EMERITUS

POSTGRADUATE DIPLOMA IN
APPLIED DATA SCIENCE

In collaboration with

COLUMBIA | ENGINEERING
EXECUTIVE EDUCATION

OVERVIEW

> Use of analytics is accelerating, and that means more data-driven decision making and fewer hunches. Evidence-based management complements analytics by adding validated cause-and-effect relationships between policies and effects.
>
> — Paul Gibbons
> Author and business scholar

Managing a business or organization in today's world is more science than art. At the core of that science is data, and the ability to unleash its power and extract value from it is critical to any company's success.

Data science and machine learning have transformed entire industries and continue to do so. The data revolution has led to a spike in the demand for data scientists and machine learning practitioners that shows no signs of slowing down.

The Postgraduate Diploma in Applied Data Science is designed to help participants master data science, from the critical foundations of statistics and probability to working hands-on with machine learning models using the world's most popular programming language.

Analytical models are more powerful when they are built with the right statistics, and this comprehensive diploma can help you learn the key statistics and probability concepts to build effective models, enhance your data interpretation skills and make well-informed decisions.

FACULTY

Vineet Goyal

Assistant Professor
Industrial Engineering and Operations Research

Professor Vineet Goyal has a Bachelor's degree in Computer Science from Indian Institute of Technology, Delhi and a Ph.D. from Carnegie Mellon University. Before coming to Columbia, he spent two years as a Postdoctoral Associate at the Operations Research Center at MIT. Professor Goyal is interested in the development of tractable approaches for dynamic optimization problems under uncertainty and their applications in electricity markets, revenue management and supply-chain and inventory management.

Hardeep Johar

Senior Lecturer of Industrial Engineering and Operations Research

Hardeep Johar received an M.A. in Economics from the Birla Institute of Technology and Science and is a Fellow of the Indian Institute of Management Calcutta. He received a Ph.D. in Information Systems from the Stern School of Business, New York University. Prior to joining Columbia, Johar has worked as a quantitative trader at Morgan Stanley, Credit Suisse and Deutsche Bank, at a tech startup (MSpoke), and has taught at NYU Stern School of Business and the Gabelli School of Business at Fordham University.

