



Introduction to the course

GRA 4142 Data Management and Python Programming

Jan Kudlicka (jan.kudlicka@bi.no)

Department of Data Science and Analytics

August 27, 2021

Hello and welcome!
Hei og velkommen!

Instructor

Jan Kudlicka

jan.kudlicka@bi.no

B3Y-096

Office hours

My door is always open for students wishing to learn!

Due to the pandemics, we will probably have to use email and Zoom.

TA

Joakim Myhre Rolfsøn

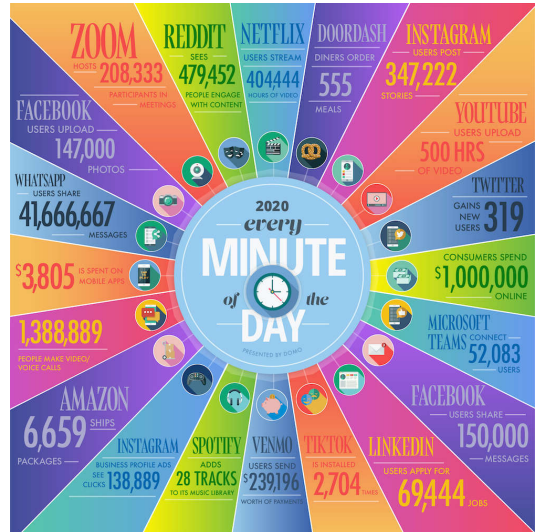
joakim.m.rolfson@bi.no

Data never sleeps!

How much data do we generate every minute?

Video from 2017:
<https://player.vimeo.com/video/226925416>

Source: Domo



With so much data around us we need a systematic approach to

- ▶ collect
- ▶ store
- ▶ process
- ▶ query
- ▶ analyze
- ▶ visualize

data, and in turn **to extract the knowledge and insights** from it.

Part I – first half of the course

Python programming with focus on data analysis

- ▶ Basic data types, conditionals, loops, functions, lists, dictionaries, comprehensions, working with text files
- ▶ NumPy and Pandas data structures, working with data files, data cleaning and preparation, merging, aggregation, visualization

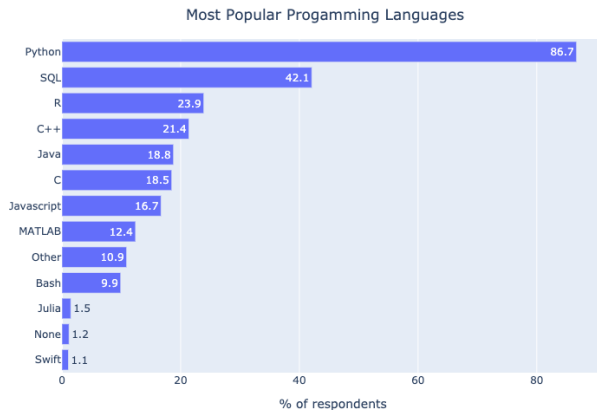
Part II – second half of the course

Data management with focus on relational databases and SQL

- ▶ SQL language
- ▶ Data modeling, normalization, indexes, transactions, programming with databases

See the course description at <https://programmeinfo.bi.no/en/course/GRA-4142/2021-autumn>.

Kaggle Machine Learning & Data Science Survey in 2020



Source: <https://www.kaggle.com/c/kaggle-survey-2020>

Asynchronous part (1/3)

- ▶ (Mandatory) Before each class you will be given a Python notebook and/or other material to go through.
- ▶ (Optional, but recommended) Additional exercises. You can work on these alone or in groups of 2–4 students.

Synchronous part (2/3)

Each Friday from today until November 12.

- ▶ Summarizing and/or discussing the topics covered by the notebooks and/or material.
- ▶ Solving practical assignments (learning by doing).

Remember to check the [itslearning](#) page regularly.

You cannot learn programming only by following lectures, reading or watching videos, **you need to practice** (a lot)!

- ▶ Remember the asynchronous learning activities.
- ▶ Discuss with your course- and group-mates.
- ▶ Use online resources, e.g.
 - ▶ Google
 - ▶ Stack Overflow and/or other online fora
 - ▶ W3Schools tutorials
 - ▶ Tutorials on YouTube
 - ▶ Additional exercises (some links will be provided later)

We are here for you to help and guide you! Don't hesitate to ask for help.

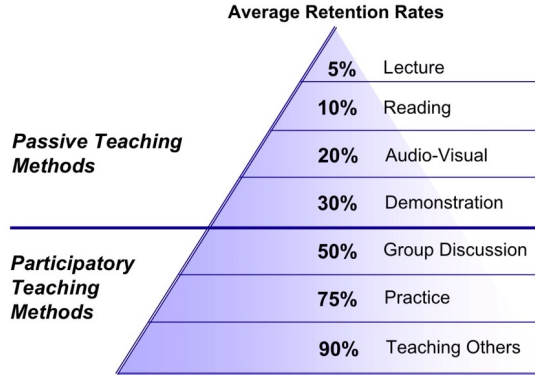
Individual take-home exam: you will be given a set of (practical) assignments that will be due 24 hours after.

You will be allowed to use books, notes and online resources.

The assignments will be divided into several parts (Python programming, Pandas, and data management) and you need to pass each section to pass the exam.

Although you have 24 hours to complete the exam, you will not expected to work round the clock!

Remember, the most important is that you learn something useful and relevant for you future!



*Adapted from National Training Laboratories. Bethel, Maine

Note: This model is criticized by some scientists, nevertheless you will learn more if you discuss and do/practice.

- ▶ John L. Viescas: *SQL Queries for Mere Mortals: A Hands-On Guide to Data Manipulation in SQL*, 2018
- ▶ Charles R. Severance: *Python for Everybody: Exploring Data in Python 3*, 2009

Alternatively:

There are plenty tutorials and free resources online.