

INFO 3200: Data Mining and Visualization

<p><u>Term and Credits:</u></p> <p>Spring 2018</p> <p>4 Credit Hours</p> <p>10:00a Section 1-CRN: 2742 12:00p Section 2-CRN: 2666 4:00p Section 3-CRN: 3458</p>	<p><u>Time and Location:</u></p> <p>Tu/Th 10:00am -11:50am in DCB 300 12:00pm - 1:50pm in DCB 300 4:00pm - 5:50pm in DCB 110</p> <p>You will be required to review some material outside of class which will be delivered through Canvas. Make sure you have a good internet connection during class for access to Canvas and Top Hat.</p>
<p><u>Instructor:</u></p> <p>Name: Kellie Keeling</p> <p>Department: Business Information & Analytics</p> <p>Office Location: DCB 590</p> <p>Email: kkeeling@du.edu</p> <p>Office Phone: 303-871-2296 (forwards to my cell)</p> <p>Office Hours: Tu/Th 2:00pm-4:00pm</p> <p>GA Office Hours: Tu 10:00am-12:00pm (Elisabeth Bernatska) Th 8:00am-12:00pm</p>	<p><u>Communication Conduct:</u></p> <p>Feel free to refer to me as Dr. Keeling, Professor, or Kellie as you feel comfortable.</p> <p>Email is usually the best way to contact me. If I haven't responded in 36 hours, feel free to resend your message.</p> <p>I will send class level communications via Canvas announcements. I will typically initiate communication with individual students directly through your DU email or through Canvas email.</p> <p>My open office hour times are available on the home page in Canvas under "My Office Hours." To specifically make an appointment with me, click the link which goes to http://doodle.com/kkeeling which will allow you to request an appointment time. If there is an open time, you can also just stop by.</p>

COURSE DESCRIPTION:

This course explores the concepts of storytelling with data, prediction modeling, and presenting statistical results. It covers the concepts of visualization terminology along with all the steps of the modeling process: define goal, get data, explore & visualize data, pre-process data, partition the data series, apply modeling technique(s), evaluation and compare performance, implement the model, and communicate the results. The modeling techniques covered include Time Series Forecasting, Clustering, Principal Components Analysis,

Decision Trees, Naïve Bayes/K-Nearest Neighbor, Multiple and Logistic Regression, and Machine Learning Approaches. This course also covers the interpretation of real-time business data in terms of dashboards and scorecards.

PREREQUISITES/CO-REQUISITES:

Prerequisite: INFO 2020

Required Materials

Textbook: *Fundamentals of Predictive Analytics with JMP*, Second Edition, 2016, Klimberg and McCullough

Soft Copy Options:

Epub ISBN# 978-1-62960-801-3, Mobi ISBN# 978-1-62960-802-0, PDF ISBN# 978-1-62960-803-7

Software: **Top Hat Account - \$26/\$38** (instructions will be emailed to you)

Microsoft Office

JMP Pro 13 (Free)

Tableau 10.5 (Free)

R and R Studio (Free))

Supplemental Texts:

These are not required but are some of the visualization resources you will see referenced throughout the class.

Now You See It, Stephen Few, 2009

Storytelling with data, Cole N. Knaflic, 2015

Show Me the Numbers, Stephen Few, 2004

The Functional Art, Alberto Cairo, 2012

Information Dashboard Design, Stephen Few, 2006

Visualize This, Nathan Yau, 2011

LEARNING OUTCOMES:

After completing this course, students will be able to:

- Define the terminology/vocabulary and methods associated with Data Mining and Visualization in order to choose and perform appropriate analyses and also intelligently talk to other analysts
- Recognize the benefit of and use a Data Modeling framework for solving problems with data including:
 - Prepare data for use in models with data manipulation and visualization techniques
 - Discuss and implement modeling and assessment of competing prediction methods/models.
 - Document the process of performing the analysis using the concepts of reproducible research.
 - Communicate data and modeling results effectively, including the use of data visualizations (graphs and tables)

GRADING STRUCTURE, SCALE, AND POLICIES:

GRADING STRUCTURE:

Performance will be evaluated on the items below. For this class, all assignments assume you are trainees for Stats Dairy. Your training score is only a measure of your performance in this class and does not reflect my opinion of you as an individual or your worth as a person.

GRADING SCALE:

Stats Dairy regularly hires more trainees than it needs. By means of this course we determine where to place the graduates of the program:

90% - 100%	A	Trainees who receive an A are considered on the "fast track" and will start out as data mining analysts. Our studies show that most trainees who fall in this group reach an executive position within 10 years.
80% - 89%	B	Trainees who receive a B will start out as assistant data mining analysts. This does not mean that they cannot reach the executive level but it will be more difficult since they will not regularly be put into career-enhancing positions such as overseas consulting assignments.
70% - 79%	C	Trainees who receive a C will be put into staff positions for further development.
60% - 69%	D	Trainees who receive a D will be offered non-management positions.
00% - 59%	F	Trainees who receive an F will be separated from Stats Dairy.