COURSE LEADER

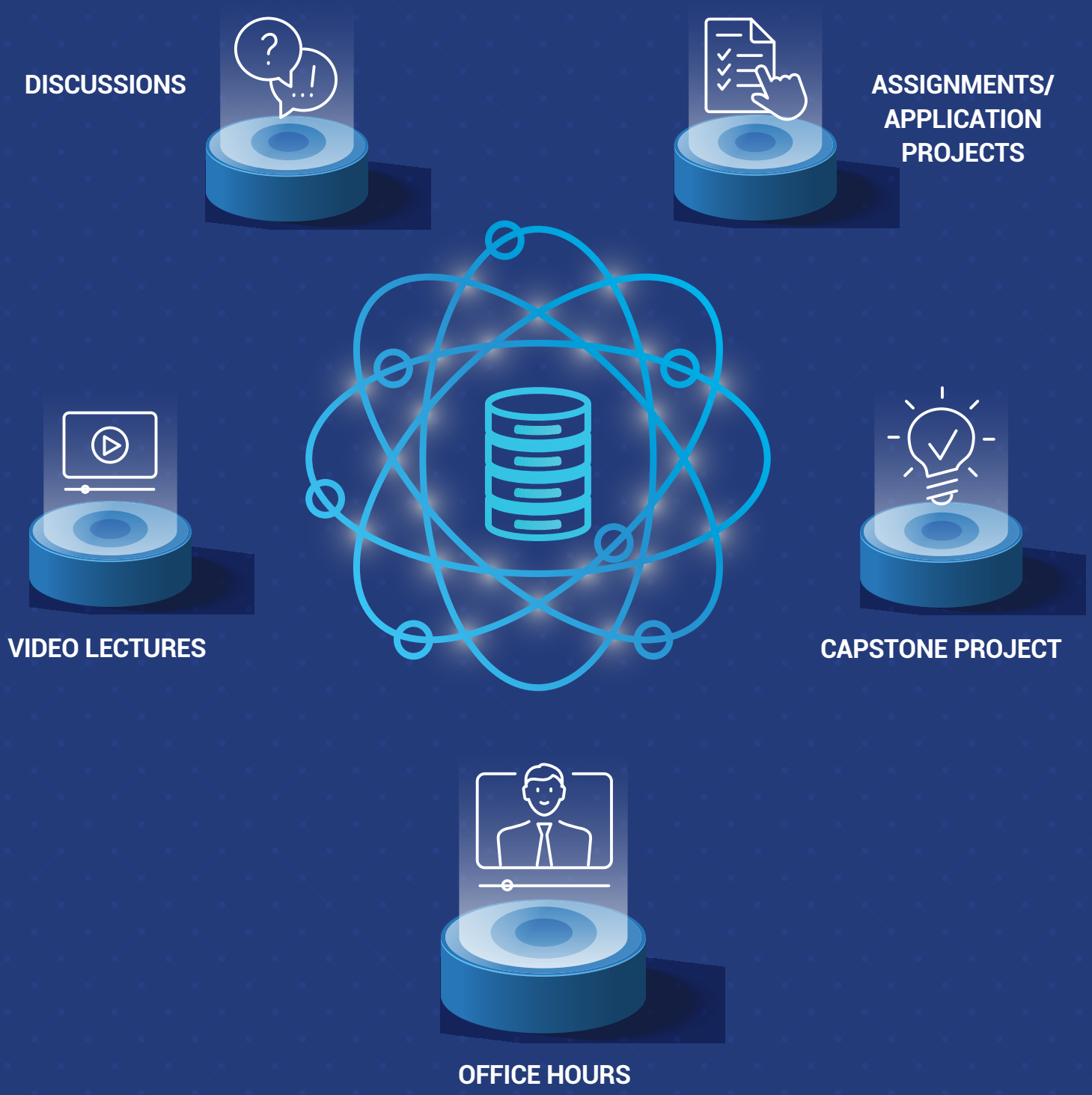
Phil Capobres

Course Leader, Emeritus

Phil has over 33 years' experience in Financial Services and Telecommunications in the areas of data analytics, marketing and finance. Most recently he was Director of Data Analytics and Insights for the Institutional Division of Fidelity Investments. Phil holds a BS in Quantitative Business Analysis from Arizona State University and an MBA in Industrial Management from the University of Dallas.

*Course Leaders are subject to change

HIGHLIGHTS


DISCUSSIONS
ASSIGNMENTS/ APPLICATION PROJECTS
VIDEO LECTURES
CAPSTONE PROJECT
OFFICE HOURS

ANALYZE. PREDICT. DECIDE. DO IT BETTER WITH DATA.

Improved statistical thinking for better analytical models and decision-making

In-depth foundational knowledge to ensure the right skills and practice

Mastery of three essential pillars of data science: Statistics, Analytics and Python

Going beyond the theory, our approach invites participants into a conversation, where learning is facilitated by Office Hours with the Course Leader.

WHAT YOU WILL LEARN

- Import data using common Python packages and conduct simple statistical analysis on your data.
- Manage relational databases (stored in tables) by accessing and querying required data using SQL commands.
- Use relevant tools and mechanisms (JSON, XML, APIs and BigSoup) to extract data from the internet and perform required analysis.
- Apply data distributions and sampling techniques to break down data into samples and represent it in multiple forms.
- Identify the different factors involved in confidence levels and assess how you may use it in hypothesis testing effectively.
- Use advanced Python libraries, such as NumPy and Pandas, to analyze discrete data sets and visualize data for better usability.
- Analyze data, form connections, and make informed data-driven decisions by using decision trees, random forests, and ensemble methods.
- Apply K-means clustering and text mining methods to classify data and organize them into categories for effective analysis.
- Identify situations where advanced methods, such as, neural networks and time-series analysis can be used to perform data analysis using multiple components and variables.

