Missouri University of Science & Technology
Department of Business and Information Technology

IST 3420 - Introduction to Data Science and Management
Course Syllabus - Fall 2017

MWF, 11:00 am – 11:50 am

Classroom: CS 204

Canvas home: https://mst.instructure.com/courses/20096

Department Mission:

"Capitalizing on the strong technological emphasis of Missouri S&T, the Department of Business and Information Technology educates individuals for careers in modern business organizations. The Department emphasizes management through technology, with particular focus on information systems and their application in a fast-changing, global, and competitive environment, to serve the economic interests of industry and the evolving needs of society."

INSTRUCTOR

Langtao Chen, Ph.D., Assistant Professor — Email: chenla@mst.edu

Department of Business & Information Technology — Phone: (573) 341-4418

Missouri University of Science & Technology — Office: Fulton Hall 106B

Office Hours: Fridays 12:00 pm – 2:00 pm or by appointment (suggest a date and time via email)

COURSE DESCRIPTION

This course introduces students to increasing business success through analyses of large-scale data collections. Topics include data import/export, summary statistics, cross-tabulation, data transformations (subsetting, merging, sorting and aggregation), modeling methods, and visualization. Significant programming in R is expected.

Prerequisite: IS&T 1552 or Comp Sci 1510.

LEARNING MANAGEMENT SYSTEM

Access Canvas (http://canvas.mst.edu/) for course materials, grades, course schedule, submission deadlines etc.

LEARNING OBJECTIVES

Course Objectives	**Program Learning Objectives**				
	Oral Communication	**Written Communication**	**Critical Thinking**	**Information Technology**	**Teamwork**
To obtain an overview of data science and data management	X	X		X	
To learn R programming language for data management and data analysis			X	X	
To obtain a set of basic methods and skills used to collect and manipulate data			X	X	
To develop an understanding of descriptive and predictive analytics			X	X	
Be able to build and evaluate time series, linear and nonlinear regression models for prediction			X	X	
Be able to build and evaluate simple classification models			X	X	
Be able to apply various techniques and skills to solve real business cases	X	X	X	X	X

GRADING

Course grading is composed of evaluations of the following components:

Grading Component	***Weight***
Participation, Exercises, and Unannounced Quizzes	15%
Homework	30%
Group Project	15%
Exam 1	20%
Exam 2	20%
Total	100%

The weighted percentage students have earned for all components will be rounded to the nearest hundredth. For example, 79.49% would be 79%, and 89.51% would be 90%. Then the final letter grade will be assigned as follows:

Grade	*Percentage*
A	>= 90%
B	< 90%
C	< 80%
D	< 70%
F	< 60%

TOOLS

1. **R**

   R is a free software environment for statistical computing and graphics.

   Obtain R from https://www.r-project.org/

2. **RStudio**

   RStudio is a powerful integrated development environment (IDE) for R.

   Get RStudio Desktop (open source edition) from
   https://www.rstudio.com/products/rstudio/download/

READINGS

The instructor will provide reading materials (refer to the course website for the detail). Those reading materials provide a basis for understanding course content and participating in class discussion. Thus, those materials must be read prior the class. Some in-classes quizzes based on reading materials will also be given prior lecturing.

TENTATIVE COURSE OUTLINE

The following is a general plan for the course; Deviations may be necessary.

1. **Introduction to Data Science**
   - 1.1. What is Data Science?
   - 1.2. Big Data
   - 1.3. Work as a Data Scientist
   - 1.4. Ethical Conduct in Data Science
2. **Introduction to R Programming**
   - 2.1. What is R?
   - 2.2. Basic Operations in R
   - 2.3. R Data Structures, Functions, and Control Structures
   - 2.4. R Programming Style and Debug
   - 2.5. Dynamic Reporting – R Markdown
3. **Data Basics**
   - 3.1. Data, Dataset
   - 3.2. Scales of Measurement
   - 3.3. Methods of Collecting Data (CSV, RDBMS, HTML, XML, JSON, APIs)
4. **Cleansing and Manipulating Data**
   - 4.1. Manipulate Strings
   - 4.2. Manipulate Datasets
5. **Data Summarization and Visualization**
   - 5.1. Tabular Methods
   - 5.2. Basic Graphical Methods
   - 5.3. Graphical Parameters
   - 5.4. Visualizing Spatial Data
6. **Data Exploration**
   - 6.1. Missing Data
   - 6.2. Outliers
   - 6.3. Correlation Analysis
   - 6.4. Hypothesis Testing
7. **Regression Analysis**

