

AIT 580 – Analytics: Big Data to Information

Section: DL4 (Online)

Collaborative Session: Wednesday 9 PM (1- to 1.5-hour session, starting 1/26)

Location: Online. Blackboard Collaborate Ultra (<https://mymason.gmu.edu>)

Instructor: Myeong Lee

Instructor E-mail: mlee89@gmu.edu

Instructor Office: ENGR 5304

Instructor Office Hours: Wednesday 2 PM to 3 PM (appointment in advance)

TA: TBA

Catalog Description

This course provides an overview of Big Data Analytics concepts, tools, and methods, and their use in commercial, scientific, governmental, social, and other application areas. Topics include technical and non-technical disciplines required to collect, process, and interpret enormous amounts of data available from numerous public sources. Course content includes discussions of, and hands-on practice with, technologies involved in collecting, mining, analyzing, visualizing data, and interpreting the results. Additional topics covered include system infrastructure and acquisition, law and policy, and ethical issues related to data collection.

This course is currently one of the core requirements for the IT Management concentration for the MS in Applied Information Technology; it can also be taken as an elective for the MS/AIT concentrations.

Prerequisite(s): Graduate Standing

Recommended recent previous courses: Programming (Python); Statistics

Topic Spectrum of the Course

The course includes but is not limited to the following topics and emphasizes the technical aspects of Data Analytics projects:

- Characteristics and representations of data
- Architecture of data analytics systems
- Tools for conducting data analytics research

- Roles of the Data Scientist and the Data Analytics Project Manager
- Public Big Data resources
- Big Data Use Cases in Social Media, Government, and Industry
- Principles of statistical and computational analysis
- Querying data using SQL
- Basic use of R, Python, and other software tools
- Data visualization concepts and methods
- Data governance, security, curation, privacy, and legal issues
- Machine learning and predictive analytics

Course assignments include:

- Big Data case study
- Topic-specific readings and exercises
- Discussions of Big Data Analytics topics
- Lab exercises using Data Analytics software
- Data Analytics project

Assignment Submission Format

Document-based Assignments: Unless otherwise specified, all assignments that are related to general description/visualization must be submitted as PDF files. Upload all assignments to Blackboard. Use the following filename format: Lastname-Assignmentname.pdf.

For example: Lee-Assignment1.pdf

Code-based Assignments: Unless otherwise specified, all assignments that consist of scripts (R or Python) need to be submitted through Github. Commit and push your assignments (i.e., code) to your Github repo and you need to add the instructor (ID: “myeong” on Github) as collaborator for your assignment repo. Then, you need to submit your assignment URL through Blackboard.

There is several Computer Labs available for general use by students, which are located on the Fairfax campus. For more information go to the web site at <http://labs.vse.gmu.edu> Otherwise, students are expected to use their personal laptops for course work, along with assignment-specific resources such as Amazon Web Services

Class Participation

Contribute actively and participate in online discussion topics posted on Blackboard. Some online discussion topics will be selected for grading.

Required Course Materials

Optional Textbook:

- The Data Science Design Manual (Texts in Computer Science) 1st ed. 2017 Edition by Steven S. Skien <http://www.data-manual.com/>

Recommended Readings:

- Big Data Science & Analytics: A Hands-On Approach (BDSA), A. Bahga & V. Madisetti, 2016 VPT <http://www.hands-on-books-series.com/bigdata.html>
- The Book of R, T. Davies, 2016, No Starch Press <https://www.nostarch.com/bookofr>
- An Introduction to Statistical Learning with Applications in R, G. James, et al., 2016 Springer <http://www-bcf.usc.edu/~gareth/ISL/index.html>

Tutorials

- A variety of tutorials (SQL, R, Statistics, Python, etc.) are available at <http://lynda.gmu.edu/>
- Seeing Theory: <http://students.brown.edu/seeing-theory/>
- The Statistics Tutor's Quick Guide to Commonly Used Statistical Tests <http://www.statstutor.ac.uk/resources/uploaded/tutorsquickguidetostatistics.pdf>

There will be required Web-based readings and tutorials, along with recommended resources. These will be listed on the course Blackboard Course Resources link and within individual assignment links.

