

COURSE SYLLABUS

1. General Information

Instructor: Amr Farahat (Amr)
<afarahat@andrew.cmu.edu>; TEP 4124
I hold virtual office hours during Mini 2 as follows (no appointment necessary):
Thursdays 6:00 p.m. – 7:00 p.m.[§]
Saturdays 4:00 p.m. – 5:00 p.m. ^{**}
Zoom link: <https://cmu.zoom.us/my/afarahat>
If necessary, you can email me to schedule an appointment outside of these hours.

Teaching assistant: Serim Hwang (Serim)
<serimh@andrew.cmu.edu>
Office hours: T.B.D.

Discussion sessions: Section M: Wednesdays 8:15 p.m. – 9:30 p.m. online
Section O: Wednesdays 6:45 p.m. – 8:00 p.m. online
See page 3 for schedule details.

Important dates: October 26th (first class); November 23rd (no class - Thanksgiving); December 14th (last class); December 18th (final exam due)

Course homepage: canvas.cmu.edu (includes links to all course content)

2. Course Description

R and Python are indispensable tools in the modern practice of data science. They are powerful yet *easy-to-learn and easy-to-use* programming languages. They enable you to import, combine, tidy, transform, explore, and model all sorts of data, big and small. They enable you to document and communicate data- and model-driven business insights and recommendations. Cutting-edge machine learning and AI algorithms are there at your fingertips, alongside the still-indispensable methods of classical statistics, through a rich ecosystem of high-quality open-source packages.

In practice, R and Python programming is used in, roughly speaking, two main modes. The first is to perform one-off analyses aimed at understanding or diagnosing a problem, identifying an opportunity, or reaching a data/model-driven recommendation. The second is to develop and test a prototype tool that, if successful, would then need to be deployed (or ‘productionized’) into a scalable analytics product. Python and R are ideal productivity languages in either mode. Note, however, that a scalable production-grade analytics product requires high-performance software design. Developing such software is the job of a professional software engineering team (not you!).

[§] All times are in Pittsburgh local time.

^{**} The Saturday office hour is dedicated to Online MSBA students only. The Thursday office hours may be joined by students from other programs/courses I teach.

This course will not magically transform you within seven weeks into a Python or R wizard. But it will give you the foundation to go far in exploratory data analysis and will prepare you well for the implementation of more advanced methodologies that you will acquire throughout your MBA program. You will get over the cold-start hump and learn how to become a life-long learner of data science software and tools. If successful, you would emerge from the course comfortable with programming and empowered to use code throughout your career. The only way to get there is through some effort and persistence. We will learn by doing and we will within the context of realistic business examples. This time investment will serve you well throughout your career.

3. Course Material

Class material will be posted on Canvas. There is no required textbook for the course. However, the following resources are recommended:

R for Data Science: Import, Tidy, Transform, Visualize, and Model Data, by Hadley Wickham and Garrett Golemund. Publisher: O'Reilly Media. ISBN: 978-1491910399. The online version is freely available at: <https://r4ds.had.co.nz/>

Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter, by Wes McKinney. 3rd Edition. Publisher: O'Reilly Media. ASIN: B0B9HY3WX7. Open edition freely available at: <https://wesmckinney.com/book/>

In addition, I will make various online notes (under development) permanently available at: <https://www.andrew.cmu.edu/user/afarahat/>

4. Grading

Your grade will be determined based on your performance on the following assessments according to the indicated weights:

Individual deliverables (equally weighted)	45%
Team deliverables (equally weighted)	20%
Final exam (take home)	30%
Attendance and online participation	5%

Details on the submission format of each deliverable and exam will be communicated during class. Online participation includes completing assigned Canvas surveys and concept-check short quizzes.

5. Note on Collaboration

The individual deliverables and the final exam should be completed independently. You should not discuss the solution details of individual work with other classmates ahead of due dates. However, you are allowed (and encouraged) to discuss *general* concepts, programming syntax, and examples covered during class with your classmates, your TA, and with me.

Work on the team deliverable with your team members only. Form teams of size 4 (or fewer) using the team self sign-up feature on Canvas.

6. Schedule of Classes

Classes 1-2 focus on the R programming language. Classes 3-7 focus on Python. Common computing concepts introduced in R will jump-start our exposition of Python.

Discussion Session	Date	Topics [†]
1	Wed, Oct 26	Intro to R; R tidyverse – data exploration
2	Wed, Nov 02	R tidyverse – data wrangling
3	Wed, Nov 09	Python programming – syntax and data structures (Part 1)
4	Wed, Nov 16	Python programming – syntax and data structures (Part 2)
	Wed, Nov 23	<i>No class – Thanksgiving week</i>
5	Wed, Nov 30	Python pandas – data exploration and wrangling (Part 1)
6	Wed, Dec 07	Python pandas – data exploration and wrangling (Part 2)
7	Wed, Dec 14	A preview of Python scikit-learn and plotly

[†] Tentative topics; may be adjusted to optimize learning during the course

The course cadence will be as follows. Except for Week 1, module content (asynchronous learning) will be posted by noon on the Thursday preceding the Wednesday discussion session. You are expected to complete the asynchronous learning portion and come with questions prior to the discussion session.

I will enable the piazza discussion app on Canvas so feel free to post questions there as well. I, the TA, and your colleagues will respond typically within 24 hours.

7. Schedule of Deliverables and Exam

Important note: Late submissions will not be accepted.

Deliverables / Exam	Individual / Team	Date and Time
Deliverable 1	Individual	Wed, Nov 09, 02:59 a.m.
Deliverable 2	Individual	Wed, Nov 16, 02:59 a.m.
Deliverable 3	Team	Wed, Nov 30, 02:59 a.m.
Deliverable 4	Individual	Wed, Dec 07, 02:59 a.m.
Final exam	Individual	Sun, Dec 18, 11:59 p.m.

8. Other General Guidelines and Resources

My goal as your instructor is to foster an interactive and stimulating classroom learning experience. You are strongly encouraged to participate in class by asking questions, answering questions, and volunteering remarks and experiences that enhance classroom discussions. If, due to time constraints, I am unable to fully address your question during class please feel free to follow up with me after class.

If at any point during the semester you feel unwell, including feeling anxious or stressed, then please know that CMU has many resources and trained professionals to help you. You can learn more here: <https://www.cmu.edu/wellbeing/index.html>.

If you have or suspect you have, a disability, I encourage you to contact CMU's Office of Disability Resources (<https://www.cmu.edu/disability-resources/>) as early as possible. We will work with you to ensure that appropriate accommodations are provided.

The University's policy on Academic Integrity (<https://www.cmu.edu/policies/student-and-student-life/academic-integrity.html>) governs all classes. Please make sure you have read it carefully and ask me when in doubt about any aspect that relates to this class. And please adhere to all classroom norms and expectations set forth by CMU Tepper.

It is imperative that we uphold CMU's high standards of professional conduct and integrity. We treat and help each other, and others, with respect, fairness, and compassion, inside and outside of the classroom.