UNIVERSITAT DE BARCELONA

Introduction to Data Science

Eloi Puertas Prats

Data

We are rendering into **data** many aspects of the world that have never been quantified before.

Digital Transformation in many aspects in our lifes and work places enhances this process of **datification**.

Data everywhere

Where **Information** comes from?

- Corporate Data Bases (structured information).
- Unstructured information in documents, Wikipedia, textbooks, journals, blogs, tweets, etc.
- Images in the web, public cameras, phones, TV, YouTube, etc.
- Public APIs: smart cities, government, search engines, etc.
- Sensor Data: GPS, accelerometer, physicochemical sensors, sociometric sensors, supercolliders, telescopes, etc.

How to handle such amount of data?

Data Science

Data Science is a **methodology** to define:

- what we want to do with data,
- how do we evaluate our actions,
- what decisions can be grounded on data,
- how do we combine evidences from several sources.

Data Science Skills V1.0

Drew Conway's Data Science Venn Diagram

Data Science V2.0

Copyright © 2014 by Steven Geringer Raleigh, NC.
Permission is granted to use, distribute, or modify this
image, provided that this copyright notice remains intact

Data Science as a Team

- **Data Scientist** (A.k.a Statisticians, Data Managers)
- **Data Engineer** (A.k.a Data Managers, Database Administrators)
- **Data Analyst** (A.k.a Bussiness analysts)

https://www.springboard.com/blog/data-science-career-paths-different-roles-industry/

Data Science Project Steps

Project Steps:

1. Ask a question.
2. Get the data from a source. Data can be heterogeneous and non structured
3. Data Processing (cleaning, ETL.).
4. Data Analysis (machine learning, statistics…).
5. Take a decision and act.

Why are you doing it?

A Business View is needed!

Data Science pipeline

Homework

- Start thinking about a question to be answered with data
- Think about how to get such data
- Find a team! (Deadline 1/10)
- Define your project proposal! (Deadline 14/10)