

Overview

Block	Session	Unit
Setting up Python for analytics	Life is easy when using the right tools: ☛ Why and how to work with Python?	01 – Getting started with Python 02 – Understanding and using code notebooks 03 – Filling your notebook with code and text 04 – Advanced features of code notebooks 05 – Working with a local code editor 06 – Installing and using packages 07 – Getting help
	In God we trust, all others must bring data: ☛ How to load data in Python?	08 – Reading in data 09 – Basic investigation of data
Visualizing data with Python	A picture is worth a thousand words: ☛ How to create a plot (matplotlib)?	10 – Generate basic plots 11 – Format basic plots 12 – Add further features to basic plots
	There is always room for improvement: ☛ How to create nicer plots (seaborn)?	13 – Creating advanced plots 14 – Colors, palettes, themes, and styles 15 – Very advanced plotting options
Wrangling data with Python	Select the needle in a haystack: ☛ How to use “select” operations?	16 – Data wrangling techniques 17 – Selecting rows 18 – Selecting columns 19 – Updating rows and columns
	Get a high-level overview: ☛ How to use “aggregate” operations?	20 – Basic aggregating techniques 21 – Advanced aggregating techniques 22 – Combined select-aggregate operations
	Leverage synergies: ☛ How to use “merge” operations?	23 – Basic merging techniques 24 – Inner joins and full outer joins 25 – Left and right outer joins 26 – Advanced merging techniques
	Play with the Pros: ☛ How to use SQL for data wrangling?	27 – Connect to a database 28 – Databases and the role of SQL 29 – Select operations in SQL 30 – Aggregate operations in SQL 31 – Merge operations in SQL
Automating workflows with Python	Don't lose control: ☛ How to use conditions and loops?	32 – If-else statements 33 – Basic looping techniques 34 – Advanced looping techniques
	First be effective and then be efficient: ☛ How to use functions?	35 – Understanding why to use functions 36 – How to write a function 37 – Tips and tricks for creating functions
	No data, no problem: ☛ How to simulate data?	38 – Creating sequences automatically 39 – Simulating and manipulating strings 40 – Simulating variables by using distributions
Practicing Python	Putting everything together: ☛ How to do a full analytics project?	41 – An introduction to scoring models 42 – Create your own scoring model

Course setup

- *Course description:*
 - People that use data analytics often spend more than 80% of their time with collecting, cleaning, and organizing data and only 20% with applying statistical models. This is not only true for real world business analytics, but also for data analyses within bachelor/master theses. This class will prepare you for those challenges by applying a non-technical approach. Meaning we start every unit with "real-world" questions and explain key programming concepts "on the way" while discussing the solutions to these questions.
 - This class is taught in workshop format, with integrated exercises. **For every session, you are required to bring your laptop (with a recent version of your operating system installed) and a pair of headphones.** We do not require any experience with Python as we start from the very beginning. However, we do require the willingness to actively participate and contribute to the class. No statistical models (besides mean and standard deviation) will be discussed in this class.
- *Learning goals:*
 - Getting to know data management and munging techniques that "scale well", i.e., that are applicable to datasets which are usually observed in real-world settings (1 to 100 GB).
 - These include loading external data (from text files, Excel files, databases), manipulating data (merging, aggregating, and selecting observations), as well as simplifying complex and repetitive data munging tasks (writing loops and conditional statements).
- *Question on lectures and exercises:*
 - If you have a question, proceed as follows:
 - First, try to find an answer by using Google / StackOverflow.
 - If this is still not successful, ask the members of your study group.
 - As a last option, get in touch with the lecturer team.
 - For questions related to slides, please be able to provide the lecture and slide number. For questions related to exercises, please be able to provide the exercise number.
- *Course registration:*
 - **As the module booking ends after the completion of the course, we ask all students to confirm their participation in the class. If you don't register in the UZH Modulbuchungstool but have officially confirmed your registration through the registration confirmation on OLAT, the class will be added to your class record by the dean's office.**
 - **Course registration is possible on Day 1, after the mock exercise, via OLAT (Course registration).**
 - **Examination material can only be accessed after this confirmation and after confirming you agree with the Honor Code for Students Completing Official Assessments at the University of Zurich (Honor Code).**
- *The daily coding exercise:*
 - Only programming questions will be asked in the daily coding exercise.
 - All exams can be started no later than 5 minutes after the communicated starting time. After 5 minutes, the exams are no longer accessible. Once an exercise is started, it must be finished within the given time.
 - **The daily exercises are in multiple choice format, each question having a single correct answer.** To determine which option is correct, you will be required to solve a given problem using Python code.
 - The exercises are open book. The code files which you create during the study sessions, can be used during the daily exam, and might prove useful to answer the questions.
 - In the exercise there will be no questions from the materials marked with "HOMESTUDY". However, such questions can come up in the final exam.
 - **In days 1-4, each daily exercise contains two questions. You will have in total 15 minutes to answer the questions. Day 5 is slightly different: you will have 4 questions and 30 minutes time to answer.**
 - **Please be aware that all exercises can be subject to randomization (e.g., random order of questions and answers) as well as of personalization (e.g., individual input datasets for each student). Further, the sequence of the questions will be fixed, i.e., there will be no option to go back to a question once this question has been answered.**

