Data Science | 2016-2017 Assessment Report

1. Please give a brief overview of the assessment data you collected this year.

    1. **MSDS program has seven program learning goals (PLG):**
    The MSDS students will be able to:
    PLG A: identify and assess the needs of an organization for a data science task
    PLG B: collect and manage data to devise solutions to data science tasks
    PLG C: select, apply, and evaluate models to devise solutions to data science tasks
    PLG D: interpret data science analysis outcomes
    PLG E: effectively communicate data science related information effectively in various formats to appropriate audiences
    PLG F: value and safeguard the ethical use of data in all aspects of their profession
    PLG G: transform findings from data resources into actionable business strategies

    2. **Assessment Committee (2016-2017)**
    Amy Kuether (UWEX)
    Gaurav Bansal (UWGB)
    Jeffrey Baggett (UW Lacrosse)
    Ethan Christensen (UW Superior)
    Robert Dollinger (UW Stevens Point)
    Erik Krohn (UW Oshkosh)
    Alex Smith (UW Eau Claire)

    3. **Assessment plan**

    MSDS Assessment committee decided to assess all the program learning goals for the year 2016-2017. the program learning goal (PLG C) for the year 2015-2016. The assessment was carried out in two courses: DS 715 and DS 730 in Spring 2016 using embedded class assessments. We assessed the following two traits of PLG C: (i) Students will be able to choose and apply tools and methodologies to solve data science tasks (DS 730), and (ii) Students will be able to assess the model used to solve data science tasks (DS 715).

    4. **Summary of the assessment:**

    We assessed the following seven traits:

Course Learning Objective Code (Traits)	Course Learning Objectives (Traits)	Course Assessed	Semester Assessed
A1	Students will be able to conduct a needs assessment.	DS700	Fall 2016

A2	Students will be able to frame tasks in the context of organizational goals.	DS700	Fall 2016
C7	Choose and apply tools and methodologies to solve data science tasks	DS700	Fall 2016
F14	Identify and analyze social, legal, and ethical issues associated with the recommendations provided.	DS700	Fall 2016
C6	Identify and classify relevant variables for data science tasks	DS745	Spring 2017
C7	Choose and apply tools and methodologies to solve data science tasks	DS745	Spring 2017
D9	Interpret data, extract meaningful information, and assess findings	DS745	Spring 2017
E12	Help non-technical professionals visualize, explore, and act on data science findings	DS745	Spring 2017

Performance of students:

Marks	DS700 Fall 2016 A1 Need Assessment	DS700 Fall 2016 A2 Frame Task	DS700 Fall 2016 C7 Tools	DS700 Fall 2016 F14 Ethics	DS745 Spring 2017 C6 Variables	DS745 Spring 2017 C7 Tools	DS745 Spring 2017 D9 Interpret	DS745 Spring 2017 E12 Act	DS760 Spring 2017 F14 Ethics
90% and above	41	50	38	27	11	11	14	10	10
80 to 90%	27	17	30	17	2	6	3	5	3
70 to 80%	5	6	5	11	4	0	0	3	1
60 to 70%	1	1	1	17	1	0	1	1	0
Less than 60	3	3	3	5	4	5	2	3	0
# of students	77	77	77	77	22	22	20	22	14

DS 700: There were 77 students who were assessed. Two students got incomplete and hence couldn't be assessed. Out of 77, 23 students got less than 60% in ethics. Also, out of 77, four students got less than 60% for need assessment, frame task and tool selection.

DS 745: Comments: There were 22 students who were assessed. Two students couldn't be assessed for the first learning goal. Out of 22, five students got less than 60% for "Choose and apply tools and methodologies to solve data science tasks".

DS 760 Comments: The instructor believes that the class did very well in satisfying the selected learning outcome. By the end of the course students have had a lot of practice analyzing

ethical issues in data science and it shows. The student who received less than 80% on the rubric item was not a strong writer, which meant that he needed to put a lot of effort into this assignment. (Instructor believes that he was capable of receiving a 40 or better out of 50 on this rubric item, but that he did not put in the effort required of him.) Note, also, that compared to the assessment done during the Fall (2016) semester, it appears that students are succeeding in this learning outcome at a higher rate.

5. Assessment Method

DS 700
Method of assessment: Case study
Data was collected by examining the performance of all the students enrolled on the course in their case study report. All students were required to submit a case study report, and it was an individual assignment. Previously it be used to be a group project (Spring 2016 and before), however it was felt that the students are dividing the work, and hence each student is learning only one component of the case study and not all. Hence this time we converted the case study into an individual assignment, and provided students additional help so that they can complete the entire case study individually.

DS 745
Method of assessment: three projects.