7.1. Simple Linear Regression

7.2. Multiple Linear Regression

7.3. Logistic Regression

8. **Predictive Analytics (Supervised Machine Learning)**

8.1. Predictive Performance Evaluation

8.2. Prediction and Classification Methods (Supervised)

9. **Unsupervised Machine Learning**

9.1. Cluster Analysis

9.2. Association Rules

10. **Advanced Topics**

10.1. Text Analytics

10.2. Analyzing Big Data (if time allows)

TENTATIVE SCHEDULE (Deviations may be necessary according to class progress)

Week	Topic	Exam/Group Project
W1: Aug 21, 23, 25	Course Introduction; 1. Intro to Data Science; 2. Intro to R Programming	
W2: Aug 28, 30; Sep 1	2. Intro to R Programming	
W3: Sep 4, 6, 8	No Class on Sep 4 (Labor Day) 3. Data Basics	Project M1: Group Proposal Due Sep 8 11:59 PM
W4: Sep 11, 13, 15	3. Data Basics	
W5: Sep 18, 20, 22	4. Cleansing and Manipulating Data	
W6: Sep 25, 27, 29	4. Cleansing and Manipulating Data	Exam 1 on Sep 29
W7: Oct 2, 4, 6	5. Data Summarization and Visualization	Project M2: Data Analysis I Due Oct 6 11:59 PM
W8: Oct 9, 11, 13	6. Data Exploration	
W9: Oct 16, 18, 20	6. Data Exploration	
W10: Oct 23, 25, 27	7. Regression Analysis	
W11: Oct 30, Nov 1, 3	8. Predictive Analytics	Exam 2 on Nov 3
W12: Nov 6, 8, 10	8. Predictive Analytics	Nov 10: Last day to drop
W13: Nov 13, 15, 17	9. Unsupervised Machine Learning	
W14: Nov 20, 22, 24	No Class (Thanksgiving Vacation)	
W15: Nov 27, 29, Dec 1	10. Advanced Topics	Project M3: Data Analysis II Due Dec 1 11:59 PM
W16: Dec 4, 6, 8	Course Review; Student Presentation etc.	Project M4: Project Presentation
W17: Dec 11, 13, 15	No Class	Project M5: Final Report Due Dec 14 11:59 PM

GROUP PROJECTS

1. Project Purpose

The purpose of the group project is to encourage students to apply and extend the techniques and methods they learn in the class to a real dataset and extract meaningful insights from the data. Students are encouraged to immerse deeply in the dataset and apply various analytical techniques to analyze the dataset.

It's important to note that the class cannot cover everything that you may need for your project. Students are expected to search online and learn new methods.

2. Datasets

Small teams (3-4 students) shall specify and analyze a small data science project by using data analytics tools.

Students should come up with their own datasets. If you have your own datasets, that is great. If you don't have datasets at hand, you can either collect your own datasets or find some existing datasets for your project from the Internet such as the following websites:

- Kaggle Datasets
  - https://www.kaggle.com/datasets
- Harvard Dataverse
  - https://dataverse.harvard.edu/dataverse/harvard

The dataset that is appropriate for the course project must satisfy the following criteria:

- The dataset must be in raw format. It has not been cleansed by other people;
- The dataset must have potential to tell an interesting story;
- There must be relatively large number of variables (>=20) that could be extracted from the dataset;
- The dataset should have a large number of observations (>=1000).

Each group must send their dataset to the instructor for approval. It is recommended that each group discuss their dataset with the instructor in the office hours.