A: 93-100%; A-: 90-92.9%; B+: 87-89.9%, B: 83-86.9%; B-: 80-82.9%; etc.

Assessment

Performance will be evaluated on the items:

Requirement	Amount
In-Class Group Tasks	10%
In-Class Top Hat Assessments	10%
Pre-Class Training Assignment Quizzes	12%
Exam 1	15%
Exam 2	20%
Course Project (3 Phases)	15%
5 Cases	18%
Total	100%

- **Module Exams:** If you are going to need to reschedule an exam for a legitimate conflict, you must receive permission from me BEFORE the exam in order to reschedule. Otherwise you will receive a zero on the exam. No make-up exams will be given. The exam will consist of two parts. The first part will be completed out of class "at home" Fri-Monday of the assigned week and the second part will be completed during class on Tuesday.
- **Pre-Class Training Assignment Quizzes:** There will be quizzes to be completed before each class that will cover the reading material and the video lectures. Late submissions will be accepted for 69% credit up to one week late.
- **Case Assignments:** There will be 5 individual case assignments.
- **Top Hat Assessment:** Top Hat will be used to assess preparation for class as well as group progress during class. Two Top Hat Assessments will be dropped. **There are no 'Make Up' Top Hat Assignments.** Note that participation counts for 75% and getting the correct answer counts for 25% of your score for each question.
- **In-class Group Tasks:** During class, mini-group discussions/problems will be completed to help supplement our course material. You may work individually or with a total of people in a group.
- **Individual Course Project:** Detailed instructions will be posted on Canvas.

Communication

If you are having difficulty with the course material, please see me at your earliest convenience. Do not wait until the first exam to see me about difficulties you are experiencing in comprehending the course material. Do not allow yourself to fall behind in covering the assigned material as this will most certainly result in a poor course grade. Keep up with your assignments and the readings in the text!

Class Schedule

TUESDAY		THURSDAY	
Mar 27	Exploring/Visualizing Your Data	Mar 29	Exploring/Visualizing Your Data/Presenting Statistical Results
Apr 3	Dashboards and Scorecards/Maps	Apr 5	Data Organizing/Principal Components Analysis
Apr 10	Time Series Forecasting	Apr 12	Time Series Forecasting
Apr 17	Clustering	Apr 19	Catch Up Day
Apr 24	Exam 1	Apr 26	Classification vs Prediction/Estimation Data Mining Models
May 1	Decision Trees	May 3	Naïve Bayes/K-Nearest Neighbor
May 8	Regression Models	May 10	Regression Models
May 15	Regression Models	May 17	Machine Learning Techniques
May 22	Machine Learning Techniques	May 24	Catch Up Day
May 29	Exam 2	May 31	Project Work Day
Jun 5	Project Presentations		

Class Structure of Topics



UNIVERSITY EXPECTATIONS, POLICIES, AND RESOURCES:

Students with Disabilities. Students who have disabilities or medical

conditions and who want to request accommodations should contact the Disability Services Program (DSP); 303.871.2372/ 2278; 1999 E. Evans Ave.; 4th floor of Ruffatto Hall. Information is also available online on the [DU Disability Services website](#); see Handbook for Students with Disabilities. Please note that academic accommodations cannot be applied retroactively, so it is important for you to register with DSP as soon as possible if you think you may need accommodations at some point while at Daniels College of Business.

DU Honor Code. All students are expected to abide by the University of Denver Honor Code. These expectations include the application of academic integrity and honesty in your class participation, assignments and assessments. The Honor Code can be viewed in its entirety on the [DU Student Conduct website](#).

All members of the University of Denver are expected to uphold the values of Integrity, Respect, and Responsibility. These values embody the standards of conduct for students, faculty, staff and administrators as members of the University community.

In order to foster an environment of ethical conduct in the University community, all community members are expected to take "constructive action," that is, any effort to discuss or report any behavior contrary to the Honor Code with a neutral party. Failure to do so constitutes a violation of the DU Honor Code. Specifically, plagiarism and cheating constitute academic misconduct and can result in both a grade penalty imposed by the instructor and disciplinary action including suspension or expulsion. As part of their responsibility to uphold the Honor Code, instructors reserve the right to have papers checked for plagiarism against a database of papers submitted previously at DU, a national database of papers, and the Internet.

Additional University Expectations. Please review all University Expectations on the [Daniels College of Business syllabus website](#).