Quantum 1: Data-driven decision making
- Python & SQL (17034)

5 ECTS

Professor Benjamin Maury
E-mail: benjamin.maury@hanken.fi

Teaching assistants: Gustav Finne, Marco Lambrecht, Thoi Mai, Elyas Saif, Joosua Virtanen, and Xiaogeng Xu

Course description

This course deals with Python and SQL in the context of data-driven decision making. Part 1 introduces frameworks for accessing web data and doing textual analysis with Python 3. The course introduces the basics of the Python syntax. After the basics, the focus is on data use and access. Data (pandas) and textual analyses are reviewed. Networked programs are reviewed including web scraping (BeautifulSoup). Interactive webpage development is introduced.

Part 2 deals with Structured Query Language (SQL) for accessing and modifying data held in databases. We will review commands for interfacing with databases. After the basics, predictions from tables in databases are dealt with. Web services are reviewed including API contracts. The integration of SQL and Python is discussed.

Students apply the frameworks in various data projects. Examples used are relevant for corporate and financial analysts.

Learning objectives

Upon completion of the course the students will

- have an understanding of the Python syntax,
- be able to perform basic data and textual analyses,
- have a framework for accessing web data,
- have skills to develop simple interactive webpages,
- have working knowledge of SQL commands,
- be able to access and modify data held in various relational databases, and
- be able to integrate SQL, Python, and dashboards (using, e.g., Tableau).

Course requirements (Scale 0-100 points)

(1) *Small projects (1-8)* 0-80 points. Projects 1 (Python basics) and 6 (SQL basics) are mandatory (unless you have documented basic knowledge of Python and SQL syntaxes approved by Elyas Saif). From the remaining 7 assignments you should

select at least 6 projects to submit. Upload through Moodle before deadline (see Moodle submission section).

1. Python basics [0-10 points] (including printing, comments, variables, numbers, math, strings, if/else, lists, loops, dictionaries, functions) [mandatory]
2. Data (pandas) and textual analysis [0-10 points] (data analysis with pandas)
3. Networked programs [0-10 points] (opening web pages to be treated like files, web scraping, BeautifulSoup)
4. Handling image and video data [0-10 points]
5. Interactive web pages [0-10 points]
6. SQL [0-10 points] (including select, update, delete, insert information, create database, alter database, create table, alter table, drop table, etc.) Accessing and modifying databases in practice [mandatory].
7. Web services [0-10 points] (APIs)
8. AI/ML methods with SQL using MindsDB [0-10 points]
9. Integrating SQL with Python, and dashboards (e.g., Tableau) [0-10 points]

(2) *Learning diary 0-20 points.* Individual work that discusses what you have learned and reviews the course content and your projects.

- Review the course content and your projects as well as discuss what you have learned.
- Discuss how you could automate repetitive tasks regarding (i) data access, and (ii) data & textual analysis for an equity fund that updates its stock portfolio regularly.
- How do you think business students can benefit from Python and SQL programming in general?

Length: 4-6 pages.

See the *Assignments* document on Moodle for more details.

Submit the diary through the course Moodle page (deadline: October 20).

Total requirements: 50 points to pass the course.

Topics and readings:

Part I:

Introduction to Part I (September 4, 8:30-10:00, online with Microsoft Teams**):**
Overview of course and requirements

- Installing Python 3 via Anaconda https://docs.anaconda.com/anaconda/install/windows/
- Using Jupyter notebook

- Elyas Saif demonstrates how to install Python

Session 1 (September 5, 8:30-10:00, online with Microsoft Teams)**: Introduction to the Python syntax**

- We will review printing, comments, variables, numbers, math, strings, if/else, lists, loops, dictionaries, functions, and writing scripts

- Elyas Saif demonstrates the basic syntax in Python

Literature:

- Jake VanderPlas "A Whirlwind Tour of Python"

https://jakevdp.github.io/WhirlwindTourOfPython/

Background video:

- Intro to Python for Business with Mattan Griffel (Columbia University)

https://www.youtube.com/watch?v=32LiJFZC484

Session 2 (September 11, 8:30-10:00, online with Microsoft Teams): **Data analysis with pandas**

- Using pandas and other packages for data analysis

- Joosua Virtanen shows how to (i) import data files to Jupyter (ii) run regressions with the data, and (iii) visualize data.