SYLLABUS

MODULE 1: TOOLS AND DATA MANAGEMENT

Python Basics - How to Translate Procedures into Code

Python data types (basic and Boolean), conditional statements, functions, assignment operations

Intermediate Python - Data Structures for Analysis

Lists, dictionaries, mutability, and iterations with examples on data structures

Relational Databases - Where Big Data Is Typically Stored

Basics of databases and normalization

SQL - Ubiquitous Database Format/Language

Using SQL for Python, SQL workbench, working with multiple tables

Data Extraction - Getting Data from the Internet - Part 1

Extracting data from the web using JSON, Google API, and XML

Data Extraction - Getting Data from the Internet - Part 2

Using the Beautiful Soup mechanism to extract data, the Epicurious example

---

MODULE 2: STATISTICS AND EXPLORATORY DATA ANALYSIS

Statistical Distributions - The Shape of Data

Types of distributions: Normal (examples), Poisson, Geometric, Exponential, Lognormal, and Bernoulli

Sampling - When You Can't or Won't Have ALL the Data

Size and sampling techniques, central limit theorem and motivation, sample size distribution (fixed sample size), polling techniques (given sample size, given target accuracy), estimating proportions

Hypothesis Testing - Answering Questions About Your Data

Calculating and interpreting confidence levels, t-tables and t-multipliers, determining P-values and A/B testing example

Data Analysis and Visualization - using Python's NumPy for analysis

Introduction to using Numpy and Pandas for data visualization, Pandas datareader, time-series analysis, risk return analysis, regression

Data analysis and visualization - using Python's Pandas for Data Wrangling

Data cleaning and data visualization using Pandas, using the groupby function to organize data

---

MODULE 3: FUNDAMENTALS OF MACHINE LEARNING

Machine Learning - Basic Regression and Classification

Machine learning using wines dataset and rocks and mines dataset, classification metrics, classification metrics using rocks and mines dataset

Linear Regression

Introduction to linear regression, using dummy variables in regression, measuring outputs of regression, making predictions with regression, collinearity, overfitting and how to prevent it

Logistic Regression

Introduction to regression, classification problems, and building a logistic regression model, and practice

Machine Learning - Decision Trees and Clustering

Understanding decision trees – example and visualization, regression trees (using the wines dataset), classification trees (rocks and mines dataset)

Ensemble Methods

Decision trees, bagging and boosting concepts, feature importance, and hyperparameter tuning

Naïve Bayes Classifiers

Discrete and conditional probabilities, Baye's theorem, spam filtering, and practicet

Neural Networks

Neural networks in keras, the perceptron, real-life examples: movie review classification and predicting housing prices

K-means Clustering

Unsupervised models, k-means clustering models and examples, Gaussian mixtures and examples

Dimensionality Reduction

Data projections, dimensionality reduction (DR), other DR techniques, principal component analysis

Text Mining - Automatic Understanding of Text

Text mining techniques: sentiment analysis, complexity analysis, and named entity analysis, text summarization, and topic modelling techniques

Time Series Analysis

Datetime and introduction to time series, exploring time series, descriptive statistics, partial autocorrelation, autoregressive models, the ARIMA model

---

MODULE 4: CAPSTONE PROJECT

Acting in the role of consultant, test the efficacy of an office supply company's telemarketing campaigns for a select audience and help them leverage the test results to their advantage.

PREREQUISITES

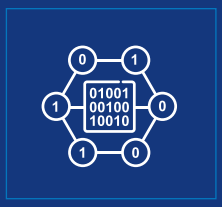
The diploma requires an undergraduate knowledge of statistics, (descriptive statistics, regression, sampling distributions, hypothesis testing, interval estimation etc.) linear algebra and probability. You would be required to possess a knowledge of programming concepts like variables, loops, functions, OOP etc.

