

**CSC 223 – Python for Scientific Programming & Data Manipulation, Fall 2023, First Day Handout**  
**TuTh 4:30 – 5:45 PM, Old Main 159, <https://faculty.kutztown.edu/parson>**

My courses are multimodal this semester, meaning you can attend in-person or remotely via Zoom. I insist on maintaining 6 feet of distance from myself in order to reduce the odds of carrying the virus home to my wife & three-year-old granddaughter. I plan to wear a mask. I strongly encourage the unvaccinated to get vaccinated unless a medical condition precludes that. The unvaccinated will not only infect each other, they will also provide an environment in which the virus can mutate to more dangerous strains. I will post Zoom videos for all classes. The class time interactive Zoom link appears on **D2L** Course CSC220 -> Content -> Overview.

**Office Hours Monday 3-5 PM, Wed. 3-5 (Zoom only), Thurs. 9:50-10:50 or by appt. All available via Zoom.**

This is my first time for this course. I taught related material to Astronomy students as an Independent Study, assisted with the course design, designed and taught CSC523 Advanced Scripting for Data Science several times, and have done professional level research programming in Python since 2007. This course is a prerequisite for CSC458 and CSC459 in our data science programs. After some Python review for students who did not take CSC123, we will cover the following topics: accessing and manipulating textual and tabular data, automating statistical analysis, deriving and aggregating data including time-series derivations, performing time-domain to frequency-domain conversion, visualizing data relationships, and related topics. We will make heavy use of Python library models for statistics, data manipulation, data relationship analysis, and visualization.

**Prerequisite: (C or better in CSC123) or (C or better in CSC135).**

There is no mandatory textbook. Our course page will include links to textbook resources as we go along.

**Grading** (A = 92:100, A- = 90:91, B+ = 87:89, B = 82:86, B- = 80:81, C+ = 77:79, C = 70:76, D+ = 67:69, D = 60:66, F = 0:59) Projects 100% divided among the project assignment deliverables. No exams.

**Programming project assignment grading criteria**

We will work mostly on the acad or mcgonagall Linux servers.

Please follow my detailed requirements in assignment handouts.

Test everything before turning it in via **make turnitin** (not the turnin script). Test it after any changes.

When you think you are finished, read the requirements to avoid missing anything.

I will deduct points for missing documentation comments required in the handout.

**Programming project assignment grading criteria**

Grading rubrics will be part of each assignment handout. Late penalty is 10% per each day late, up until I go over the solution. Any assignment turned in after that is worth 0%.

We will use the CS&IT documentation requirements:

<https://faculty.kutztown.edu/parson/CSCDocumentationStandards.pdf>

**The academic integrity policy:**

<https://faculty.kutztown.edu/parson/AcademicIntegrityPolicy.pdf> Please read the policy statement.

You may openly discuss ideas, algorithms, pitfalls, and the use of programming tools.

You may not share code, test drivers or test data except within groups for group projects.

Please let me know if you have a preferred name or preferred personal pronoun not given in MyKU.

Attendance is not graded, but I will be teaching using data sources and concepts both inside and outside the scope of the textbook. You are responsible for all material covered in class, including technical information, coding standards and conventions, verbal specification of assignments, and your questions about topics that are not clear to you. Please, there should be no classroom conversations, cell phones, text messaging, eating, sleeping, obscenities, smoking (tobacco or artificial), vaping, listening to music or other disruptions of the class. I will deduct 5% from an assignment for each infraction. If you have already disclosed a disability to the Disability Services Office (215 Stratton Administration Building) and are seeking accommodations, please feel free to speak with me privately so that I may assist you. If you have an injury sustained during military service including PTSD or TBI, you are also eligible for accommodations under the ADA and should contact the Disability Services Office. If you have preferred pronouns for yourself, or a name that differs from the MyKU roster, please let me know.

Any course work submitted to the instructor (including but not limited to assignments, tests, and projects) may be photocopied and retained for the purpose of assessment, accreditation and quality improvement, after removal of any information identifying the student.

The plan in this page is somewhat tentative since this is the first time I am teaching this course. We will cover most of these topics in approximately this order. There will be one in-class work session each time I hand out an assignment. Please read it ahead of the session.

W1	Python dynamic data types, primitive & aggregate types, type conversion via construction. Implicit parametric polymorphism!
2	Named and anonymous functions, functions as first-class objects, list comprehensions. Higher order functions, built-in generator types, custom closures and generators.
3	Statistical measures and statistical distribution libraries. <b>Hand out assn1.</b>
4	Trigonometric and inverse trig functions. Identifying and deriving linear and cyclic data relationships
5	Generating periodic data. Transforming data from time domain to frequency domain for analysis. <b>Assn2.</b>
6	A Python data analysis framework for the analysis half of Assignment 2. Class work time.
7	Python classes & objects, static & dynamic methods, events, object-oriented programming.
8	Python generators, closures, coroutines, and their application in iterating over CSV datasets. <b>Assn3</b>
9	Introduction to Pandas.
10	Using Pandas to simplify and extend a previous project. <b>Assn4.</b>
11	Exceptions, the Exception and Type library extensions to the base language. Work session.
12	Capstone project discussion and initiation. <b>Assn5.</b>
13	Overview of machine learning topics and data visualization. CSC523 and the grad data analytics cert.
14	Consolidation. Work session.
15	<b>Final project work session.</b>

Each of the **assn[1,5]** above is a planned assignment handout. It will be due during the week of the following assignment handout. I am trying to pace these to have a 2.5 week turnaround time,  $5 \times 2.5 = 12.5$  weeks, + 2.5 week introduction.

**Assn1** – Statistical measures and statistical distribution generation and characterization.

**Assn2** – Generating periodic data. Transforming data from time domain to frequency domain for analysis.

**Assn3** – Python classes, generators, closures, and their application to data manipulation.

**Assn4** – Using Pandas to simplify and extend a previous project.

**Assn5** – Individual, per-student capstone project to test and extend your knowledge.

We will be using Zoom for remote attendance during class time. Recorded archives of class sessions will be available within a day. We will go over Zoom & recording permissions in the first class.