Other Expectations

- Some familiarity with Linux operating system
 - Adding Linux VM to Windows or to OS X using VirtualBox
- Programming concepts
 - Using R, Python, Tableau, other data analytics tools
 - Basic statistics
- Self-study, tutorials, video lectures etc.

Course Duration (Online)

- Dates: Week of August 26 through Week of December 16 - 2019
- Total Duration: 16 weeks

Form of Lecture

This course is designed for "asynchronous" delivery (see <https://masononline.gmu.edu/course-delivery-methods>). The instructor will have an online meeting on Blackboard Collaborate Ultra; there will be no penalty for non-participation in the optional online meetings. These online meetings will be recorded and available to students for the semester.

Grading Policy

Student grades will be determined based on class participation/blackboard discussion board, general assignments, lab exercises, case studies, reports, and final project.

Grade Component	Weight
Class Participation (graded and ungraded online discussions)	10%
General assignments & lab exercises	40%
Case studies and reports	20%
Final Project	30%

Grading Guidelines

Some grade components could be evaluated *subjectively*

- A: consistently above and beyond the course/assignment requirements
- B: meets and occasionally exceeds the course/assignment requirements
- C: minimally meets the course/assignment requirements
- F: fails to meet the course/assignment requirements

Grades will be awarded in accordance with the Mason Grading System for graduate students. See the university catalog for policies: <http://catalog.gmu.edu> for more information.

Grading Scale/Schema

The grading scale for this course is:

Grade Distribution		
97 – 100%	A+	Passing
94 – 96.99%	A	Passing
90 – 93.99%	A-	Passing
87 – 89.99%	B+	Passing
83 – 86.99%	B	Passing
77 – 82.99%	B-	Passing
70 – 76.99%	C	Passing
0 – 69.99%	F	Failing

Raw scores may be adjusted by the Instructor to calculate final grades.

Students are responsible for checking the currency of their grade books. Grade discrepancies must be brought to instructor's attention within one week of assignment submission and 48 hours of exam submission.

Honor Code

All work performed in this course will be subject to [GMU's Honor Code](#). Students are expected to do their own work in the course unless a group project is approved by the instructor. In papers and project reports, students are expected to write in their own words, rather than cutting-and-pasting from sources found on the Internet. The goal of assignments is to demonstrate what you have learned, not what you can google. When you do use text or graphical material from books, articles, and the Web, enclose the material in quotes and provide a complete and proper reference. If a paragraph is used then it should be indented in the text (both left and right margins). In-text citation can use the [Author, Year] format or the Numerical [1] format which must refer to the source in the References section of your assignment. Use [Chicago Manual of Style](#) for guidance on citation style, usage, etc. (Don't buy the big CMS. See the smaller [A Manual for Writers](#) by Kate Turabian). Regardless of the citation method used, proper citations always include: Author(s), Title, Publication Date, Publisher, and URL (if from the Web, along with Last Accessed Date). BlackBoard's SafeAssign service will be used to review selected student assignments.

Notes

- [Wikipedia is not a primary reference](#). Use it for initial discovery, but use and cite primary references (which Wikipedia itself might use).
- If you need assistance with writing an assignment, you can get assistance here: <http://writingcenter.gmu.edu>
- The first item instructor looks at in assignments is the References section!
- Any programming/coding assignments must adhere to the [CS Honor Code](#).

Other Notes

- There will be reading assignments along with the assumption that you have actually read them.
- Lecture slides from instructor's material will be posted on Blackboard.
- Some material you may already know (good! that's review!) Some material you may have learned earlier and have forgotten (good! you'll be reminded) Some material you may know more than the instructor (good! share it!).

- Call or email the instructor if you anticipate being unable to meet any course requirements in a timely manner.
- Personal Safety and Security: The Mason Alert system provides emergency information of various sorts. Students can sign up for it by visiting the website <https://alert.gmu.edu>. Students are also reminded that an emergency poster exists in each classroom explaining what to do in the event of crises and that further information about emergency procedures exists on <https://ready.gmu.edu/be-prepared/>
- Computer and IT Security: Visit GMU's IT <http://itsecurity.gmu.edu/> web site regularly. Norton AntiVirus Software is free to download for all GMU students/faculty/staff.