Some hands on knowledge with Python Language and Jupyter Notebook IDE will be necessary. All assignments/application projects will be done in Jupyter Notebooks using the Python programming language. Emeritus offers a complimentary Python for Data Analytics certificate course to meet this prerequisite. Participants who successfully complete this certificate course will receive a certificate of completion from Emeritus Institute of Management.

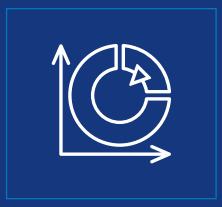
APPLICATION ASSIGNMENTS

Data extraction & web scraping using Beautiful Soup

Data distribution & sampling using office supply data



Hypothesis testing



Data analysis & visualization using lending club data

Linear regression using house pricing data

Machine learning classification using handwritten digits data

Decision trees using Abalone data

Text mining using Amazon review data

Unsupervised ML using Iris data

Time series analysis using Amazon stock prices

THE LEARNING EXPERIENCE

The diploma is spread over 38 weeks and consists of Video Lectures, Assignments/Application Projects, Discussions, and a Capstone Project.

Orientation
The first week is orientation week. During this week you will be introduced to the other participants in the class from across the world and you will learn how to use the learning management system and other learning tools provided.

Goal Setting
You have learning goals set throughout the duration of the diploma, including watching the video lectures and completing the assignments. All assignments have weekly deadlines.

Recorded Video Lectures
The recorded video lectures are by faculty from the collaborating university.

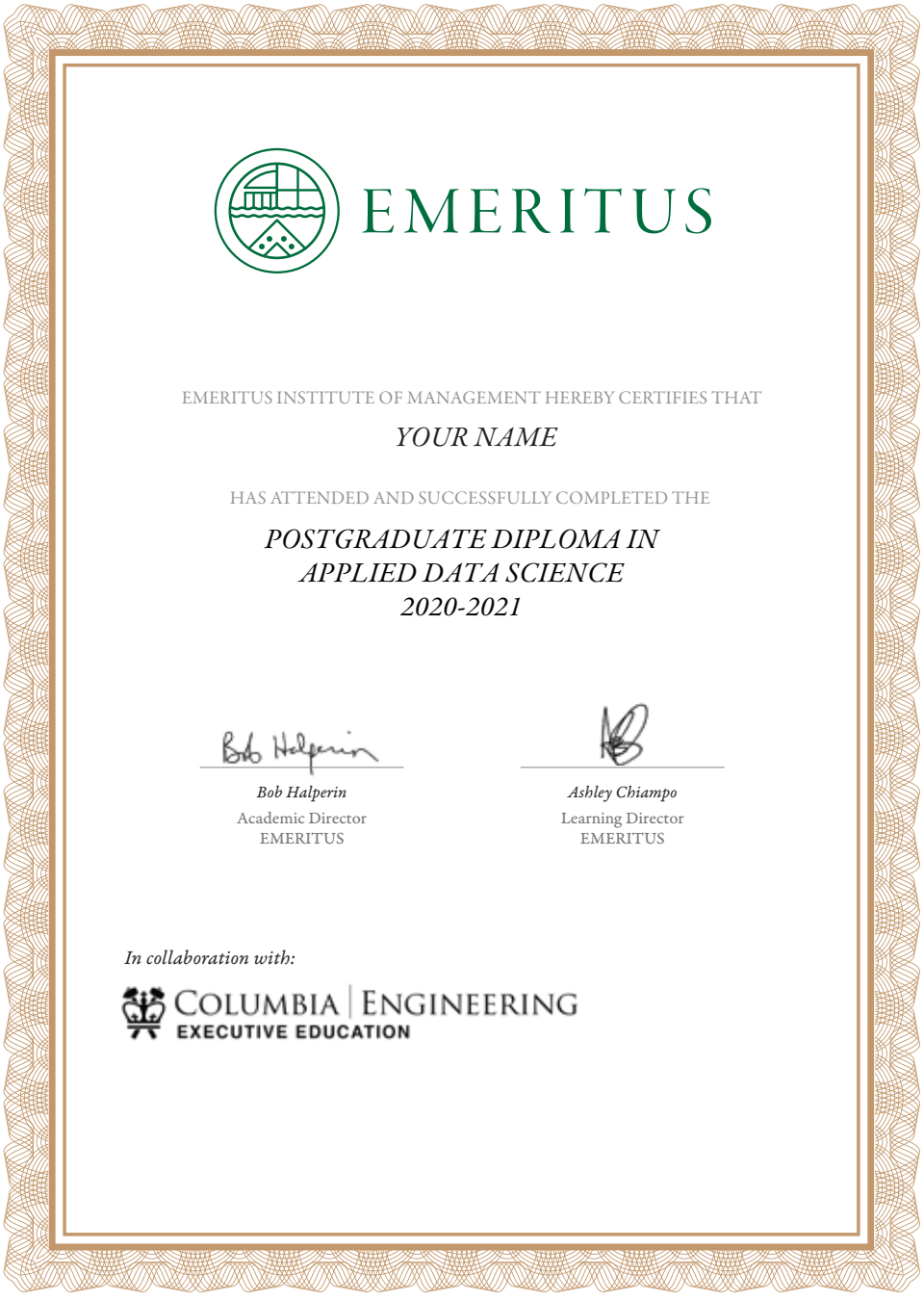
Clarifying Doubts
In addition to the live webinars, for some diplomas, the Course Leaders conduct Office Hours, which are webinar sessions that are open to all learners. During Office Hours, learners ask questions and Course Leaders respond. These are usually conducted every alternate week to help participants clarify their doubts pertaining to the content.

