



# Data Science & Big Data

A part-time course to train the new generation of data scientists.



[www.ub.edu/datascience](http://www.ub.edu/datascience)

Provisional Wireless Network  
[wifi.ub.edu](http://wifi.ub.edu)

Usuari: **agvzhh.tmp**

Paraula clau: **bkiq78**

Data d'inici de validesa	2017/10/03
Data de finalització de validesa (no inclosa en el període de validesa)	2017/10/18

# GDC@UB



DATA SCIENCE  
PROGRAMS AND  
COURSES AT  
UNIVERSITY OF  
BARCELONA



MASTER  
OF  
FOUNDATIONS  
OF DATA  
SCIENCE



DEEP  
LEARNING  
SUMMER  
COURSE



DATA  
SCIENCE  
AND BIG  
DATA  
COURSE

# jordi.vitria@ub.edu

Departament de Matemàtiques i Informàtica de la Universitat de Barcelona



Since 2007, I am a Full Professor at the Mathematics & Computer Science Department, **Universitat de Barcelona**. Before that I spent 20 years on the faculty of the CS Department at the **Universitat Autònoma de Barcelona**. I am the Director of the **Data Science & Big Data Postgraduate Course** at UB. I am the leader of the **DataScience@UB** group, whose objective is to promote technology transfer.

**My research statement:** To understand the fundamental processes underlying the visual perception of objects to derive new computer vision algorithms that, one day, may enable machines to see.

# Instructors



Laura Igual



Francesc Dantí



Oriol Pujol

Jordi Vitrià



Lluis Garrido



Jordi Nin



Albert Diaz-Guilera



Eloi Puertas



Sergio Escalera



Santi Seguí



Josep Perelló



Montse Guillen



UNIVERSITAT DE  
BARCELONA

with the collaboration of



BBVA



## Introduction

Universitat de Barcelona's **Data Science and Big Data** course offers students a program that covers the concepts and tools you will need throughout the entire data science pipeline: asking the right questions; wrangling and cleaning data; generating hypothesis; making inferences; visualizing data; assessing solutions; and building data products.

## Schedule

Every Tuesday 18h-20h and Thursday 18h-20h

## Workload

8 hours per week (including lectures and homework)

## Location

Aula T1, Facultat de Matemàtiques i Informàtica,  
Edifici Històric de la Universitat de Barcelona,  
Gran Via de les Corts Catalanes 585, 08007, Barcelona.

# Calendar & Content

Date	Session
dt. 3 d'oct. 2017	Introduction, Jordi Vitrià
dj. 5 d'oct. 2017	Python, Jordi Vitrià
dt. 10 d'oct. 2017	Software Carpentry, Eloi Puertas
dj. 17 d'oct. 2017	Software Carpentry, Eloi Puertas
dj. 19 d'oct. 2017	Kick-Off Project + MasterClass1 - KING
dt. 24 d'oct. 2017	Data Gathering, Oriol Pujol
dj. 26 d'oct. 2017	Data Science Toolbox, Eloi Puertas
dt. 31 d'oct. 2017	Data Gathering, Oriol Pujol
dj. 2 de nov. 2017	Data Science Toolbox, Eloi Puertas
dt. 7 de nov. 2017	Data Cleaning, Sergio Escalera
dj. 9 de nov. 2017	Data Cleaning, Sergio Escalera
dt. 14 de nov. 2017	CAPSTONE PROJECT
dj. 16 de nov. 2017	MasterClass Satellogic
dt. 21 de nov. 2017	Computational Statistics, Petia Radeva
dj. 23 de nov. 2017	Computational Statistics, Petia Radeva
dt. 28 de nov. 2017	Statistical Estimation, Jordi Vitrià
dj. 30 de nov. 2017	Statistical Estimation, Jordi Vitrià
dl. 4 de des. 2017	MasterClass SOCIALPOINT
dt. 12 de des. 2017	Regression, Laura Igual
dj. 14 de des. 2017	Regression, Laura Igual
dt. 19 de des. 2017	Montse Guillen
dj. 21 de des. 2017	Montse Guillen
dt. 9 de gen. 2018	Bayesian Estimation, Jordi Vitrià
dj. 11 de gen. 2018	Supervised Learning, Oriol Pujol
dt. 16 de gen. 2018	Supervised Learning, Oriol Pujol
dj. 18 de gen. 2018	Supervised Learning, Oriol Pujol
dt. 23 de gen. 2018	Supervised Learning, Oriol Pujol
dj. 25 de gen. 2018	Unsupervised Learning, Petia Radeva
dt. 30 de gen. 2018	Unsupervised Learning, Petia Radeva
dj. 1 de febr. 2018	Capstone Project
dt. 6 de febr. 2018	Recommenders, Santi Seguí
dj. 8 de febr. 2018	Recommenders, Santi Seguí
dt. 13 de febr. 2018	Multicore computing, Lluís Garrido
dj. 15 de febr. 2018	Multicore computing, Lluís Garrido
dt. 20 de febr. 2018	Cloud Computing, Francesc Dantí
dj. 22 de febr. 2018	Big Data Analytics: Spark, Jordi Nin
dt. 27 de febr. 2018	MasterClass BBVA Data&Analytics
dj. 1 de març 2018	Big Data Analytics: Spark, Jordi Nin
dt. 6 de març 2018	Graph Analysis, Laura Igual
dj. 8 de març 2018	Big Data Analytics: Spark, Jordi Nin
dt. 13 de març 2018	Graph Analysis, Laura Igual
dj. 15 de març 2018	Complex Networks, Albert Diaz-Gilera
dt. 20 de març 2018	Data Crowdsourcing, Josep Perelló
dj. 22 de març 2018	MasterClass KernelAnalytics
dt. 3 d'abr. 2018	Visualization, Santi Seguí
dj. 5 d'abr. 2018	Visualization, Santi Seguí
dt. 10 d'abr. 2018	Deep Learning
dj. 12 d'abr. 2018	Deep Learning
dt. 17 d'abr. 2018	Capstone Project
dj. 24 de maig 2018	Capstone Project