3. Project Milestones

- ***Milestone 1: Project Proposal***
  Discuss with your classmates and form a project team with 3 or 4 members in total. Submit your project proposal. See the website for the detailed requirement on project proposal. If you need to discuss your project proposal with the instructor, please book a meeting with the instructor.
- ***Milestone 2: Data Analysis I***

See the website.

- ***Milestone 3: Data Analysis II***
  See the website.
- ***Milestone 4: Project Presentation***
  See project presentation guideline below.
- ***Milestone 5: Submit Project Final Report***
  See the website.

4. Project Presentation Guideline

Project presentations will take place at the end of the semester. Each presentation will last about 10-15 minutes (depending on the size of the class).
Note:

- This is a very short time, so be prepared to be concise.
- This is not the time to air complaints about the class, your classmates, or downplay the strengths of your work. This is the time to show the fruit of your semester's labor.

Your creative efforts should be realized within the following format:

- ***Teamwork*** (1 minute)
  Introduce your team members. Briefly describe their contributions.
- ***Background*** (1-2 minutes)
  Briefly explain the context and setting.
- ***Data Collection*** (1-2 minutes)
  Briefly explain the context and your data collection procedure or methods.
- ***Data Analyses and Results*** (5-8 minutes)
  Briefly explain your data analysis methods and results. If you have many things, try to present the important methods and interesting findings.
- ***Conclusion*** (1 minute)
  Summarize your project: What insights you get from the data analysis? What limitations does the project have? Can the project be extended? What are the next phases of development for this project?
- ***Open Floor*** (1 minute)
  Finally, the presenting team will respond to questions from the class. Your instructor will moderate the time remaining.

5. Project Grading

- Your milestone submissions will be graded through the semester, with temporal grades at the whole group level assigned to all group members.
- At the end of the semester, you will evaluate your group members on contribution to the group project.
- Each student's final grade for the group project will be assigned according to both group

performance (the cumulative points obtained from all milestones) and the level of his/her contribution to the group.

POLICIES

1. Attendance and Participation

Participation and in-class activity points cannot be made up if the student is absent.

The instructor will check attendance during the semester. Four missing classes will be ignored in calculating your final grade at the end of the semester. Because students are given a leeway of four missing classes, doctor's notes and other excuses are not accepted for absences. Missing one more additional class will lower your final grade by 1%. For example, if you miss 6 classes in the semester for whatever reason, your final grade will be lowered by 2%.

Attendance is a prerequisite, not a substitute for class participation. Students should be fully prepared for each class. The instructor will ask students questions in the class. Your participation points obtained will reflect the quality of your answers and your motivation for class participation.

In-class activities such as assignments, exercises, and unannounced quizzes should be submitted before the class session ends if not specified otherwise. Late submissions will not be accepted.

2. Readings

Reading materials will be assigned for corresponding sessions. Readings provide the basis for lecture and class discussions. Thus, corresponding materials **must** be read prior each class.

Check the course website for reading assignments.

3. Exams

Exams are closed book and note. No make-up exam allowed unless you ask in advance (and the instructor agree) that a significant life-event prevents you from attending the exam.

4. Assignment Submission Policy

Instruction for homework and project assignments will be posted on the course's website. Each assignment must be submitted by the due time. The assignment must be submitted through the course website unless the instructor requests another channel. Students who fail to submit an assignment before the deadline will be given additional 3 days to submit the assignment to the instructor for legitimate reasons (e.g., documented illness). Late submissions will receive a penalty of 20% of the points assigned to that specific assignment "per day". No credit will be given for assignments submitted more than 100 hours after their original due time. It is your responsibility to make sure that you properly submit the correct files.

5. Title IX

Missouri University of Science and Technology is committed to the safety and well-being of all members of its community. US Federal Law Title IX states that no member of the university community shall, on the basis of sex, be excluded from participation in, or be denied benefits of, or be subjected to discrimination under any education program or activity. Furthermore, in accordance with Title IX guidelines from the US Office of Civil Rights, Missouri S&T requires that all faculty and staff members report, to the Missouri S&T Title IX Coordinator, any notice of sexual harassment, abuse, and/or violence (including personal relational abuse, relational/domestic violence, and stalking) disclosed through communication including but not limited to direct conversation, email, social media, classroom papers and homework exercises.

