AND BEST PRACTICES

Apply advanced SQL concepts like windowing and procedures to derive insights from data and answer pertinent business questions

#### ADVANCED EXCEL

Learn the advanced concepts in Excel and start to perform data analysis like a pro!

#### SYNTACTIC PROCESSING ASSIGNMENT

Build a POS tagger for tagging unknown words using HMM's & modified Viterbi algorithm.

#### STRUCTURED PROBLEM SOLVING USING FRAMEWORKS

Learn how to attack a business problem using various structured frameworks like 5W, 5WHYS, and SPIN.

#### HYPOTHESIS FORMULATION

The module will equip you with a stepwise process for understanding a business problem and building hypotheses around it.

#### ASSIGNMENT

Apply your learnings from the course to solve a real-life business problem.

#### REVENUE AND OPERATIONAL COST MODELLING

Understand financial concepts such as revenue, cost of goods sold, profit, balance sheets, cash flow statements, and EBITDA. Also, get a brief introduction to economics concepts including supply and demand curves, cost curves, and a lot more!

#### EFFECTIVE COMMUNICATION STRATEGIES, FORMATS, AND TEMPLATES

Learn how to effectively strategise, communicate, and fine grain your data analysis projects.

#### PRESENTATIONS TO TECHNICAL AND NON-TECHNICAL STAKEHOLDERS

Understand how to optimally present your findings to technical and non-technical stakeholders and upgrade your storytelling skills.

#### BUSINESS CASE STUDY

Understand how a project in the industry is taken up and solved through a comprehensive business case study.

### COURSE 5 - CAPSTONE PROJECT

#### CAPSTONE PROJECT

Solve an end-to-end real-life industry problem from a wide variety of domains like Marketing, Retail, E-Commerce, Supply Chain, Healthcare, BFSI, and many more.



# SPECIALISATION 4: BUSINESS INTELLIGENCE/ DATA ANALYTICS

## COURSE 3 - DATA ANALYST TOOLKIT

### INTRODUCTION TO DATABASES

Learn how data is stored and which database is optimal to use in a particular scenario

### ADVANCED SQL AND BEST PRACTICES

Apply advanced SQL concepts like windowing and procedures to derive insights from data and answer pertinent business questions

### DATA ACQUISITION ASSIGNMENT

Reinforce the concepts learnt in SQL and solve this comprehensive assignment involving hundreds of queries.

### NOSQL DATABASES AND BEST PRACTICES

Take your knowledge of query languages a step further by learning about MongoDB - a NoSQL database which is becoming more and more popular in the industry.

### INTRODUCTION TO CLOUD AND HIVE

Understand the basics of cloud and learn about the Hive Query Language.

### SQL CASE STUDY

Understand how a project in the industry is taken up and solved through a comprehensive business case study.

## COURSE 4 - STORYTELLING WITH ADVANCED VISUALIZATION

### ADVANCED EXCEL

Learn the advanced concepts in Excel and start to perform data analysis like a pro!

### VISUALISATION USING TABLEAU

Learn advanced visualisation techniques using the most in-demand visualization tool in the industry.

### INTERACTIVE MARKETING CAMPAIGN ANALYSIS

Apply the new found Excel and Tableau skills to solve an exciting business assignment.

### VISUALISATION USING POWERBI

Take your visualization game a step forward by understanding how to operate PowerBI.

### INTRODUCTION TO R AND RSHINY

Get a brief introduction to another popular data science language and learn how to manipulate dataframes in R and learn to create attractive dashboards and web applications using ShinyR.

### EFFECTIVE COMMUNICATION STRATEGIES, FORMATS, AND TEMPLATES

Learn how to effectively strategise, communicate, and fine grain your data analysis projects.

### PRESENTATIONS TO TECHNICAL AND NON-TECHNICAL STAKEHOLDERS

Understand how to optimally present your findings to technical and non-technical stakeholders and upgrade your storytelling skills.

### BUSINESS CASE STUDY

Understand how a project in the industry is taken up and solved through a comprehensive business case study.

## COURSE 5 - CAPSTONE PROJECT

### CAPSTONE PROJECT

Solve an end-to-end real-life industry problem from a wide variety of domains like Marketing, Retail, E-Commerce, Supply Chain, Healthcare, BFSI, and many more.

# SPECIALISATION 5: DATA ENGINEERING

## COURSE 3 - DATA ENGINEERING I

### INTRODUCTION TO HADOOP AND MAPREDUCE PROGRAMMING

Understand the world of distributed data processing and storage with Hadoop. Learn to write MapReduce jobs in Python.

