MSDS 402: Introduction to Data Science

Course Description

This course introduces the field of data science, which combines business strategy, information technology, and analytic modeling. The course reviews applications and methods in data science, explores implementation and development of analytics in organizations, and introduces foundational concepts in data engineering. Students evaluate business problems and solutions in traditional and contemporary data management systems, selecting appropriate tools for data collection, preparation, warehousing, and analysis. The course culminates with students producing an original, professionally presented research plan that includes sampling and survey design for a contemporary research problem.

Course Objectives

By the end of this course, you will be able to:

- Articulate the many applications of data science.
- List and evaluate a variety of methods used in data science.
- Articulate the value of data science in organizations.
- Rank an organization's analytic maturity.
- Address data quality issues.
- Discuss basic principles of data engineering, including database architecture and data warehousing.
- Define and describe Big Data, machine learning, and artificial intelligence.
- Articulate means of collecting, preparing, storing, and analyzing data.
- Evaluate and select appropriate sample, survey, and data collection methods.
- Create a professionally presented research plan and proposal for data collection.
- Develop and create a survey questionnaire.
- Recognize and address sources of bias and error.
- Determine sample size using basic probability sampling methods.
- Identify and discuss core research ethics issues and compliance requirements.
- Use the Northwestern library online to locate refereed articles for data science research.
- Develop professional-level written communications skills.

Prerequisites

There are no prerequisites for this course. This is an introductory course that does not require any special skills.

Software

You do not need R and/or Python for this course.

Required and Optional Books and Resources

Required Books

Davenport, T. H., & Harris, J. G. (2017). *Competing on analytics: The new science of winning (Updated)*. Boston, MA: Harvard Business School.

Miller, T. W. (2015). *Modeling techniques in predictive analytics with Python and R: A guide to data science.* NJ: Pearson FT Press.

Provost. F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking.*Sebastopol, CA: O'Reilly.

Course Reserves

Some readings will be available through the Course Reserves in the left navigation menu. Assignment and Discussion forum instructions will note which readings are to be accessed through Course Reserves. For assistance with Course Reserves, e-mail e-reserve@northwestern.edu. To ask a librarian for assistance, visit Northwestern's Ask A Librarian page.

Optional Readings and Resources

If available, posted in the weekly module.

Assignment Overview and Grading Breakdown

Grading and feedback turnaround will be one week from the due date. You will be notified if turnaround will be longer than one week. The discussion forums, written assignments, and midterm and final presentation will be graded based on specific criteria or a rubric. The criteria or rubric for each type of assessment is available in the course shell. Your final grade will be determined as follows:

Discussion Forums	Each week, students will participate in group discussions, answering posted questions and relating personal and professional experience to course materials. Weekly discussion board participation is required, with at least one unique individual post per week along with discussion and responses to other students' posts. See additional guidelines for discussion board in the syllabus. (10 points/week. 100 points total.) Due: End of week for each week of term.	12.5%
Individual Short Assignments	One shorter assignment is due at the end of the first. Week 1 - "Moneyball" Review (40 points).	5%
Short quizzes	Four short quizzes: Week 2 - APA style prep (25 points). Wee 4: Data engineering essentials quiz (25 points). Week 5: Data processing quiz (25 points). Week 7 - Sampling (25 points).	12.5%
Individual Assignments	Two longer individual assignments are due the third and sixth weeks of the term. Week 4 – Applications and Methods in Data Science (100 points). Week 6 – Data Processing/Machine Learning assignment (100 points).	25%
Team Project	Teams will be assigned during the fourth week. The first deadline for the group's Survey Design and Implementation	32.5%

	Project will be in the sixth week, and the final project is due in Week 9, with team checkpoints along the way in which drafts of the sampling design, survey, and final paper are submitted for review. By the end of Week 6: first team checkpoint - Teams submit topic proposals (30 points for timely submission). By the End of Week 7: second team checkpoint - Sampling Design/Survey DRAFT (40 points for timely submission). By the End of Week 8: third team checkpoint - Rough Draft of the Paper or Survey Design (40 points for timely submission). End of Week 9 - Final Survey Design and Implementation Team Project Paper Due (150 points).	
Final Exam	Comprehensive proctored final exam to be taken during the 10th week (100 points).	12.5%
Total	800 points.	100%

