Syllabus

Introduction to Python and SQL
1 credit/2 ECTS

Brian Kim, PhD

Video lecture by Brian Kim, PhD

June 06, 2021 – June 29, 2021

Short Course Description

This course introduces students to the basics of Python and SQL for data analysis. Students will explore real publicly-available datasets, using the data analysis tools in Python to create summaries and generate visualizations. Students will learn the basics of database management and organization, as well as learn how to code in SQL and work with SQL databases. By the end of the class, students should understand how to read in data from CSV files or from the internet and be comfortable using either SQL or Python to aggregate, summarize, describe, and visualize these datasets.

Course Objectives

By the end of the course, students will…
- Understand the basic structure of how Python and object-oriented programming works
- Be able to write basic Python code, including functions and loops
- Know how to use Pandas and matplotlib packages in Python to analyze data and create visualizations
- Be comfortable reading error messages and Python documentation to diagnose and debug code
- Understand how relational databases work
- Be able to construct a query to answer questions about the data
- Understand how joins work and how to use them.

Prerequisites

No prerequisites.

Class Structure and Course Concept

This is an online course, using a flipped classroom design. It covers the same material and content as an on-site course but runs differently. In this course, you are responsible for watching video-recorded lectures and reading the required literature for each unit prior to participating in mandatory weekly one-hour online meetings where students have the chance to discuss the materials from a unit with the instructor.

Although this is an online course where students have more freedom in when they engage with the course materials, students are expected to spend the same amount of time overall on all activities in the course – including preparatory activities (readings, studying), in-class-activities (watching videos, participating in online meetings), and follow-up activities (working on assignments and exams) – as in an on-site course. As a rule of thumb, you can expect to spend approximately 3h/week on in-class-activities and 9 hours per week on out-of-class activities (preparing for class, readings, assignments, projects, studying for quizzes and exams). Therefore, the workload in all courses will be approximately 12h/week. This is a 1-credit/2-ECTS course that runs for 4 weeks. Please note that the actual workload will depend on your personal knowledge.

Mandatory Weekly Online Meetings

Tuesdays 1-2 PM EDT / 7-8 PM CEST, starting June 8, 2021

Meetings will be held online through Zoom. Follow the link to the meeting sessions on the course website on mannheim.instructure.com. If video participation via Internet is not possible, arrangements can be made for students to dial in and join the meetings via telephone.

In preparation for the weekly online meetings, students are expected to watch the lecture videos and read the assigned literature before the start of the meeting. In addition, students **must post questions or comments** about the materials covered in the videos and readings of the week in the forum before the meetings (deadline for posting questions is Tuesday, 9:00 AM EDT/3:00 PM CEST).

Students have the opportunity to use a different Zoom meeting room to connect with peers outside the scheduled weekly online meetings (e.g., for study groups). Detailed information is posted on the course page in Canvas. Students are encouraged to post the times that they will be using the room to the course website forum to avoid scheduling conflicts. Students are not required to use Zoom and can use other online meeting platforms, such as Microsoft Teams, Google Hangout or Skype.

Grading

Grading will be based on:
- 4 online quizzes (5% each)
- Participation in discussion during the weekly online meetings and submission of questions to the weekly discussion forums (deadline: Tuesday 9:00 AM EDT/3:00

PM CEST before class) demonstrating understanding of required readings and video lectures (20% of grade)
- 4 homework assignments (15% each)

A+   100 - 97
A   96 - 93
A-   92 - 90
B+   89 - 87
B   86 - 83
B-   82 - 80
Etc.

The grading scale is a base scale recommended by the MDM. Variations for grading on a scale are at the discretion of the instructor.

Dates of when assignments will be due are indicated in the syllabus. Extensions will be granted sparingly and are at the instructors' discretion.

Technical Equipment Needs

The learning experience in this course will mainly rely on the online interaction between the students and the instructors during the weekly online meetings. Therefore, we encourage all students in this course to use a web camera and a headset. Decent quality headsets and web cams are available for less than $20 each. We ask students to refrain from using built-in web cams and speakers on their desktops or laptops. We know from our experience in previous online courses that this will reduce the quality of video and audio transmission and therefore will decrease the overall learning experience for all students in the course. In addition, we suggest that students use a wire connection (LAN), if available, when connecting to the online meetings. Wireless connections (WLAN) are usually less stable and might be dropped.

Long Course Description

Python has recently seen a huge surge not only as a programming language, but also as a tool for data analysis. In this course, we will introduce the basics of programming in Python for the purposes for data analysis. We will explore the Longitudinal Employer-Household Dynamics (LEHD) datasets, specifically the LEHD Origin-Destination Employment Statistics (LODES) datasets, using Python to read in datasets, explore the datasets, find statistical summaries, and create visualizations. By the end of the course, students should be comfortable with using Python for data analysis, as well as be capable of using their general knowledge of the Python language for other applications.