Literature:

- Dr. Charles R. Severance: "Python for Everybody: Exploring Data Using Python 3" (Chapter 11)

https://www.py4e.com/book.php

- Pandas:

https://pandas.pydata.org/pandas-docs/stable/getting_started/index.html

Session 3 (September 14, 8:30-10:00, online with Microsoft Teams): **Textual analysis**

- Joosua Virtanen shows how to us nltk and other packages for data analysis
- Literature:
    - https://www.py4e.com/html3/11-regex
    - https://web.stanford.edu/~jurafsky/slp3/

Session 4 (September 18, 8:30-10:30, online with Microsoft Teams): **Networked programs**

- Networked programs are reviewed (HTTP) including web scraping (BeautifulSoup)

- Accessing web data that can be treated like a file

- Joosua Virtanen and Thoi Mai demonstrate the use of BeautifulSoup

Literature:
- Dr. Charles R. Severance: ”Python for Everybody: Exploring Data Using Python 3” (Chapter 12)

https://www.py4e.com/book.php

Session 5 (**September 21, 10:15-12:30**, room 309 and online with Teams): **Introduction to analysis of image data & ChatGPT**

- Thoi Mai will show how to collect, read, edit, extract data, and save image data
- Video data will also be dealt with.

- ChatGPT for data processing such as code generating, code debugging, using ChatGPT to analyze textual data like measuring sentiment, extract information from text

Pre-assignment: Read the uploaded code and learn basically how to use Google Colab

Instructions about Google Colab:
- https://www.youtube.com/watch?v=inN8seMm7UI&ab_channel=TensorFlow
- https://towardsdatascience.com/getting-started-with-google-colab-f2fff97f594c
- https://www.youtube.com/watch?v=i-HnvsehuSw&ab_channel=ProgrammingKnowledge

Session 6 (September 25, 8.30-10:00, Hanken room 411 & Teams**)**: **Basics of developing web-based surveys**

- Introduction to oTree (version 5) (a software to design experiments based on Python)
- Structure of oTree (variables, functions, pages, templates)
- A simple example of a survey
- Marco Lambrecht will demonstrate the basics of developing web-based surveys

Session 7 (September 27, 8:30-10:00, Teams**)**: **Development of interactive webpages**

- Motivation: Why program webpages instead of using easy platforms like Qualtrics?
- Developing a simple interactive app
- Deploying an app on an external host

- Styling webpages

- Xiaogeng Xu will demonstrate the development of interactive webpages (JavaScript, HTML, and CSS)

Optional session (**September 29, 8:30-10:00**, Teams):
- The session is not mandatory.
- Xu and Marco will be available to provide help on the webpage assignment.

Part II:

Introduction to Part II (September 29, 12:30-14:00, online with Microsoft Teams**): Overview of course and requirements**
- Brief introduction to relational databases
- Installing PostgreSQL

- Elyas Saif demonstrates how to set up the free PostgreSQL (https://www.postgresql.org/).

- PostgreSQL is supported by easy-to-use database management tools such as pgAdmin (also free and open source).

Pre-assignment: Watch video on Moodle

Session 8 (October 2, 12:30-14:00, online with Microsoft Teams**): Introduction to the SQL syntax**

- Elyas Saif demonstrates the basic syntax in SQL
- The session covers data retrieval from a given database (select, select distinct, count, where, order by, between, like, group by, having, joins).

Pre-assignment: Watch video on Moodle

Literature:
"SQL: practical guide for developers" (by Michael J. Donahoo and Gregory D. Speegle), chapters: 1-5. The book can be accessed through Hanken's library: https://hanken.finna.fi/Record/hanna.514914.

Session 9 (October 6, 8:30-10:00, online with Microsoft Teams): **Accessing and modifying databases in practice**

- Elyas Saif shows how to work with SQL
- This session deals with creating and modifying a database (create table, insert, delete, update, alter table, drop table).

Pre-assignment: Watch video on Moodle

Literature:

"SQL: practical guide for developers" (by Michael J. Donahoo and Gregory D. Speegle), chapters: 8-9. The book can be accessed through Hanken's library: https://hanken.finna.fi/Record/hanna.514914.

Session 10 (October 9, 8:30-10:00, online with Microsoft Teams**)**: **Web services**

- Accessing web data via APIs
- We will use Python to access databases in Quantum throughs APIs

- Gustav Finne will demonstrate the use of APIs

Pre-assignment: Create a personal account on WRDS. Your account will be approved by Vilhelm Lönnberg or colleagues during working hours.

Literature:

- Dr. Charles R. Severance: "Python for Everybody: Exploring Data Using Python 3" (Chapter 13)

https://www.py4e.com/book.php

Session 11 (October 13, 16:00-17:45, online with Microsoft Teams**): Databases and ML/AI** (guest lecture, 1.5 bonus points for participation).

- Cosmo Jenytin discusses how to make predictions from tables in databases using standard SQL in MindsDB (https://mindsdb.com/)

- Questions concerning this session and the assignment can be directed to cosmo.jenytin@iki.fi

Session 12 (October 17, 8:30-10:00, online with Microsoft Teams**)**: **SQL, Python, and dashboards**

- Elyas Saif demonstrates how to integrate SQL and Python
- Dashboards are developed using Tableau that make use of SQL and Python.

Updated: August 30, 2023.
Changes are possible.