Data was collected by examining the performance of all the students enrolled in the course in three different projects: visualization, network analysis and text mining. All students were required to submit a report for each of the project. All the projects were individual assignments. We arrived at the assessment data by averaging the grades obtained by a student for each learning objective across the three projects. Refer to the Appendix B for the rubrics used for DS 745 assessment.

DS 760
Method of Assessment: Final Project

Examples: The final project is a paper with a very specific format. For the paper, the student selects a current topic related to ethics in data science. This topic is to be chosen from among all the topics presented by members of the class in video presentations (due earlier). The paper must specifically analyze an ethical issue by taking an ethical point of view that emphasizes equality, justice, and respect, applying existing policies and conventions (including any laws), contemplating what one or more ethical theories might conclude, and consulting at least one relevant professional code of ethics. The assignment requires that the student also consider at least one point of view that is contrary to her own with regard to the ethical issue under consideration. The student must make the case that while she understands the alternative point of view, there are strong reasons for favoring her own. The project therefore offers a very strong platform for showing mastery of Outcome #16: Students will be able to identify and analyze social, legal, and ethical issues in data science.

Results: Of the 14 students in the course who submitted a final project, all of the students enrolled except for one received at least a 40/50 (or 80%) on the rubric item that reads:

Show a nuanced understanding of ethical dimensions through a strong analysis which (1) incorporates ethical frameworks, code of ethics, and other systematic analysis techniques to lay out two alternative views; and (2) defends one of them.

2. How will you use what you've learned from the data that was collected?

As we are just getting the program up and going there are many gaps in measures and results. This is a work in progress. Most of the CLOs are being covered in multiple courses. Several of the gaps correspond to courses that are being piloted during 2016-17.

An important task for MSDS assessment committee for 2017-2018 will be to put a comprehensive assessment plan in place to start assessing most of the PLG, and also to take corrective action based on the results from 2016-2017 assessment data.

In preparing to report on our current assessment plan we've realized a few **shortcomings:**
   - o The learning outcomes are too granular and frequently overlapping. We will be work to simplify the learning outcomes and curriculum map in Summer 2017.
   - o The curriculum map needs to distinguish between introducing, continuing coverage, and assessment of learning outcomes. Not every one of the outcomes will be assessed in every class. We hope to add this to the simplified curriculum map in Summer 2017.
   - o Right now the assessment plan is essentially to ensure that every course includes assessment of at least some of its corresponding learning outcomes. So every course incudes assessment, but not every learning outcome is being assessed somewhere. We will be working to address this during the 2017-18 year.
   - o Each instructor is called upon to complete an assessment report for their course, but the template for these reports isn't sufficiently detailed so that the task and rubric are not always reported.

Appendix A

Master of Science in Data Science - Program Level Learning Goals & Course Mapping

Learning Goal	7005	7010	7015	7105	7130	7345	7405	7450	7605	7705	7805	7885
A. Students will be able to…identify and assess the needs of an organization for a data science task.												
1) Students will be able to conduct a needs assessment.	X											
2) Students will be able to frame tasks in the context of organizational goals.	X											
3) Students will be able to communicate data science options and limitations that could meet organizational needs.										X		
B. Students will be able to collect and manage data to devise solutions to data science tasks.												
4) Students will be able to collect, clean, and prepare data.			X	X		X						
5) Students will be able to evaluate data in terms of source, volume, frequency, and flow.			X	X	X	X				X		
C. Students will be able to…select, apply, and evaluate models to devise solutions to data science tasks.												
6) Students will be able to identify and classify relevant variables for data science tasks.			X	X	X	X	X		X	X		
7) Students will be able to choose and apply tools and methodologies to solve data science tasks.	X		X	X	X	X	X					
8) Students will be able to assess the model used to solve data science tasks.			X	X		X						
D. Students will be able to…interpret data science analysis outcomes.												
9) Students will be able to interpret data, extract meaningful information, and assess findings.			X	X	X	X	X		X	X		
10) Students will be able to evaluate the limitations of data science findings.			X	X		X				X		
E. Students will be able to…effectively communicate data science related information effectively in various formats to appropriate audiences.												
11) Students will be able to write, format, disseminate, and orally communicate technical materials.				X		X						X
12) Students will be able to help non-technical professionals visualize, explore, and act on data science findings.				X		X	X	X	X			
13) Students will be able to facilitate data-informed discussions through listening, questioning, and presenting.				X		X	X		X			
F. Students will be able to…value and safeguard the ethical use of data in all aspects of their profession.												
14) Students will be able to identify and analyze social, legal, and ethical issues in data science.	X											
15) Students will be able to interpret and apply a professional code of ethics relevant to the data science profession.								X	X			
16) Students will be able to interpret the activities and choices of others within an ethical framework and determine an appropriate action based on standards of professional conduct.								X				
G. Students will be able to… transform findings from data resources into actionable business strategies.												
17) Students will be able integrate data science capabilities into the formation of a situation analysis.											X	
18) Students will be able to explain how data assets can be used to develop competitive advantage.											X	
19) Students will be able to identify and appraise the leadership and management skills required to direct a team of data science professionals toward meeting organizational goals.											X	

Appendix B

Following rubrics were used to examine the four learning objectives for three different projects (A: Visualization Project, B: Network analysis project, C: Text Mining project).