UNIVERSITAT DE  
BARCELONA

## Requirements:

The program is specially designed for students with a background in computer science, mathematics, and applied statistics, but other scientific and engineering backgrounds can be considered.

We will require to follow lessons and complete class exercises using personal laptops. You will not be able to complete all your assignments in class if you rely solely on desktop equipment at home.

## Information:

Campus Virtual UB (after **official enrollment**):

[campusvirtual2.ub.edu](http://campusvirtual2.ub.edu)

# Raw Data

## 1. Processing: How I do clean and separate my data?

- Identification: filter data.
- Outliers.
- Imputation: missing value processing.
- Reduction: dimensionality reduction.
- Normalization: duplicates, ranges, format, coordinates, units, etc.
- Feature extraction.

## 3. Enrichment: How do I add more information to my data?

- Feature engineering.
- Search for additional data sources.

## 5. Analyze: How do I model my data?

- Variable selection (How do I determine important variables?)
- Probabilistic modeling (How are my variables related?)

## 7. Evaluate: Are the outcomes generic and robust?

- Statistical Testing.
- Model performance.

Data Science Path

## 2. Acquire: How do I get my data?

- Web Scraping.
- Data Base queries.
- Access to bulk data stores.

## 4. Aggregation: How do I collect and summarize my data?

- Basic Statistics: mean, std, box plots, scatter plots, counts, etc.
- Distribution fitting.
- Feature aggregation.

## 6. Discover: What are the key relationships in my data?

- Clustering (How do I segment the data to find natural groupings?)
- Visualization (Are there unexpected relationships?)

## 8. Predict: What are the likely future outcomes?

- Regression (How do I predict the future?)
- Classification (How do I predict a category?)
- Recommendation (How do I predict relevant conditions?)

Insights

## Activities

**twitter:** @datascienceub | @databeersbcn



## Evaluation

**Capstone Project:** An important part of the course is the IPython process notebook. This notebook details your steps in developing a solution to a real problem, including how you collected the data, alternative solutions you tried, describing statistical methods you used, and the insights you got.

## Master Classes



<https://github.com/DataScienceUB/Postgrau>

The screenshot shows the GitHub repository page for 'DataScienceUB / Postgrau'. The repository has 18 commits, 1 branch, 0 releases, and 1 contributor. The latest commit was made 24 seconds ago. The repository contains files named 'algorimes', 'images', 'README.md', and 'README.md'. A large callout box highlights the repository's purpose: 'Data Science and Big Data Postgraduate Course at UB'. The box states: 'Data Science and Big Data course offers students a program that covers the concepts and tools you will need throughout the entire data science pipeline: asking the right questions; wrangling and cleaning data; generating hypothesis; making inferences; visualizing data; assessing solutions; and building data products.'

This repository has 18 commits, 1 branch, 0 releases, and 1 contributor.

Branch: master ▾ New pull request Create new file Upload files Find file Clone or download ▾

algorimes Update README.md

Latest commit 5b812d6 24 seconds ago

images Add files via upload 3 days ago

README.md Update README.md 23 seconds ago

README.md

## Data Science and Big Data Postgraduate Course at UB

Data Science and Big Data course offers students a program that covers the concepts and tools you will need throughout the entire data science pipeline: asking the right questions; wrangling and cleaning data; generating hypothesis; making inferences; visualizing data; assessing solutions; and building data products.