Grading Scale

A 93-100 %

A- 90-92 %

B+ 87-89 %

B 84-86 %

B- 80-83 %

C+ 77-79 %

C 74-76 %

C- 70-73 %

F Below 70%

Late Work Policy

Unless otherwise noted, all work is due on the assigned day by 11:55 pm CDT (central time). This includes papers and participation in the discussions.

Late work is not accepted without prior agreement of the instructor. If you think you might be late, email me and we can figure out a solution. Do it early - I am pretty flexible, but don't have too much sympathy for someone who waits to the last moment.

Try not to fall behind in this course. We cover a lot of material, and falling behind is a primary reason folks don’t succeed. Plan ahead, and always contact your instructor or academic advisor if you begin to fall behind or encounter an unanticipated event that may interfere with your coursework

Online Communication and Interaction Expectations

Discussion Forums

Discussion board participation is an essential and important part of this class, and is designed to allow free exchange of ideas in a respectful and open environment. How often you post is less important than the contents of your contribution, although a minimum level of engagement is expected. You are encouraged to post actively and frequently, but please try not to clutter the board with irrelevant or insignificant material, which could work against you. Stay on topic, keep your language professional (abbreviated texting language is not appropriate), and try always to offer something new when you post (a “me too” type post doesn’t count). When relevant, please remember to cite all sources, and avoid plagiarism.

I want to see a minimum of three posts in discussion session per week. One of those posts an original contribution to the discussion that when necessary includes references and citations, while being careful not to plagiarize or violate copyright. Also demonstrate engagement in the discussion forum by responding to others’ posts. This will make the forum easier to read. Explain, clarify, politely ask for details, provide details, persuade, and enrich communications for a great discussion experience. Also, it is a discussion board - give everyone a chance to discuss your posts and take the chance to discuss others' posts, i.e. don't wait until the last moment to do your posts. Original postings done early in the week (by early I mean before Thursday) have more opportunity for back and forth which is the best way to learn from each other.

Participation in discussion is worth up to 10 points per week (100 for the course).

It is helpful to read this guide to netiquette .

Participation and Attendance

This course will not meet at a particular time each week. All course goals, session learning objectives, and assessments are supported through classroom elements that can be accessed at any time. Your participation in threaded discussion boards is required, graded, and paramount to your success in this course. Please note that any scheduled synchronous meetings are optional and are always recorded. While your attendance is highly encouraged, it is not required and you will not be graded on your attendance or participation in sync sessions. However, the material presented in the sync sessions is important to doing well in the course, so please review the recordings and/or PowerPoint slides if you don’t attend live.

Sync Sessions

Sync sessions are an important and useful part of the course. However, attending a sync session live is optional in the MSDS program, due to its asynchronous nature. Sync sessions are always recorded for those who cannot attend live. You will be given information a day or two ahead about how to connect to the live sync session in Blue Jeans. Note that I usually schedule sync sessions from 7:30 pm to 8:30 pm CDT (Central). The sync sessions are your best time to ask me questions and get an immediate answer - they are not "lectures", they are times for us to discuss the course work.

PowerPoint files from the sync sessions are posted along with the recordings, if they exist.

Valuable information about assignments and course requirements is often communicated during sync sessions, so if you are unable to attend live, please listen to the recordings.

Sync sessions for this course are scheduled for weeks 2, 5, and 8. I will be traveling in week 1, so will post a recorded information session instead of meeting live.

Student Support Services

AccessibleNU

This course is designed to be welcoming to, accessible to, and usable by everyone, including students who are English-language learners, have a variety of learning styles, have disabilities, or are new to online learning. Be sure to let me know immediately if you encounter a required element or resource in the course that is not accessible to you. Also, let me know of changes I can make to the course so that it is more welcoming to, accessible to, or usable by students who take this course in the future.

