

70-374: Data Mining & Business Analytics

Fall 2017

Instructor: John Gasper

Office: CMUQ 2160

Email: gasper@cmu.edu

Course Time / Location:

MW 09:00 - 10:20AM, CMUQ 2035

Office Hours:

Tuesday 1-3pm & by appointment

In general, I have an open door policy: if my door is open and I'm not meeting with someone, you are welcome to come in and meet with me. If my door is closed, I am not available (out of the office, working, etc). On non-teaching days, I'm often quite busy with research and not available. I highly encourage you to **set up an appointment** to make sure I will be available.

Required Textbooks:

- *Machine Learning In R* by Bret Lantz
- *Data Science for Business* by Foster Provost & Tom Fawcett
- various handouts

Course Description and Goals

This course is an introduction to data analytics, data mining, and data-driven decision-making. Interest in big data analytics has exploded in the past few years. This growth has been partially driven by the availability and quality of data that enables managers to ask and answer questions they were never able to consider. Data mining enables one to extract useful insights, which then can be utilized for data-driven decision-making and competitive advantage.

Data mining and data analytics involve a collection of techniques for extracting patterns and trends in large databases. This course is a hands-on introduction to these areas with

an emphasis on aspects useful to business managers. At the end of this course, students will better understand the need and appropriate place for data mining, the major techniques used in data mining, and the important pitfalls to avoid.

Throughout the course, we will use two software packages that are commonly used throughout industry. The first is Tableau, a data visualization package, and the second is R, a powerful open-source statistical language. Both of these tools are becoming mainstay workhorses of business analytics.

Course Logistics

This course has a Canvas site. The sites should set up and functioning. Our class page can be accessed via the following URL:

<https://cmu.instructure.com/courses/1002>

Handouts, problem sets, updated syllabi and announcements will be posted to Canvas and you are responsible for checking the site regularly. I will also maintain the Canvas discussion board. If you have any questions about the techniques, problem sets, etc, ask them on the discussion board. It has been my experience that one of the best ways to learn something is to try to explain it to someone else. So *I will expect you to try to answer the questions that other students ask*; doing so will aid the participation element of your grade.

I welcome questions during class: if you have a question or a comment, please let me know. I will generally pause after each slide and ask if there are any questions. *Please feel encouraged to raise questions during class*. I am also fairly accessible via email, but you should not expect a reply immediately (within 24 hours).

Grades

Each student's grade for the course will be based on the following:

1. *Visualization Challenge* 10%
2. *Data Prediction Challenge* 10%
3. *Homework Problem Sets/Quizzes* 20%
4. *Exam* 20%
5. *Case study* 15%
6. *Final Project Competition* 25%

The only way to learn the material is to do it. There will be multiple problem sets distributed that will be graded. Some of the problem sets will be graded on a "check-minus / check-plus" system where credit will be given for completing the problem set, and others will be graded on a correctness scale, with each problem set indicating the grading scheme.

For those graded on the check-system, a check will mean that you've reasonably attempted the problems; a check-plus is awarded for exemplary work (i.e., I could use it as a solution set next year) and a check-minus for a poor and deficient attempt. Solution sets will be posted and *you will be responsible for checking that your work is correct*. There will be no personal extensions to the homework deadlines because I will post the solutions right after the homework deadline. On the top of the first page of the assignment, you **must list everyone you worked with on the assignment**.

I know that your schedule during the semester can be hectic. I also know that various events can happen during the semester that make finishing projects on time difficult. I also, however, expect you to know these things as well. I have a no personal extensions and no make-up policy, unless it is a university approved absence. **If you attend 90% of the classes, your lowest quiz score will be dropped when calculating your final grade.** Again, you are required to bring your clicker to every class once it is assigned. I will be using these to take attendance for the course. If you forget to bring your clicker and we use it that day, you will be counted as absent.

Accommodations for Students with Disabilities

Carnegie Mellon University is committed to providing reasonable accommodations for all persons with disabilities. To access accommodation services you are expected to initiate the request and submit a Voluntary Disclosure of Disability Form to the office of Health & Wellness or CaPS-Q. In order to receive services/accommodations, verification of a disability is required as recommended in writing by a doctor, licensed psychologist or psycho-educational specialist. The office of Health & Wellness, CaPS-Q and Office of Disability Resources in Pittsburgh will review the information you provide. All information will be considered confidential and only released to appropriate persons on a need to know basis.