Follow-Up
Emeritus Program Support Team follow-up over email and phone calls with learners who are unable to submit their assignments on time.

Continuous Course Access
You will continue to have access to the diploma videos and learning material for up to 12 months from the diploma end date.

CERTIFICATE

Upon successful completion of the diploma, participants will receive a verified digital diploma from Emeritus Institute of Management.



EMERITUS

EMERITUS INSTITUTE OF MANAGEMENT HEREBY CERTIFIES THAT

YOUR NAME

HAS ATTENDED AND SUCCESSFULLY COMPLETED THE

POSTGRADUATE DIPLOMA IN
APPLIED DATA SCIENCE
2020-2021

Bob Halperin
Academic Director
EMERITUS

Ashley Chiampo
Learning Director
EMERITUS

In collaboration with:

COLUMBIA | ENGINEERING
EXECUTIVE EDUCATION

PARTICIPANT PROFILE

WHAT WE ARE LOOKING FOR

Our participants demonstrate the following qualities:

- Leadership initiative - going beyond their defined scope of work
- An ability to “make things happen;” strong execution skills
- A visionary with desire to shape the future
- Multicultural experience and sensitivity
- Ambitious and passionate learner
- At least three years of professional experience

APPLICATION REQUIREMENTS

Applicants will be required to:

- Submit a completed Application Form
- Submit a current resume
- Submit proof of diploma/degree in any field of study (highest qualification should be submitted)
- Have completed 21 years of age on or before start of the diploma

ENGLISH LANGUAGE PROFICIENCY REQUIREMENT

All candidates who have received their bachelor's or other degree or diploma from an education institution where English is NOT the primary language of instruction are required to demonstrate English language proficiency through ANY ONE of the following methods

- Obtain a TOEFL minimum score of 550 for the paper based test or its equivalent
- Obtain an IELTS minimum score of 6.0
- Obtain a Pearson Versant Test minimum score of 59
- Obtain a Certificate of Completion for a Certificate course offered by Emeritus
- Submit a document which shows that the candidate has, for the last 24 months or more, worked in ANY ONE of these countries: Antigua and Barbuda, Australia, The Bahamas, Barbados, Belize, Canada, Dominica, Grenada, Guyana, India, Ireland, Jamaica, New Zealand, Singapore, South Africa, St Kitts and Nevis, St Lucia, St Vincent and the Grenadines, Trinidad and Tobago, United Kingdom, United States of America

DURATION

9 Months, Online

APPLICATION DEADLINE

Round 1 (Early bird)
August 24, 2020

Round 2
September 28, 2020

DIPLOMA COMMENCES

September 30, 2020

PAYMENT DETAILS

FEES

$3000 (payable in 2 equal installments)

Application Fee: (Non-Refundable) $50

Note: Singapore residents who wish to enroll for this diploma will be charged GST.