Northwestern University and AccessibleNU are committed to providing a supportive and challenging environment for all undergraduate, graduate, professional school, and professional studies students with disabilities who attend the University. Additionally, the University and AccessibleNU work to provide students with disabilities and other conditions requiring accommodation a learning and community environment that affords them full participation, equal access, and reasonable accommodation. The majority of accommodations, services, and auxiliary aids provided to eligible students are coordinated by AccessibleNU, which is part of the Dean of Students Office.

SPS Student Services

The Department of Student Services supports the academic and professional growth of SPS students. The Student Services team guides students through academic planning, policies, and administrative procedures, and promotes a supportive environment to foster student success. Students are encouraged to actively make use of the resources and staff available to assist them: Academic and Career Advisers, Counseling and Health Services, Student Affairs, Legal Services, Financial Aid and Student Accounts, among other services.

For a comprehensive overview of course and program processes and policies and helpful student resources, please refer to your SPS Student Handbook.

Academic Support Services

Northwestern University Library

As one of the leading private research libraries in the United States, Northwestern University Library serves the educational and information needs of its students and faculty as well as scholars around the world. Visit the Library About page for more information or contact Distance Learning Librarian Tracy Coyne at 312-503-6617 or tracy-coyne@northwestern.edu.

Program-Specific Library Guides

- Data Science

The Writing Place

The Writing Place is Northwestern's center for peer writing consultations. Consultations are free and available to anyone in the Northwestern community: undergraduates, graduate students, faculty, or staff. To book an appointment, go to The Writing Place website.

Academic Integrity at Northwestern

Students are required to comply with University regulations regarding academic integrity. If you are in doubt about what constitutes academic dishonesty, speak with your instructor or graduate coordinator before the assignment is due and/or examine the University website. Academic dishonesty includes, but is not limited to, cheating on an exam, obtaining an unfair advantage, and plagiarism (e.g., using material from readings without citing or copying another student's paper). Failure to maintain academic integrity will result in a grade sanction, possibly as severe as failing and being required to retake the course, and could lead to a suspension or expulsion from the program. Further penalties may apply. For more information, visit The Office of the Provost's Academic Integrity page.

Some assignments in SPS courses may be required to be submitted through Turnitin, a plagiarism detection and education tool. You can find an explanation of the tool here.

Examity Proctoring Tool

This course requires a proctored exam. Your identity will be verified prior to the exam, and your activity during the exam monitored by a remote proctor.

SPS partners with Examity for proctored exams. Students are responsible for proctoring, late scheduling, cancellation, and no-show fees.

- Two-hour exam: $23.00

Other possible fees:

- Scheduling within 24 hours of exam: $5.00 per hour.
- Cancellations or schedule changes within 24 hours of exam: $5.00 per exam.
- No-shows: Full payment of all proctoring fees.

To avoid additional fees, please schedule your exam appointment more than 24 hours prior to the exam.

Visit our course's Examity page for detailed instructions on how to set up an account and access the exam. Examity's privacy statement is also available.

Course Technology

This course will involve a number of different types of interactions. These interactions will take place primarily through the Canvas system. Please take the time to navigate through the course and become familiar with the course syllabus, structure, and content and review the list of resources below.

Canvas

The Canvas Student Center includes information on communicating in Canvas, navigating a Canvas course, grades, additional help, and more. The Canvas at Northwestern website provides information of getting to know Canvas at Northwestern and getting Canvas support. The Canvas Student Guide provides tutorials on all the features of Canvas. For additional Canvas help and support, you can always click the Help icon in the lower left corner to begin a live chat with Canvas support or contact the Canvas Support Hotline. The Canvas Accessibility Statement and Canvas Privacy Policy are also available.

BlueJeans

We will use BlueJeans for optional synchronous meetings. The Northwestern IT YouTube channel on Blue Jeans Video conferencing and the Canvas Learning Center BlueJeans page provide additional guidance for using BlueJeans.
The Blue Jeans Privacy Policy and the Accessibility Features on BlueJeans are also available.