Once the accommodations have been approved, you will be issued a Summary of Accommodations Memorandum documenting the disability and describing the accommodation. You are responsible for providing the Memorandum to your professors at the beginning of each semester. For more information on policies and procedures, please visit <https://scotty.qatar.cmu.edu/qword/student-affairs/office-of-health-and-wellness/assistance-for-individuals-with-disabilities/>

Health & Well-being

Take Care of Yourself: Do your best to maintain a healthy lifestyle this semester by eating well, exercising, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress.

All of us benefit from support during times of struggle. You are not alone. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is often helpful.

If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support. Counseling and Psychological Services (CaPS-Q) is here to help: call 4454 8525 or make an appointment to see the counselor by emailing student-counselling@qatar.cmu.edu. Consider reaching out to

a friend, faculty or family member you trust for help. If you or someone you know is feeling suicidal or in danger of self-harm, call someone immediately, day or night at 5554 7913

If the situation is life threatening, call 999

Academic Integrity

You should feel encouraged to talk with your class mates about the problems on the problem sets, but do not copy even parts of someone else's work. The homework is graded on a check system to encourage you to attempt the homework yourself. ***While I highly encourage you to use the Piazza discussion site, if you speak with anyone else (including a TA/CA or the ARC) regarding the homework, I require that you list it.*** Many students also do not realize that using a homework assignment from a previous iteration of the course to aid them in attempting their problem sets is also an academic integrity violation.

The CMU-Q policy on cheating and plagiarism has been updated and I would like to point out the following text: *In all academic work to be graded, the citation of all sources is required. When collaboration or assistance is permitted by the course instructor(s) or when a student uses the services provided by Academic Development, the Global Communication Center, and the Academic Resource Center (CMU-Q), the acknowledgement of any collaboration or assistance is likewise required. This citation and acknowledgement must be incorporated into the work submitted and not separately or at a later point in time. Failure to do so is dishonest and is subject to disciplinary action.*

I am very sensitive to cheating and plagiarism; my policy is that cheating of any kind will not be tolerated. My automatic penalty for any offense is a one letter grade reduction in your final course grade. If you have any doubt about your actions, please ask me. I strongly encourage you to review Carnegie Mellon's policies regarding academic integrity.

Weekly readings and course outline

The following is a **very tentative** schedule for the semester. This is the first iteration of this course for me and you should expect that quite a few wrinkles will need to be ironed out. The following gives you an idea of what we'll cover and the approximate workload I envision for the course:

Class	Date	Topic	Reading	Assignment
1	8/21/17	Intro / Data Vis	handouts	
2	8/23/17	Tableau	handouts	
3	8/28/17	Tableau	handouts	PS0 Due
4	8/30/17	Tableau	handouts	PS1 Due; Quiz 1
	9/4/17	Eid Break		
	9/6/17	Eid Break		
5	9/11/17	Tableau	handouts	
6	9/13/17	Tableau	handouts	Viz Challenge
7	9/18/17	Intro to R / Managing Data	ML Chp 2	PS2 due
8	9/20/17	Intro to R / Regression	ML p186-200	
9	9/25/17	Classification: Decision Trees	ML Chp 5; DS Chp 3	
10	9/27/17	Classification: KNN	ML Chp 3	
11	10/2/17	Classification: KNN	ML Chp 4	PS3 due
12	10/4/17	Classification: Nave Bayes	ML Chp 4	
13	10/9/17	Classification: Nave Bayes	ML Chp 5	
14	10/11/17	Classification: Logistic Regression	DS p 97-102; handout	
15	10/16/17	Classification: Logistic Regression	handouts	PS4 due
16	10/18/17	Classification: SMVs	ML p 239-256	
17	10/23/17	Classification: SMVs		
18	10/25/17	Market Basket Analysis	ML Chp 8	
19	10/30/17	Market Basket Analysis	ML Chp 9	PS5 due
20	11/1/17	Clustering	ML Chp 9	
21	11/6/17	Clustering	ML Chp 9	
22	11/8/17	Extra Topics (time permitting)	TBA	
24	11/13/17	Extra Topics (time permitting)	TBA	PS6 due
25	11/15/17	Extra Topics (time permitting)	TBA	
26	11/20/17	Extra Topics (time permitting)	TBA	
27	11/22/17	Visualizing performance	DS Chp 8	
28	11/27/17	Visualizing performance	TBA	
29	11/29/17	Final Presentations		