CSC 223 – Python for Scientific Programming & Data Manipulation, Fall 2024, First Day Handout

Monday & Wednesday 3:00-4:15 and 4:30-5:45 PM, Old Main 159, https://faculty.kutztown.edu/parson
My courses are multimodal, 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 any virus home to my wife & four-year-old granddaughter. I plan to wear a mask. I will post Zoom videos for all classes.
The class time interactive Zoom link appears on **D2L** Course CSC223 -> Content -> Overview.
Office Hours Monday 11 AM-1 PM, Wednesday 12-2 PM, Thursday 4-5 PM, or by appt. All available via Zoom.
This course is a prerequisite for CSC458 and CSC459 in our data science major & minor programs. After some Python review for students who did not take CSC123, we will cover the following topics: statistical data generation and / or analysis, numeric data correlation, accessing and manipulating textual and tabular data, deriving and aggregating data, visualizing data relationships, Python features supporting dataflow architectures, and related topics. We will make heavy use of Python library models for statistics, trigonometry, 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 constitute 100% of the grade divided among the project assignments. No exams.
Programming project assignment grading criteria

We will work mostly on the **K120023GEMS.kutztown.edu** Linux server, ssh'ing in from acad.
Please follow my detailed requirements in assignment handouts.
Test everything via **make test** before turning it in via **make turnitin** (not the turnin script).
Test it after any changes. When you think you are finished, reread the specs 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.

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. Control constructs. Overview of Python as used in CPSC Department's Scripting Certificate & Data Science major & minor.
2	Named and anonymous functions, functions as first-class objects, list comprehensions. Higher order functions, built-in generator types, custom closures and generators.
3	Classes & objects, static & dynamic methods, generating statistical distributions, libraries. **Hand out assn1**.
4	Work time. Trigonometric and inverse trig functions. Deriving linear data relationships from non-linear relationships.
5	Identifying and deriving linear and cyclic data relationships. **Assn2**. Work time.
6	Manipulating sequences and tables of data. Input & output from comma-separated-value (CSV) files.
7	More on libraries related to Assignment 3. **Assn3**.
8	Using Python closures, generators, and coroutines to build a *pull*-dataflow software architecture.
9	Transforming data from time domain to frequency domain for analysis via a dataflow architecture.
10	Statistical averaging and smoothing techniques for data series. **Assn4**.
11	Work time. Using Python closures, generators, and coroutines to build a *push*-dataflow software architecture.
12	Overview of additional libraries & topics as time allows.
13	Machine learning and sci-kit learn overview. **Assn5**
14	Consolidation. Work session.
15	**Final project work session.**

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 X 2.5 = 12.5 weeks, + 2.5 week introduction.

Assn1 – Generating & extracting cyclic statistical data.
Assn2 – Statistical and trigonometric distributions for numeric values, extracting statistical scalars, polynomial -> root AND exponential -> logarithmic data compression and decompression, Pearson correlation coefficient.
Assn3 – Tabular data manipulation including reading, examining, deriving, and writing data.
Assn4 – Building a *pull*-dataflow software architecture.
Assn5 – Building a *push*-dataflow software architecture

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.