Please note that any scheduled synchronous meetings are optional. While your attendance is highly encouraged, it is not required and you will not be graded on your attendance or participation. These synchronous sessions will be recorded, so you will be able to review the session afterwards.

Minimum Required Technical Skills

Students in an online program should be able to do the following:

- Communicate via email and Canvas discussion forums.
- Use web browsers and navigate the World Wide Web.
- Use the learning management system Canvas.
- Use integrated Canvas tools (e.g., BlueJeans, YellowDig, ARC, Course Reserves).
- Use applications to create documents and presentations (e.g., Microsoft Word, PowerPoint).
- Use applications to share files (e.g., Box, Google Drive).

Systems Requirements for Distance Learning

Students and faculty enrolled in SPS online master's degree programs should have access to a computer with the Minimum System Requirements.

Technical Help and Support

The SPS Help Desk is available for Faculty, Students and Staff to support their daily IT needs. For additional technical support, contact the Northwestern IT Support Center.

Course Schedule

Getting Started

- Welcome page
- Syllabus and Course Calendar
- About the Instructor
- Course Q & A
- Course FAQ
- Course Rubrics
- About Your TA
- University Resources

Module 1: Data Science and Data Scientists in Organizations

Objectives

After this session, you will be able to:

- Define data science, and describe what data scientists do.
- Compare how analytics are used in different organizational verticals.
- Assess the value of data science in business competition.

In addition the following topics may be discussed in future modules and/or appear on the final exam

- Review the methods and types of analyses used in the bobblehead promotion example.
- List and compare several data visualization tools used to present marketing and advertising findings.
- Describe a basic training and test regime for model evaluation.
- Define the DELTA model and enterprise level management.

Required Readings

Textbook Readings:

(Just for the first week, the textbook readings are also available in Course Reserves, in case you don't have the books yet.)

- *Competing on Analytics*, Chapters 1–3 (pgs. 3-56).
- *Modeling Techniques in Predictive Analytics*, Chapters 1-2 (pgs.1-12, and 15-25).
- *Data Science for Business*, Chapter 1 (pgs. 1-17).

Required Media

- Watch *Moneyball* (found in Course Reserves/Library Media).
- Video: What is data science?
- Video: Data science and analytically competitive organizations.
- Audio: Interview with Tom Miller.
- Infographic: Overview of Data Science.

Activities/Assessments

- Watch *Moneyball* (stream from library & post comment in graded discussion forum online).
- Discussion board participation.
- Introduce yourself in the Introductions forum.
- A recorded sync session of the term.

Module 2: Analytic Methods for Business Processes

Objectives

After this session, you will be able to:

- List and define the stages toward becoming an analytically competitive organization.
- Compare and contrast internal and external business processes and related analytic techniques.
- Distinguish between refereed and non-refereed sources for researching a paper topic.
- Use APA style for writing a research paper, and know how to find and use a style guide.
- List common data science methods.
- Define an application in data science for the paper.

In addition the following topics may be discussed in future modules and/or appear on the final exam

- Define conjoint analysis and explain its use.
- Differentiate ways to measure and analyze preference and choice.
- Describe data science methods for operations research, including queueing theory and mathematical programming.
- Review data science techniques for exploring and presenting findings on preference and choice, economic data, and operations research.

Required Readings

Textbook reading

- *Competing on Analytics*, Chapters 4, 5 (pgs 57-103).
- *Modeling Techniques in Predictive Analytics*, Chapters 3, 5, 6 (pgs. 29-33, 53-61, 67-76).

Course Reserves Reading

- Siegel, *Predictive Analytics*: Introduction; Chapter 1; Insert: 182 Examples of Predictive Analytics; Appendix B. Read in preparation for the Applications in Predictive Analytics assignment.

Optional Reading

- *Modeling Techniques in Predictive Analytics*, Chapter 9 (“Sports Analytics”; pgs. 149-170).

Required Media

- Video: Data Science Methods.
- Read through this online resource: Purdue OWL: APA Formatting & Style Guide: (to prep for Quiz 1 and the project).
- Audio: Interview with Thomas Davenport.

