

BMGT 404: Essential Data Skills for Business Analytics (Spring 2020)

Decisions, Operations & Information Technologies
Robert H. Smith School of Business

Instructor: Kunpeng Zhang (kzhang@rmsmith.umd.edu)
Lectures: Monday/Wednesday, 8:00 – 9:15 (0401) and 9:30 – 10:45 (0501)
Room: VMH 3522
Office Hour: TBD
Room: VMH 4316

Teaching Assistant: TBD



Textbook 1: **Think Python – How to Think Like a Computer Scientist**
[Buy this book at Amazon.com](#)
[Download it here in PDF](#)
[Read it in HTML](#)

Textbook 2: **The Python Tutorial**, available from the Python website. This is good for explaining the nuts and bolts of how Python works.

About the course

Since the data grows exponentially and becomes complex, we need computational methods to collect, store, and analyze them in order to be successful in science, engineering, business, and other professions. BUDT758X, “Data Processing and Analysis in Python”, is an introductory programming class that meets this need. You will learn how to write computer programs in Python language to solve real-world problems and how to use tableau to explain your results as a report in a more readable way. This will be useful in your research and your jobs in the future.

This class is designed for students that want to learn to computer programming for data science. This course guides students through the basic Python programming language, from initial concepts to final data analysis using python and external packages.

The (tentative) list of topics that we plan to cover:

- Basic programming concepts
- Install and run Python programs
- Variables and expressions
- Control structures: sequence, branch, iteration
- List, tuple, dictionary in Python
- Functions, parameters, and recursion in Python
- String operations
- Regular expression
- File operations and modules
- Visualization
- Data manipulation: Pandas
- Database operations (MySQL)
- Scientific computing: NumPy
- Natural language processing: NLTK
- Machine learning: scikit-learn

Prerequisites

The course does not require any prior programming background or experiences. **Students that enroll in the class are expected to have some basic familiarity with programming in other languages**, at the introductory level (i.e., R, matlab, SAS, etc.), **however no prior knowledge of *Python* is required. Since this course contains hands-on labs, (I prefer) you are expected to bring your laptop to every class (and remember to charge it, so that it lasts for the duration of the class).**

Course Objectives

At the completion of this course, students will be able to:

1. Understand the architecture, basic elements, and the planning of data science
2. Understand and use standard sequential, conditional, and iterative control structure of automated data analysis through computers
3. Understand and use off-the-shelf packages to solve business related applications, such as resource allocation, finance, accounting, information system management, and many others.
4. Understand how to manipulate data (store, query, and summarize) using database for analyzing structured data
5. Understand and use computer programming for basic interactive web applications
6. Understand and use computer programming to collect, analyze, and visualize business data

Tools we will use

The tools that we will learn to use in the class include:

- Jupyter Notebook (<https://jupyter.org/>)

Scientific computing package NumPy (<https://numpy.org>)

- Using Python and SQL to create and query relational databases, in particular MySQL
 - Pandas (<http://pandas.pydata.org/>) for handling and visualizing structured data and time series
 - Machine learning packages: scikit-learn (<http://scikit-learn.org/stable/documentation.html>)
-

Resources

While we do not have any required textbook for the course, the following books will be useful references for the material that we will be covering in class.

- nbviewer: A simple way to share Jupyter Notebooks (<http://nbviewer.jupyter.org/>).
 - [Learn Python the Hard Way](#) This book is also available [for free on the web](#). We will not be really using this book but it is a useful textbook for those that have never been exposed to programming before.
 - [Python for Data Analysis](#): This is a useful book, discussing in details the “Python pandas” library, which provided much useful functionality for data management in Python.
 - [Pandas](#) (Python extension), Transforms Python into a great tool for spreadsheet and data frame manipulation
 - [Python Data Visualization Cookbook](#): This is a book for those interested in studying deeper visualization in Python.
 - [Google’s Python class](#)
 - [Learn Python the Hard Way](#)
 - [Code Academy, Python track](#)
 - [Useful IPython notebooks](#)
-

Assignment

You are free to submit late, but there is a **10%** grade penalty for every additional day after the deadline. Given the generous late submission policy, penalties are strictly enforced, and no extensions are granted. Please plan accordingly, and do not leave submission for the last minute.