- Please be aware that for each question you will have to “submit” your answer and go to “next question”. Once you have finished the exercise you have to click on “finish test” and “close test”.
- **Logging out from OLAT before finishing the exercise (e.g., caused by a laptop restart during the exercise) will cause that you will not be able to submit answers for the respective exercise. It will be your responsibility to ensure this does not happen.**
- *The final exam:*
 - The final exam will test both, knowledge questions and programming questions (e.g., what is the correct command for a certain task).
 - **All content (incl. home study sessions) will be covered in the final exam.**
 - **The final exam has 20 multiple-choice questions and covers all lectures. You have 35 minutes of time for completing the exam.**
 - **Questions may have multiple correct answers or only one correct answer. There always is at least one correct answer.**
 - The wording of the question, e.g., the usage of the plural “are”, does not necessarily imply that multiple answer options are correct. This wording has been chosen for readability reasons.
 - Preparing for the final exam will take some time, please start as early as possible.
 - **Please be aware that all exams can be subject to randomization (e.g., random order of questions and answers) as well as of personalization (e.g., individual hyperparameter or input datasets for each student).**
- Finally, you will have to complete several **DataCamp exercises**.
 - During class, you will receive an invitation to join a group on www.datacamp.com. Create an account and please use your real name (e.g., not MickeyMouse1995).
 - Work on the assigned online exercises to strengthen your programming skills at your own pace
 - Please check the deadline for the online exercises on www.datacamp.com.
 - The grading for this task depends on your performance in solving these online exercises.
- Contact details:
 - Email/telephone:
Email: radu.tanase@business.uzh.ch
Telephone: +41 44 634 2918
 - Address:
Department of Business Administration
URPP Social Networks
Andreasstrasse 15
8050 Zurich
Switzerland

Course schedule

DAY 1

<i>Time</i>	<i>Subject</i>	<i>Medium</i>
09:00 – 09:30	Course logistics & outlook day 1	Classroom
09:30 – 10:30	Study session 1	Classroom
10:30 – 10:45	Break	
10:45 – 11:45	Study session 2	Classroom
11:45 – 12:00	Break	
12:00 – 13:00	Study session 3	Classroom
13:00 – 13:15	Mock 1: daily exercise	OLAT
13:15 – 14:00	Lunch break	
14:00	Deadline to officially register for the course	OLAT
14:00 – 15:00	Individual preparation for daily coding exercise	Classroom
15:00 – 15:15	Daily coding exercise	OLAT (on-site supervision)
after 15:15	Home study session	-

Session 1: Why and how to work with Python?

- 01 – Getting started with Python
- 02 – Understanding and using code notebooks
- 03 – Filling your notebook with code and text (HOMESTUDY)
- 04 – Advanced features of code notebooks (HOMESTUDY)
- 05 – Working with a local code editor (HOMESTUDY)
- 06 – Installing and using packages
- 07 – Getting help (HOMESTUDY)

Session 2: How to load data in Python?

- 08 – Reading in data
- 09 – Basic investigation of data

Session 3: How to create a plot (“matplotlib”)?