Activities/Assessments

- Start working on Applications and Methods in Data Science assignment (due end of third week).
- Quiz 1 to prep for Applications and Methods in Data Science paper (due end of second week).
- Discussion board participation.
- Sync session on upcoming assignments and Q&A.

Module 3: Analytic Leadership & Multivariate Models

Objectives

After this session, you will be able to:

- Assess the analytic capabilities of an organization.
- Describe how to evaluate analytic initiatives.
- Evaluate how organizations navigate the stages of becoming an analytic competitor.
- Compare and describe the roles played by analytic executives, analytic professionals, and analytic amateurs.

In addition the following topics may be discussed in future modules and/or appear on the final exam

- Define: text analytics and related methods; bag of words and natural language processing (NLP); principal components analysis; sentiment analysis and related methods and measures; iteration.
- Explain the role of the data scientist in enterprise level management.

Required Readings

Textbook Readings

- *Competing on Analytics*, Chapters 6–7 (157-216).
- *Modeling Techniques in Predictive Analytics*, Chapters 7- 8, 12 (pgs 93-100, 113-128, 231-235).

Required Media

- Video: Roles and Current Prospects in Data Science.

Activities/Assessments

- Discussion board participation.
- Applications and Methods in Data Science paper due.

Module 4: Data Engineering Essentials

Objectives

After this session, you will be able to:

- Evaluate the CRISP-DM cycle.
- Compare and contrast structured and unstructured data.
- Identify types of data and data sources for analytic applications: databases, weblogs, social media, enterprise, and public databases.
- Explain data preprocessing tasks: cleaning, integration, reduction, transformation, and discretization.
- Identify and distinguish between types of data storage and retrieval: text, binary, CSV, JSON, and XML.
- Compare and contrast common programming languages and technologies used for data munging and wrangling.

Required Readings

Textbook Readings

- *Business Problems and Data Sciences Solutions*, pgs. 19-42.

Course Reserves

- *Introduction to Data Mining*, Chapter 2, “Data”. Tan, P., Steinbach, M., & Kumar, V. (2005). Addison-Wesley, ISBN 9780321321367. Chapter 2 (pages 19-36).
- *Breaking Data Science Open*, 1st Edition, Chapter 7, “The Open Data Science Landscape.” Doig, C., Chambers, M., Stokes-Rees, I. (2017), O'Reilly Media, ISBN: 9781491990735. Chapter 7 (pgs. 33-43).

Required Reading available via link.

- Visit CRISP-DM, read only pgs 10-12.

Optional Resources

Available in Course Reserves:

Data Mining: Concepts and Techniques, Third Edition (2011), J. Han, M. Kamber & J. Pei, ISBN: 978-0-12-381479-1, Chapter 1, “Advanced Pattern Mining” (pgs. 279-325).

Required Media

- Infographic overview of data science and MSDS program.

Activities/Assessments

- Quiz 2 on Data Engineering.
- Discussion Board participation.
- Begin working on the Data Processing/Machine Learning Assignment - due week 6.
- Teams assigned for Survey Design & Implementation Project.

Module 5: Database Architecture, Data Warehousing, and the Web

Objectives

After this session, you will be able to:

- Delineate the role of data management in data architecture and warehousing.
- Outline the client–server architecture for database management systems (DBMS).
- Articulate and define the architecture and basic concepts of a data warehouse.
- Evaluate data warehouse design and usage.
- Compare and contrast SQL and NoSQL technologies for data storage and retrieval.
- Distinguish between data mining and data warehousing.
- Assess and debate the challenges of Big Data technologies.

In addition the following topic may be discussed in future modules and/or appear on the final exam

- List and define the three (or four) V's of Big Data.

Required Readings

Textbook reading

- *Competing on Analytics*, Chapter 8: "The Architecture of Analytics and Big Data," (pgs 217-247; 30 pages).

