Social Data Science Base Camp (15 ECTS)

Course Description

This course introduces students to the interdisciplinary degree programme of Social Data Science. In the first week, students are introduced to the group-based learning and working practices, which are core elements of the degree program. For the rest of the term, students are introduced to the fundamentals of programming, data collection, and data analysis in Python including regression analysis. This will be combined with lectures and exercises that focus on elementary statistical modelling techniques and integrated quali-quant methods. Overall, the course will teach students the basic skills to program, collect and process data from a variety of online sources and structure them into a dataset, and to conduct basic analyses on that dataset.

More details here: https://kurser.ku.dk/course/asdk20001u/2021-2022

Course Schedule

Mondays and Wednesdays (+ 2 Fridays)

Lectures: 8:00 – 10:00
Exercises: 10:00 – 13:00

Teachers

Samantha Breslin (SB) (course responsible) – samantha.breslin@sodas.ku.dk
Friedolin Merhout (FM) – fmerhout@sodas.ku.dk
Gregory Eady (GE) – gregory.eady@sodas.ku.dk
Asger Andersen (AA) – asa@sodas.ku.dk

Teaching Assistants

Veronica Correa Bayas (Class I) – bjn802@alumni.ku.dk
María José Romero Lado (Class II) – wfs758@alumni.ku.dk
Mònica Aguilà i Sans (Class III) – ptl441@samf.ku.dk

Course Literature

Brooker, Philip D. 2020. *Programming with Python for Social Scientists*. London: Sage.

The book is available for purchase in Academic Books at CSS. All other literature will be uploaded as PDFs on Absalon, under the module page. Recommended literature is not mandatory.

Assignments/Exercises

To be eligible for the exam in Social Data Science Base Camp, it is a requirement that students have completed and submitted all of the exercise assignments via Absalon prior to the exam start date. Each class-day will have an associated exercise assignment (max. 28 Jupyter Notebooks).

See: https://kurser.ku.dk/course/asdk20001u/2021-2022

The assignments are the exercises you will work through during the exercise portion of class. They are due at noon on Friday of the same week. TAs will then review the solutions and any common errors or misunderstandings on the following Monday.

You may submit the exercises after the initial deadline with an absolute **final deadline of Monday, December 20, 2021**.

Course Plan

	Title	Content	Readings	Teachers	Room
Week 36 – Beginnings					
Mon, Sept. 6	Base Camp Introduction	• Intro to Base Camp • What is coding • Programming community • Landscape of tools and applications • Elementary data types + strings Mathematical Operations	Required: Chapters 0, 1, 2 & 3 in Brooker (Note: read but do not do the setup instructions in ch. 3) Chapter 4 (pgs 57-65) in Brooker	SB	1.1.18
Mon, Sept. 6	Exercises				1.1.12 1.0.10 2.1.02
Wed, Sept. 8	Pandas Intro & Data Ethics	• A brief history of programming • Data Ethics • Why Pandas First • Getting Data I (downloading) • CSV	Required: Chapter 6 (Ethics) in Salganick, Matthew. 2017. *Bit by Bit: Social Research in the Digital Age.* Princeton University Press. https://www.bitbybitbook.com/en/1st-ed/ethics/ Recommended: Friedman, Linda Weiser. 1992. "From Babbage to Babel and Beyond: A Brief History of Programming Languages." Computer Languages 17 (1): 1– 17.	SB	1.1.18
Wed, Sept. 8	Exercises				1.1.12 1.0.10 2.1.02
Week 37 - Programming Structures					
Mon, Sept. 13	Flow control and more	• If statements • Calling methods and functions • Debugging • Commenting • Pseudocode • Categorical & binary variables	Required: Chapter 4 in Brooker (pg 70 – 78) Recommended: Grus. 2019. Ch 2: "A Crash Course in Python" in Data Science from Scratch.	FM	1.1.18
Mon, Sept. 13	Exercises				1.1.12 1.0.10 2.1.02

