

Data Science Study Group
Virginia Tech National Capital Region
Kenneth H. Wong, Ph.D.

Meeting Time: Wednesdays, 11 am – 12 noon. Location varies to support different sites in the National Capital Region. Online option available through Virginia Tech WebEx, meeting number 649 837 338.

Canvas Site (used for discussion board, announcements, project/team support):

<https://canvas.vt.edu/enroll/Y8AYBK>

This site requires you to log in with VT credentials

Description:

The goal of the study group is to gather like-minded learners, ask questions, share project ideas and results, help each other, and encourage everyone to build up their skills with Data Science. While the convenience of asynchronous online learning is very useful, it's also great to interact with actual people!

The target audience for the study group are people who desire to learn more about data science and how it can be applied, but who are not computer scientists. Computer scientists are certainly welcome to participate, but may find the content non-challenging.

The group is open to all members of the Virginia Tech community, including students, faculty, and staff.

Topics:

What is Data Science? What value can it bring to a problem space?

The Data Science Process (Acquire, Explore/Clean, Analyze, Report, Action)

Understanding and manipulating data structures (How is your data organized?)

Data Science Tools: Python, Jupyter, Tableau

Visualization Libraries: Matplotlib, Bokeh, Seaborn

Machine Learning and Bots

Prerequisites:

Although there are no formal prerequisites, you should know how to use the command line interface on whatever computer system you use. Also, you should be comfortable with installing software packages and using online learning resources.

Grading: None

Materials:

The core "textbook" for this study group will be the online EdX course from UCSDx in "Python for Data Science", DSE200x, which can be found at:

<https://courses.edx.org/courses/course-v1:UCSanDiegoX+DSE200x+1T2018/course/>

This is a free-to-enroll course, although you can also sign up for a paid version where you can get a verified certificate and access to other perks.

The general format is a series of short videos, written materials, quizzes, forum discussions, and sample problems. The course assumes that you have some experience with programming and know how to use the command line interface. There are project components. This course also serves as a gateway to three additional UCSDx courses that cover data science in more depth.

Professional Ethics and Honor System:

Although this is not an academic course, participants are still expected to adhere to the principles laid out in the Graduate Honor System, which can be found at

<https://graduateschool.vt.edu/academics/expectations/graduate-honor-system.html>