Course Reserves Reading

- *Data Mining: Concepts and Techniques, Third Edition* (2011), J. Han, M. Kamber & J. Pei, ISBN: 978-0-12-381479-1, Chapter 4 (124-135, 11 pages).
- *DataBase Systems: A Practical Approach to Design, Implementation and Management, 5th Edition* (2010). T. Connolly, and C. Begg, New York, NY: Addison Wesley, ISBN: 978-0321523068, Chapter 3: "Database Architectures and the Web" (page 57-97; 40 pages).
- *Database Systems: Design, Implementation, & Management, 12th Edition* (2016) , C. Coronel, and S. Morris, Cengage Learning. ISBN: 9781305627482, Chapter 14: "Big Data Analytics and NoSQL" (648-677; 29 pages)

Optional Resources

Available in Course Reserves:

- Data Mining: Concepts and Techniques, Third Edition (2011), J. Han, M. Kamber & J. Pei, ISBN: 978-0-12-381479-1, Remainder of chapter 4 (pgs 136-185; 60 pages on specifics of OLAP and Data Cube)

Available via link:

- Manyinka, J., Chui, M., Brown, B., Dobbs, R., Roxburgh, C. Hung Byers, A. (May 2011). Big Data: The Next Frontier for Innovation, Competition, and Productivity. Report, McKinsey Global Institute.

Required Media

Video: Big Data.

Activities/Assessments

- Quiz 3 on Data Warehousing.
- Discussion Board participation.
- Continue working on assignment due end of Week 6.
- Live sync session.

Module 6: Machine Learning and Artificial Intelligence

Objectives

After this session, you will be able to:

- Articulate the basic principles of data mining, machine learning (ML), and artificial intelligence (AI).
- Compare and contrast supervised and unsupervised machine learning.
- Define an algorithm.
- List and describe machine learning algorithms and methods: classification, regression, association, clustering.
- Define market basket analysis, and describe its applications and frequent patterns.
- Discuss various evaluation metrics: confusion matrix, precision, recall, and ROC metrics.
- Compare and contrast specificity and sensitivity.
- Discuss model evaluation and selection.
- List and discuss techniques for improving classification accuracy.
- Collaborate with your team to formulate your research question and population of interest.

Required Reading

Textbook Reading

- Data Science for Business, Chapters 3 (pgs 43-66), 4 (pgs 81-94), 6 (pgs 141-156), 7 (pgs. 187-193), & 12 (pgs 291-313) (75 pages).
- Modeling Techniques, Chapter 4, “Market Basket Analysis” (12 pages).

Course Reserves Reading

- Appendix B: Marketing Data Sources, from T. Miller, (2015). *Marketing data science*. Old Tappan, NJ: Pearson Education, Inc. (pgs 291-293; 3 pages)
- Appendix B: Marketing Data Sources, from T. Miller, (2015). *Marketing data science*. Old Tappan, NJ: Pearson Education, Inc. (pgs 305-337)
- "Ubiquitous Artificial Intelligence." N. Nilsson (2009), Cambridge University Press, ISBN: 9780521122931, Chapter 33 (pgs 615-620; 5 pages). A free online web version of this book is available from the author homepage at: The Quest for Artificial Intelligence.

Optional Additional Sources

None.

Required Media

None.

Activities/Assessments

- Machine Learning assignment due.
- First team checkpoint due for Survey Research Project.
- Discussion Board participation.

Module 7: Samples and Sampling Problems

Objectives

After this session, you will be able to:

- Define essential sampling terms and concepts: population, sample, sampling frame, simple random sample, population parameter.
- Identify key elements and symbols used in sample size formulas.
- Compute sample size for commonly used probability sampling methods.
- Compare and contrast simple random sample, stratified random sample, cluster sample, and quota sample, and state the appropriate uses and drawbacks of each.
- Define and contrast probability and nonprobability sampling, and identify problems arising in nonprobability sampling.
- Identify and discuss survey and sampling issues with underrepresented, rare, or stigmatized populations.
- Evaluate and identify sampling problems that result in faulty election predictions.
- Develop and create a survey questionnaire.
- Refine your sampling frame and sampling methods.
- Create a rough draft of the final team paper.