Missouri S&T's Title IX Coordinator is Vice Chancellor Shenethia Manuel. Contact her directly (manuels@mst.edu; (573) 341-4920; 113 Centennial Hall) to report Title IX violations. To learn more about Title IX resources and reporting options (confidential and non-confidential) available to Missouri S&T students, staff, and faculty, please visit http://titleix.mst.edu.

6. Student Honor Code and Academic Integrity:

Please take a few minutes to stress the importance of academic integrity in class. Discuss why it should matter to the student, why it matters to you and your discipline, why it matters to Missouri S&T, and why it matters to future employers. Include a statement on your syllabus about the Honor Code developed and endorsed by the Missouri S&T Student Council: the Honor Code can be found at this link: http://stuco.mst.edu/honor-code/. Encourage students to read and reflect upon the Honor code and its emphasis on HONESTY and RESPECT.

Page 30 of the Student Academic Regulations handbook describes the student standard of conduct relative to the University of Missouri System's Collected Rules and Regulations section 200.010, and offers descriptions of academic dishonesty including cheating, plagiarism or sabotage (http://registrar.mst.edu/academicregs/index.html). Additional guidance for faculty, including the University's Academic Dishonesty Procedures, is available on-line at http://ugs.mst.edu. Other informational resources for students regarding ethics and integrity can be found online at http://ugs.mst.edu/academicintegrity/studentresources-ai.

7. S&Tconnect: https://canvas.mst.edu/ (S&Tconnect icon on left toolbar)

S&Tconnect provides an enhanced system that allows students to request appointments with their instructors and advisors via the S&Tconnect calendar, which syncs with the faculty or staff member's Outlook Exchange calendar. S&Tconnect will also facilitate better communication overall to help build student academic success and increase student retention. S&Tconnect Early Alert has replaced the Academic Alert system used by Missouri S&T. If training is needed, please contact Rachel Morris at rachelm@mst.edu or 341-7600.

8. Classroom Egress Maps

Faculty should explain where the classroom emergency exits are located. Please include a statement in your course syllabus asking the students to familiarize themselves with the classroom egress maps posted on-line at: http://designconstruction.mst.edu/floorplan/.

9. Accessibility and Accommodations

It is the university's goal that learning experiences be as accessible as possible. If you anticipate or experience physical or academic barriers based on disability, please contact Disability Support Services at (573) 341-6655, dss@mst.edu, or visit http://dss.mst.edu/ for information, or go to mineraccess.mst.edu to initiate the accommodation process.

**Please be aware that any accessible tables and chairs in this room should remain available for students who find that standard classroom seating is not usable.*

10. LEAD Learning Assistance http://lead.mst.edu

The Learning Enhancement Across Disciplines Program (LEAD) sponsors free learning assistance in a wide range of courses for students who wish to increase their understanding, improve their skills, and validate their mastery of concepts and content in order to achieve their full potential. LEAD assistance starts no later than the third week of classes. Check out the online schedule at http://lead.mst.edu/assist, using zoom buttons to enlarge the view. Look to see what courses you are taking have collaborative LEAD learning centers (bottom half of schedule) and/or Individualized LEAD tutoring (top half of the schedule). For more information, contact the LEAD office at 341-7276 or email lead@mst.edu.

11. The Burns & McDonnell Student Success Center

The Student Success Center is a centralized location designed for students to visit and feel comfortable about utilizing the campus resources available. The Student Success Center was developed as a campus wide initiative to foster a sense of responsibility and self-directedness to all S&T students by providing peer mentors, caring staff, and approachable faculty and administrators who are student centered and supportive of student success. Visit the B&MSSC at 198 Toomey Hall; 573-341-7596; success@mst.edu; facebook: www.facebook.com/SandTssc; web: http://studentsuccess.mst.edu/

If you have any questions about the information listed above, please contact the Office of Undergraduate Studies at 573-341-7276.