EARLY BIRD TUITION ASSISTANCE

Apply before August 24, 2020 and avail
an early bird tuition assistance of
$400

Use code **PGDDS0920EB** while applying to avail the same

BENEFITS TO THE LEARNER

Deepen Your Intellectual Capital

- World-class curriculum from Columbia Engineering Executive Education
- Selective entrance criterion ensures you learn with the best

Manage Your Brand Capital

- A global education on your resume
- Participants scoring 90% or above in all courses, and 'Exceeds Expectations' on the Capstone assignment will be awarded 'Emeritus Scholars' status

Enhance Your Social Capital

- Make new, life-long connections with experienced business people from a wide variety of cultures, industries and backgrounds
- Inclusion in the Emeritus Network
- Invitations to Emeritus Network events globally including career panels, CXO speaker series, and industry interactions

Enhance Your Career Capital

- Learn while you earn
- Get noticed. Get ahead.
- Understand how to manage your career and personal brand

EMERITUS NETWORK

On successful completion of this program, join a community of learners on the Emeritus Network. The Emeritus Network is your platform to connect to a global network of individuals. Benefits of the Emeritus Network include:

- Tuition assistance
- Global network that includes over 400 CEOs, presidents, vice presidents, directors, founders, and managing directors
- Start-up corner to help connect, collaborate, raise capital, invest, or identify talent