### DATA MANAGEMENT AND RELATIONAL DATABASE MODELLING

"Understand the concepts of Data Management and learn to model data from a Relational Database"

### NOSQL DATABASES AND APACHE HBASE

Learn the concepts of NoSQL databases. Understand the workings of Apache HBase.

### DATA WAREHOUSING (OPTIONAL)

Understand the intricacies behind designing a data warehouse and a data lake for your use case.

### DATA INGESTION WITH APACHE SQOOP AND APACHE FLUME

Get familiar with the challenges involved in data ingestion. Use Sqoop and Flume to ingest structured and unstructured data into Hadoop.

### BUILDING AND QUERYING DATA WAREHOUSE WITH APACHE HIVE

Manage and query a data warehouse with Apache Hive. Learn to write optimized HQL for large scale data analysis.

### CASE STUDY: INGESTION & WAREHOUSING

Make use of Sqoop, Flume, Hive and HBase to design an ETL data pipeline.

## COURSE 4 - DATA ENGINEERING II

### DATA PROCESSING WITH PYSPARK

Get introduced to Apache Spark, a lightning fast big data processing engine. Use PySpark to create large scale data processing applications.

### REAL-TIME DATA STREAMING WITH APACHE KAFKA

Understand the producer-consumer architecture of Apache Kafka. Learn to set up a Kafka cluster for managing real-time data.

### "REAL-TIME DATA PROCESSING USING SPARK STREAMING"

Learn about the real-time data processing architecture of Apache Spark. Build Spark Streaming applications to process data in real-time.

### ASSIGNMENT (OPTIONAL)

Use Kafka and Spark to develop a real-time data processing application.

### BUILDING AUTOMATED DATA PIPELINES WITH OOZIE/AIRFLOW

Automate your Data Pipeline with Apache Airflow

### ANALYTICS USING PYSPARK

Use PySpark to do EDA and Predictive Analysis of large datasets.

### CASE STUDY: KAFKA, SPARK STREAMING AND PYSPARK

Build an end-to-end real-time data processing application using Spark Streaming and Kafka

## COURSE 5 - CAPSTONE PROJECT

### CAPSTONE PROJECT

The capstone project will stitch all the components of data engineering together.

# SUCCESS STORIES



“

The assignments and case studies are the ones that will really give you a mileage during the job interviews. Through upGrad and IIIT-B career support, I have secured an offer from an MNC. I have learnt quite a lot from peer discussions. All-in-all upGrad and IIIT-B program can really help you transition to data science.

- **Antarik, Network Engineer, USEReady**



“

I was skeptical even after my admission in the upGrad & IIIT-B PGDDS program thinking it is an online pilot program. But I soon realised otherwise. This course covers deep concepts in DA and develops confidence in topics through challenging assignments. Team upGrad is highly reachable and very prompt in addressing course-related concerns. Managing my work and concentrating on the course has never felt like a burden because of how thoughtfully the program is developed for working professionals.

- **Kishore M, Sr. Associate | Projects, Cognizant**



“

The peer-to-peer interactions help a lot. It is brilliant to collaborate and learn with others even though it's an online program. IIIT-B is a good brand to have and the curriculum of the program is solid.

- **Sajal Roychowdhury, Risk Analyst, Amazon India**

## PROGRAM STARTS

Please refer to the [website](#) for program start dates

## DURATION

12 months

## PROGRAM FLOW

**5-6 months** - Learn about data science and machine learning along with the necessary tools.

**5-6 months** - Specialize in one track according to the hottest data roles in the industry.

## WEEKLY COMMITMENT

12-15 hours per week

4-7 hours of asynchronous learning time

4-7 hours assignments and projects

1 live session every 2 weeks

## PROGRAM FEE

₹2,85,000 (Incl. of all taxes)

Flexible Payment Options Available

## ELIGIBILITY

Bachelor's/Master's degrees with a minimum of 50% marks in graduation

## SELECTION PROCESS

Selection test to assess quantitative and analytical reasoning

For further details contact us:



**KAVERI SHIVKAR**

Chief Admissions Counsellor

kaveri.s@upgrad.com

+91-9372201982



**ABDUL SAYED**

Chief Admissions Counsellor

abdul.sayed@upgrad.com

+91-9324900862

**upGrad**

### COMPANY INFORMATION

upGrad Education Private Limited  
Nishuvi, 75, Dr. Annie Besant Road  
Worli, Mumbai - 400018