- 10 – Creating basic plots
- 11 – Formatting basic plots
- 12 – Add further features to basic plots (HOMESTUDY)

DAY 2

<i>Time</i>	<i>Subject</i>	<i>Medium</i>
09:00 – 09:15	Recap day 1 & outlook day 2	Classroom
09:15 – 10:15	Study session 4	Classroom
10:15 – 10:30	Break	
10:30 – 11:30	Study session 5	Classroom
11:30 – 12:00	Overview essential concepts	Classroom
12:00 – 13:00	Lunch break	
13:00 – 14:00	Study session 6	Classroom
14:00 – 14:15	Break	
14:15 – 15:00	Individual preparation for daily coding exercise	Classroom
15:00 – 15:15	Daily coding exercise	OLAT (on-site supervision)
after 15:15	Home study session	-

Session 4: How to create nicer plots (“seaborn”)?

- 13 – Creating advanced plots
- 14 – Colors, palettes, themes, and styles
- 15 – *Very advanced plotting options (HOMESTUDY)*

Session 5: How to use “select” operations?

- 16 – Data wrangling techniques
- 17 – Selecting rows
- 18 – Selecting columns
- 19 – *Updating rows and columns (HOMESTUDY)*

Session 6: How to use “aggregate” operations?

- 20 – Basic aggregating techniques
- 21 – Advanced aggregating techniques
- 22 – *Combined select-aggregate operations (HOMESTUDY)*

DAY 3

<i>Time</i>	<i>Subject</i>	<i>Medium</i>
09:00 – 09:15	Recap day 2 & outlook day 3	Classroom
09:15 – 10:15	Study session 7	Classroom
10:15 – 10:30	Break	
10:30 – 11:30	Study session 8	Classroom
11:30 – 12:00	Overview essential concepts	Classroom
12:00 – 13:00	Lunch break	
13:00 – 14:00	Study session 9	Classroom
14:00 – 14:15	Break	
14:15 – 15:00	Individual preparation for daily coding exercise	Classroom
15:00 – 15:15	Daily coding exercise	OLAT (on-site supervision)
after 15:15	Home study session	-

Session 7: How to use “merge” operations?

- 23 – Basic merging techniques
- 24 – Inner joins and full outer joins
- 25 – Left and right outer joins
- 26 – *Advanced merging techniques (HOMESTUDY)*

Session 8: How to use SQL for data wrangling? (1/2)

- 27 – Connect to a database
- 28 – Databases and the role of SQL

Session 9: How to use SQL for data wrangling? (2/2)

- 29 – Select operations in SQL
- 30 – Aggregate operations in SQL
- 31 – *Merge operations in SQL (HOMESTUDY)*

DAY 4

<i>Time</i>	<i>Subject</i>	<i>Medium</i>
09:00 – 09:15	Recap day 3 & outlook day 4	Classroom
09:15 – 10:15	Study session 10	Classroom
10:15 – 10:30	Break	
10:30 – 11:30	Study session 11	Classroom
11:30 – 12:00	Overview essential concepts	Classroom
12:00 – 13:00	Lunch break	
13:00 – 14:00	Study session 12	Classroom
14:00 – 14:15	Break	
14:15 – 15:00	Individual preparation for daily coding exercise	Classroom
15:00 – 15:15	Daily coding exercise	OLAT (on-site supervision)
after 15:15	Home study session	-

Session 10: How to use conditions and loops?

- 32 – If-else statements
- 33 – Basic looping techniques
- 34 – *Advanced looping techniques (HOMESTUDY)*

Session 11: How to use functions?

- 35 – Understanding why to use functions
- 36 – How to write a function
- 37 – *Tips and tricks for creating functions (HOMESTUDY)*

Session 12: How to simulate data?

- 38 – Creating sequences automatically
- 39 – *Simulating and manipulating strings (HOMESTUDY)*
- 40 – Simulating variables by using distributions

DAY 5

<i>Time</i>	<i>Subject</i>	<i>Medium</i>
09:00 – 09:30	Recap day 4 & outlook day 5	Classroom
09:30 – 09:45	Programming project - Introduction	Classroom
09:45 – 10:00	Break	
10:00 – 12:00	Programming project	Classroom
12:00 – 13:00	Lunch Break	
13:00 – 15:00	Programming project + Individual exam preparation	Classroom
15:00 – 15:30	Final exam	Classroom (on-site supervision)
15:30 – 16:00	Break	
16:00 – 16:30	Daily coding exercise	OLAT (on-site supervision)

Programming project

- 41 – An introduction to scoring models
- 42 – Create your own scoring model