50,000+

PARTICIPANTS

160+

COUNTRIES

70+

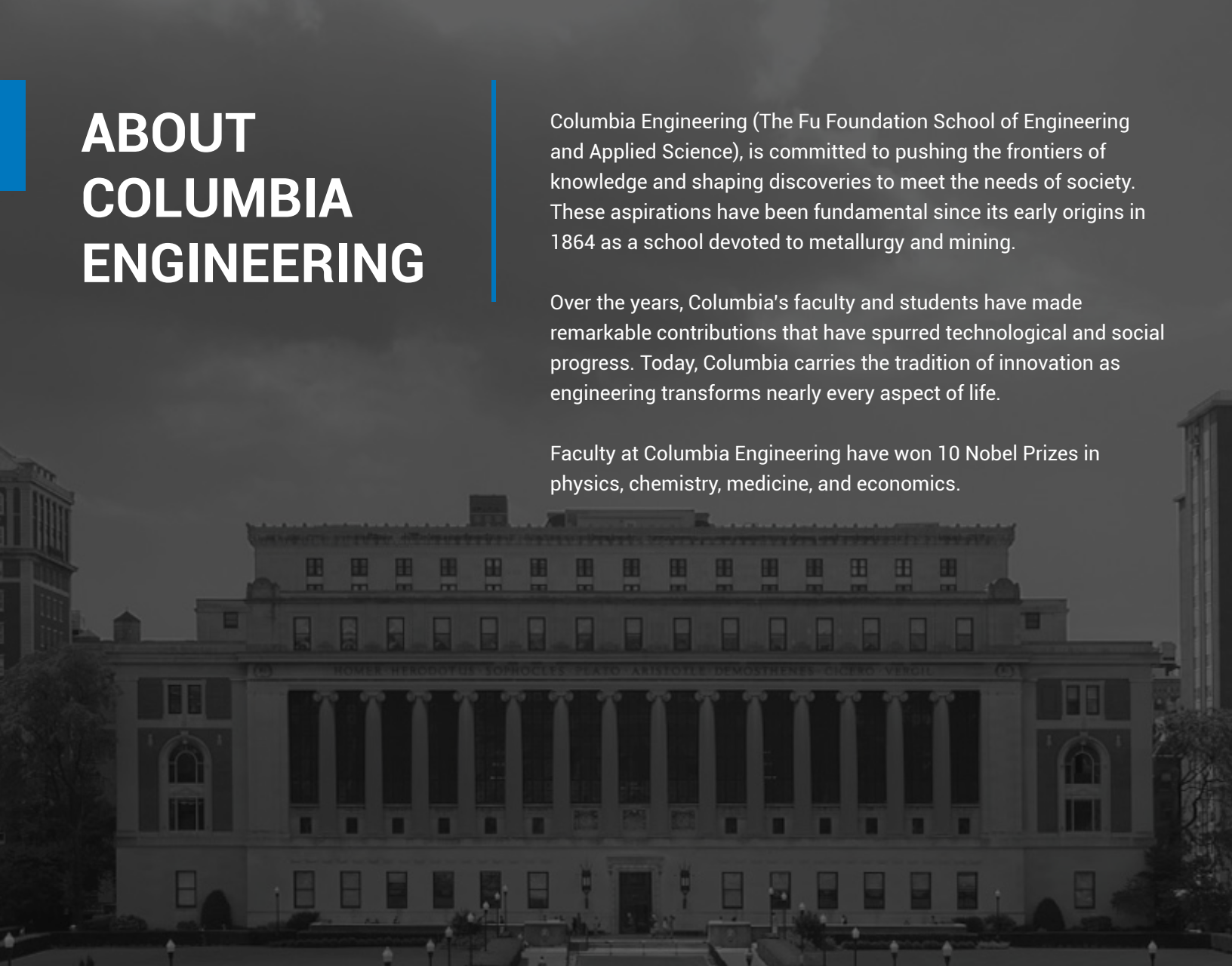
PROGRAMS

ABOUT COLUMBIA ENGINEERING

Columbia Engineering (The Fu Foundation School of Engineering and Applied Science), is committed to pushing the frontiers of knowledge and shaping discoveries to meet the needs of society. These aspirations have been fundamental since its early origins in 1864 as a school devoted to metallurgy and mining.

Over the years, Columbia's faculty and students have made remarkable contributions that have spurred technological and social progress. Today, Columbia carries the tradition of innovation as engineering transforms nearly every aspect of life.

Faculty at Columbia Engineering have won 10 Nobel Prizes in physics, chemistry, medicine, and economics.



ABOUT EMERITUS

Columbia Engineering Executive Education is collaborating with online education provider Emeritus to offer executive education courses through a dynamic, interactive, digital learning platform. These courses leverage Columbia Engineering Executive Education's thought leadership in Engineering practice developed over years of research, teaching and practice.

An Emeritus Certificate course created in collaboration with Columbia Engineering Executive Education is based on syllabus approved by Columbia Engineering Executive Education, and contains video content created and recorded by Columbia Engineering Executive Education faculty, combined with assessments, assignments, projects, cases, and exercises delivered by Emeritus. Upon successful completion of the course, learners will be awarded a certificate jointly by Emeritus and Columbia Engineering Executive Education.

An Emeritus Postgraduate Diploma contains multiple Emeritus Certificate courses created in collaboration with Columbia Engineering Executive Education, and may also include courses created independently by Emeritus. Upon successful completion, students will be awarded a Postgraduate Diploma by Emeritus.

www.emeritus.org

Schedule a call with a Program Advisor to learn how this diploma can help you

SCHEDULE A CALL

Apply for the diploma here

APPLY NOW

E-mail: info@emeritus.org
Call: +1 315 871 0015
We attempt to respond to queries in 24 hours or less. However, over weekends and holidays, our responses may take up to 72 hours.

© Copyright 2020 Emeritus. All Rights Reserved.
Emeritus | Council for Private Registration No. 201510637C | Period: 29 March 2018 — 28 March 2022