Plagiarism Policy: Inevitably in a programming course, it seems that a few people will turn in work that is not their own. You should understand that it is usually easy to detect copying of programs – even when a program is modified to try to disguise its source. **Copying a program, or letting someone else copy your program, is a form of academic dishonesty and the penalties can be found [here](http://www.rhsmith.umd.edu/about-us/academic-integrity) (<http://www.rhsmith.umd.edu/about-us/academic-integrity>).**

Grading

Your final grade for the course will be composed from the following items:

Participation	$5\% * 1 = 5\%$
Midterm exam	$25\% * 1 = 25\%$
Quizzes	$10\% * 2 = 20\%$
Final project	$15\% * 1 = 15\%$
Assignments	$5\% * 7 = 35\%$

Letter grades are assigned as follows:

Letter	Points
A+	100 – 97
A	96.9 – 93
A-	92.9 – 90
B+	89.9 – 87
B	86.9 – 83
B-	82.9 – 80
C+	79.9 – 77
C	76.9 – 73
C-	72.9 – 70
D+	69.9 – 67
D	66.9 – 63
D-	62.9 – 60
F	Below 60

Academic Integrity

The University is an academic community. Its fundamental purpose is the pursuit of knowledge. Like all other communities, the University can function properly only if its members adhere to clearly established goals and values.

The University's Code of Academic Integrity is designed to ensure that the principles of honesty and integrity are upheld. You are expected to adhere to this Code. The Smith School does not tolerate academic dishonesty. All acts of academic dishonesty will be dealt with in accordance with the provisions of this Code. Anyone suspected of academic dishonesty will be referred to the Office of Student Conduct immediately. Please visit the [website](#) for more information on the University's Code of Academic Integrity.

Academic Dishonesty: any of the following acts, when committed by a student, shall constitute academic dishonesty:

- Cheating: intentionally using or attempting to use unauthorized materials, information, or study aids in any academic exercise.
- Fabrication: intentionally and unauthorized falsification or invention of any information or citation in any academic exercise.
- Facilitating academic dishonesty: intentionally or knowingly helping or attempting to help another to violate any provision or the Code of Academic Integrity.
- Plagiarism: intentionally or knowingly representing the words or ideas of another as one's own in any academic exercise. To avoid plagiarism, always cite your sources. You can view more about citing sources at: <http://www.lib.umd.edu/tl/guides/citation-tools>. Feel free to use any of the styles listed here.

To help you avoid unintentional cheating, the following table lists levels of collaboration that are acceptable for each type of graded exercise. If you are ever unclear about acceptable levels of collaboration, please ask!

	 OPEN NOTES	 USE BOOK	 SEARCH ONLINE	 ASK FRIENDS	 WORK IN GROUPS
Assignments & labs	✓	✓	✓	✓	✗
Midterm	✗	✗	✗	✗	✗
Final	✗	✗	✗	✗	✗

Attendance

University policies excuse the absences of students for illness, religious observances, participation in University activities at the request of university authorities and compelling circumstances beyond the student's control. Regular attendance and participation in this class is the best way to grasp the concepts and principles being discussed. However, in the event that a class must be missed due to illness, the policy in this class is as follows:

1. For every medically necessary absence from class (lecture, recitation, or lab), a reasonable effort should be made to notify the instructor in advance of the class. When returning to class, students must bring a note identifying the date of and reason for the absence, and acknowledging that the information in the note is accurate.
2. If a student is absent more than 3 times, the instructor may require documentation signed by a health care professional.

3. If a student is absent on days when tests are scheduled, he or she is required to notify the instructor in advance, and upon returning to class, bring documentation of the illness, signed by a health care professional.
 4. No students can receive A if **absence 5 or more classes**.
-
-

Disability Services

If you are a student with a disability that requires accommodation, you must provide me with documentation from Disability Support Services (DSS) by *the second week of class*, so we can make appropriate arrangements. If you have a disability, but have not yet registered with the University, please do so by contacting DSS as soon as possible (301.314.7682 or 301.405.7683 TDD). Accommodations cannot be made without appropriate documentation.

Technology

The use of cellular phones, tablets, or computers for non-course purposes will not be allowed without the prior consent of the presiding faculty member. Students using phones during class or participating in other disruptive activities will be asked to leave out of respect for fellow students and faculty. Eating or drinking during skills activities, texting/web surfing, working on other courses' material, or other activities that distract from course activities are not allowed. Audio or video recording of any course activity needs express permission of the instructors.

Tentative Schedule

Here is a tentative schedule of lectures, readings, and labs for this course. We will try to keep approximately to this schedule.

(Note that we may change the schedule during the semester.)

Session	Topic
1	Introduction to Python Install and run a Python program (Jupyter Notebook)
2	Variables, expressions, and statements
3	Condition statements
4	Loop structure
5	Lists, tuples, and dictionaries
6	Functions, parameters, and recursion
7	File operation and modules
8	String operations
9	Midterm
10	Regular expression
11	Visualization
12	Database operations, MySQL

13	Data manipulation in Python: Pandas
14	Text mining in Python
15	Scientific computing: NumPy
16	Machine learning in Python
17	Project