Required Reading

Textbook Reading

None.

Course Reserves Reading

“Sampling,” Chapter 6 (pgs. 62-81) from Watt, J. & van den Berg, S. A. (2002) Research Methods for Communication Science.

Meyer, I. H. & Wilson, P. A. (2009). Sampling lesbian, gay, and bisexual populations. *Journal of Counseling Psychology*, 56(1), 23-31.

Required Reading available via link:

Siegel, E. (2016, November 9). The science of error: How polling botched the 2016 election.

Required Media

Video: Sampling

Activities/Assessments

- Discussion Board participation.
- Quiz 4
- Second team checkpoint due.

Module 8: Creating and Using a Questionnaire

Objectives

After this session, you will be able to:

- Consider the processes and problems of constructing and administering surveys.
- Identify and mitigate sources of error and bias writing survey questions, and administering surveys.
- Evaluate the benefits and limitations of data collection methodologies.
- Assess the applications and advantages of a multimethod approach.
- Continue to develop and create a survey questionnaire.
- Continue to refine your sampling frame and sampling methods.
- Create a rough draft of the final team paper.
- Address sources of bias and error in the team project.

Required Readings

Textbook reading

None.

Course Reserves reading

Survey Methodology, Chapter 2: “Inference and Error in Surveys” (pg 39-64) and Chapter 7: “Questions and Answers in Surveys” (pgs. 217-255).

Couper, M. P. (2011). The Future of Modes of Data Collection. *Public Opinion Quarterly* 75(5):889-908.

Required Media

None.

Activities/Assessments

- Third Team Checkpoint.
- Discussion Board Participation.
- Live sync session.

Module 9: Data Quality, Missing Values, Error, and Bias

Objectives

After this module students will be able to:

- Evaluate sources of poor quality data and its possible effects.
- Identify potential sources of data in an organization.
- Assess the relationship between usability and availability of data.
- Identify and discuss causes of ethical problems in social research.
- Review ethical standards for social research in the US and at Northwestern University.
- Evaluate contemporary issues in data science and the internet in historical perspective.
- Design a professional-level research project.
- Collaborate with your team to propose appropriate sampling and survey methods.
- Perfect your final survey, sampling design, and final paper.

Required Reading

Textbook reading

None

Course Reserves reading

Groves et al., Survey Methodology: Chapter 6 (pgs 183-210): "Nonresponse in Sample Surveys," Chapter 9 (pgs 291-323): "Survey Interviewing," Chapter 10 (pgs 329-365): "Postcollection Processing of Survey Data."

Buchanan, E. A. & Zimmer, M. (2016).Internet research ethics, Sections 1-3, *The Stanford Encyclopedia of Philosophy* (E. N. Zalta, ed.; online link:

Berinato interview with Sandy Pentland.

Optional Resources

Course Reserves

- Shuster, E. (1997). Fifty years later: The significance of the Nuremberg Code. *New England Journal of Medicine* 337(20):1336-1440.
- Buchanan, E. A. & Zimmer, M. (2016). Internet research ethics, Section 4, *The Stanford Encyclopedia of Philosophy* (E. N. Zalta, ed.; online link:

Via Link

- Northwestern's Human Research Policies, Guidance & Oversight page:

Required Media

Available in Library Media/Course Reserves

Boudrias, M. & McDougall, I. (Producers) (2000) *Nuremberg*. [Docudrama]

Activities/Assessments

- Final Team Survey Design & Implementation Project Due.
- Discussion Board participation.

Module 10: Course Wrap-Up

Objectives

- List and evaluate a variety of methods used in data science.
- Articulate the value of data science in organizations.
- Discuss basic principles of data engineering, including database architecture and data warehousing.
- Define and describe Big Data, machine learning, and artificial intelligence.
- Articulate means of collecting, preparing, storing, and analyzing data.

Required Reading

None.

Required Media

None.

Activities/Assessments

- Proctored final exam.
- Discussion board participation.