Introduction to SQL and Python

October 02 – October 13, 2023

Course Objectives

This course introduces participants to the basics of Python and SQL for data analysis. Participants will explore real publicly-available datasets, using the data analysis tools in Python to create summaries and generate visualizations. Participants will learn the basics of database management and organization, as well as learn how to code in SQL and work with SQLite databases. By the end of the class, participants 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.

Learning Outcomes:

After this course, participants should be able to:

- 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

Who Should Attend

Statisticians and statistical analysts in government, universities, business, and nonprofit organizations interested in learning how to use SQL and Python to aggregate, summarize, and describe data. The focus of this short course will be on gaining familiarity with these programming languages and empowering participants to learn more about how to use these languages in their own work. The instructor will provide guided video tutorials, and cloud-based tools will allow participants to practice their coding without installing anything on their own computers.

No prior programming experience is necessary for this course. Some basic knowledge of Statistics is expected.
Instructor

Brian Kim is an Assistant Research Professor at the Joint Program in Survey Methodology and Co-Director of the Social Data Science major at the University of Maryland. He received his Ph.D. in Statistics from the University of California, Los Angeles in 2018. He has taught various data science courses at the University of Maryland, including Data Science for Social Science, Machine Learning for the Social Sciences, Fundamentals of Computing and Data Display, and Introduction to Python and SQL. In addition, he has taught workshops and short courses on data science at NSF, USDA, and state level governments as part of the Applied Data Analytics program with the Coleridge Initiative.

Class Structure

The course will be in an online format from October 02 to October 13, 2023. Participants will have online access to the lecture videos walking through the course notebooks. Participants can watch the videos and work through the notebooks at their own pace. Live group online discussions are scheduled for Friday 10/06/23 and 10/13/23, from 10:00AM to 11:30AM EST. These sessions will be used to discuss questions that came up over the week when watching the videos. Students will have the opportunity to submit their questions in advance. The course notebooks will contain exercises to provide the participants with hands on experience. While these exercises are not mandatory and will not be collected, participants are encouraged to work through the assignments prior to the online meetings.