Wed, Sept. 15	Naming and Organizing Data in Python	• Politics of names • Lists • Tuples • Dictionaries • More Strings	Required: Chapter 5 in Brooker Recommended: Eglash, Ron. 2007. "Broken Metaphor: The Master-Slave Analogy in Technical Literature." Technology and Culture 48 (2): 360–69. D'Ignazio, Catherine and Laura Kline. 2020. "What Gets Counted Counts." In: *Data Feminism*. https://data-feminism.mitpress.mit.edu/pub/c zq9dfs5	SB	1.1.18
Wed, Sept. 15	Exercises				1.1.12 1.0.10 2.1.02
Week 38 – More Structures					
Wed, Sept. 20	Digital Data Structures	• HTML • CSS • Tabular & nontabular data • JSON	Required: Chapter 12 in Brooker	FM	1.1.18
Wed, Sept. 20	Exercises				1.1.12 1.0.10 2.1.02
Wed, Sept. 22	Loops & more	• For loops • While loops • List comprehension • Tracing	Required: Chapter 6 (Loops and list Comprehension, pgs 118-124) in Brooker	SB	1.1.18
Wed, Sept. 22	Exercises				1.1.12 1.0.10 2.1.02
Week 39 – Review and Pandas					
Mon, Sept. 27	Review & Structures	• Review • Reading code • Building bigger programs • Shell	Required: Chapter 8 in Brooker	SB	1.1.18
Mon, Sept. 27	Exercises				1.1.12 1.0.10 2.1.02
Mon, Sept. 29	Data with Pandas	• Creating pandas data frames (from csv, from data structures, etc.) • Accessing and modifying data • Merging	Required: Chapter 14 in Brooker (pg 255 – 269) Recommended https://realpython.com/pandasdataframe/ (up to and including Accessing and Modifying Data)	SB	1.1.18

		DataFrames.			
Mon, Sept. 29	Exercises				1.1.12 1.0.10 2.1.02
Week 40 - Getting Data					
Mon, Oct. 4	Data from APIs	• Different APIs • Twitter API • Getting data II (querying)	Required: Chapter 11 in Brooker	FM	1.1.18
Mon, Oct. 4	Exercises				1.1.12 1.0.10 2.1.02
Wed, Oct. 6	Text & Data Cleaning	• Working with text files • RegExp • Cleaning DF (finding, replacing values)	Required: Chapter 10 in Brooker	FM	1.1.18
Wed, Oct. 6	Exercises				1.1.12 1.0.10 2.1.02
Fri, Oct. 8	Netnography	• Why Netnograpy • Conducting Observations • Fieldnotes	Required: Kozinets, Robert V. 2918. *Netnography*. Sage. (Chapter 5)	SB	2.0.63
Fri, Oct. 8	Exercises				18.01.11
Week 41 – Getting Data II					
Mon, Oct. 11	Scraping I	• Basic scraping techniques • Beautiful Soup	Required: Chapter 13 in Brooker	FM	1.1.18
Mon, Oct. 11					1.1.12 1.0.10 2.1.02
Wed, Oct. 13	Getting Data Practice	• More Twitter API • Build the SODAS Data Frame	Required: Cursor Tutorial, Tweepy: https://docs.tweepy.org/en/stable/cursor_tutorial.html	SB	1.1.18
Wed, Oct. 13	Exercises				1.1.12 1.0.10 2.1.02
Week 42 & 43 - NO CLASS					
Week 44 – Visualization, Functions, & Review					

Mon, Nov. 1	Basic Visualization	• Politics & practices of visualization • Visualization with Matplotlib • Plotting with Pandas • Group by & Summary Stats	Required: Chapter 14 (from pg. 270) in Brooker Recommended: Grus. 2019. Ch 3: "Visualizing Data" In Data Science from Scratch D'Ignazio, Catherine and Laura Kline. 2020. "The Numbers Don't Speak for Themselves." In: *Data Feminism*. https://datafeminism.mitpress.mit.edu/pub/c zq9dfs5	FM	1.1.18
Mon, Nov. 1	Exercises				1.1.12 1.0.10 2.1.02
Wed, Nov. 3	Functions	• Writing Functions	Chapter 6 (pgs 110 – 117) in Brooker	SB	1.1.18
Wed, Nov. 3	Exercises				1.1.12 1.0.10 2.1.02
Fri, Nov. 5	Review	• Review	Required: Chapter 15 in Brooker	FM	2.0.63
Fri, Nov. 5	Exercises				2.1.36 2.1.55 2.2.55
Week 45 - Basic Data Analysis					
Mon, Nov. 8	Counting Text	• Word Counting • Content Analysis	Required: http://somatosphere.net/forumpost/covid19-danish-twittercomputational-map/	SB & HBC	1.1.18
Mon, Nov. 8	Exercises				1.1.12 1.0.10 2.1.02
Schedule for Part II TBA					