In addition, as more and more data becomes available, relational database management systems (RDBMS) have become increasingly popular because it allows people to relatively easily organize large amounts of data. In many cases, knowledge of SQL is crucial to being able to access this data. In this course, we will introduce the basics of programming in SQL

using PostgreSQL. We will explore the Longitudinal Employer-Household Dynamics (LEHD) datasets, specifically the LEHD Origin-Destination Employment Statistics (LODES) datasets, using SQL to explore the datasets and find statistical summaries. By the end of the course, students should be comfortable with constructing basic queries of the database and linking multiple tables together using SQL.

Readings

Primary Readings
There is no textbook for this course.

Required and Recommended Readings
List of required and recommended readings for each class are provided below for each specific unit.

Academic Conduct

Clear definitions of the forms of academic misconduct, including cheating and plagiarism, as well as information about disciplinary sanctions for academic misconduct may be found at

https://www.president.umd.edu/sites/president.umd.edu/files/documents/policies/III-100A.pdf (University of Maryland) and

https://www.uni-mannheim.de/en/research/good-research-practice/ (University of Mannheim).

Knowledge of these rules is the responsibility of the student and ignorance of them does not excuse misconduct. The student is expected to be familiar with these guidelines before submitting any written work or taking any exams in this course. Lack of familiarity with these rules in no way constitutes an excuse for acts of misconduct. Charges of plagiarism and other forms of academic misconduct will be dealt with very seriously and may result in oral or written reprimands, a lower or failing grade on the assignment, a lower or failing grade for the course, suspension, and/or, in some cases, expulsion from the university.

Accommodations for Students with Disabilities

In order to receive services, students at the University of Maryland must contact the Accessibility & Disability Service (ADS) office to register in person for services. Please call the office to set up an appointment to register with an ADS counselor. Contact the ADS office at 301.314.7682; https://www.counseling.umd.edu/ads/.

Students at the University of Mannheim should contact the Commissioner and Counsellor for Disabled Students and Students with Chronic Illnesses at http://www.uni-mannheim.de/studienbueros/english/counselling/disabled_persons_and_persons_with_chronic_illnesses/

Course Evaluation

In an effort to improve the learning experience for students in our online courses, students will be invited to participate in an online course evaluation at the end of the course (in addition to the standard university evaluation survey). Participation is entirely voluntary and highly appreciated.

Sessions

Week 1: Introduction to Python and Pandas

Video lecture and workbook: available Tuesday, June 1, 2021

Online meeting: Tuesday, June 8, 2021, 1-2 PM EDT / 7-8 PM CEST

Online quiz 1: due Thursday, June 10, 2021, 1-2 PM EDT / 7-8 PM CEST

Homework 1: due Thursday, June 10, 2021, 1-2 PM EDT / 7-8 PM CEST

Required Readings:
LEHD Origin-Destination Employment Statistics (LODES) OnTheMap: Data Overview (LODES Version 7): https://lehd.ces.census.gov/doc/help/onthemap/OnTheMapDataOverview.pdf

LEHD Origin-Destination Employment Statistics (LODES) Dataset Structure: https://lehd.ces.census.gov/data/lodes/LODES7/LODESTechDoc7.3.pdf

Week 2: Functions, Loops, and Visualizations

Video lecture and workbook: available Tuesday, June 8, 2021

Online meeting: Tuesday, June 15, 2021, 1-2 PM EDT / 7-8 PM CEST

Online quiz 2: due Thursday, June 17, 2021, 1-2 PM EDT / 7-8 PM CEST

Homework 2: due Thursday, June 17, 2021, 1-2 PM EDT / 7-8 PM CEST

Week 3: Introduction to SQL

Video lecture and workbook: available Tuesday, June 15, 2021

Online meeting: Tuesday, June 22, 2021, 1-2 PM EDT / 7-8 PM CEST

Online quiz 3: due Thursday, June 24, 2021, 1-2 PM EDT / 7-8 PM CEST

Homework 3: due Thursday, June 24, 2021, 1-2 PM EDT / 7-8 PM CEST

Week 4: Joins

Video lecture and workbook: available Tuesday, June 22, 2021

Online meeting: Tuesday, June 29, 2021, 1-2 PM EDT / 7-8 PM CEST

Online quiz 4: due Thursday, July 1, 2021, 1-2 PM EDT / 7-8 PM CEST

Homework 4: due Thursday, July 1, 2021, 1-2 PM EDT / 7-8 PM CEST