A. Visualization Project Rubric

Criteria	Excellent	Strong	Competent	Needs Improvement	Needs Significant Revision
You have a well-defined purpose. The story portrayed in your visualization aligns with the purpose of your visualization. The story reflected in your visualization is novel and the shows interesting pattern(s).	10 points	8 points	6 points	4 points	2 points
Choice of dataset used is rich and multivariate. Your work is original and rigorous.	40 points	30 points	20 points	15 points	5 points
Choice of variables aligns well with the purpose of the visualization	10 points	8 points	6 points	4 points	2 points
Choice of visualization tools and methodology adheres well with design principles learned in this course.	10 points	8 points	6 points	4 points	2 points
Provide and explain three progressively improving iterations. Explain pros and cons for each iteration. Iterations are thoughtful and rigorous.	40 points	30 points	20 points	15 points	5 points
The visualization is self – explanatory, and has no puzzling (confusing) pieces. The visualization stimulates viewer engagement and attention.	10 points	8 points	6 points	4 points	2 points
Overall Score	A 108 or more	B 96 or more	C 72 or more	F 0 or more	

B. Networking Project Rubric

Interpret data, extract meaningful information, and assess findings.	Excellent	Strong	Competent	Needs Improvement	Needs Significant Revision
Choice of dataset used is rich and multivariate (with several node and edge attributes). Data and the subset selected are well explained.	20 points	16 points	12 points	8 points	4 points
Visualizations are easy to understand. Accompanying text highlights the key points observed in the visualizations.	40 points	30 points	20 points	15 points	5 points
Community structures are well explained.	10 points	8 points	6 points	4 points	2 points
The path coefficients and p value of the ERGM model are well explained.	10 points	8 points	6 points	4 points	2 points
ERGM Model goodness of fit is well explained using several different criteria.	10 points	8 points	6 points	4 points	2 points
Identify and classify relevant variables for data science tasks.	**Excellent**	**Strong**	**Competent**	**Needs Improvement**	**Needs Significant Revision**
There is well explained hypothesis for possible relationship between node / edge attributes and linkage possibilities.	10 points	8 points	6 points	4 points	2 points

Choose and apply tools and methodologies to solve data science tasks.	Excellent	Strong	Competent	Needs Improvement	Needs Significant Revision
Network visualizations are prepared, and network summary is presented.	10 points	8 points	6 points	4 points	2 points
Community structures are detected.	40 points	30 points	20 points	15 points	5 points
ERGM modeling is carried out.	10 points	8 points	6 points	4 points	2 points
Help non-technical professionals visualize, explore, and act on data science findings.	**Excellent**	**Strong**	**Competent**	**Needs Improvement**	**Needs Significant Revision**
The network visualizations and the accompanying explanations along with the report are detailed but still clear and concise. Follows the project guidelines carefully.	30 points	24 points	18 points	12 points	6 points
Overall Score	**A** 171 or more	**B** 152 or more	**C** 114 or more	**F** 0 or more	

C.   Text Mining Project Rubric

Dataset	Excellent	Strong	Competent	Needs Improvement	Not Present
Choice of dataset used is rich and multivariate (with several node and edge attributes). Data and the subset selected are well explained.	30 points	24 points	18 points	12 points	0 points
Dataset is explained well, using visual and descriptive methods.	10 points	8 points	6 points	4 points	0 points
Report: Meet basic requirements	**Excellent**	**Strong**	**Competent**	**Needs Improvement**	**Not Present**
Cover page.	8 points	6 points	4 points	2 points	0 points
Index.	8 points	6 points	4 points	2 points	0 points
References.	8 points	6 points	4 points	2 points	0 points
Report: Interpret data, extract meaningful information, and assess findings.	**Excellent**	**Strong**	**Competent**	**Needs Improvement**	**Needs Significant Revision**
Report highlights the key points observed in the analysis.	8 points	6 points	4 points	2 points	1 point
Report: Identify and classify relevant variables for data science tasks.	**Excellent**	**Strong**	**Competent**	**Needs Improvement**	**Needs Significant Revision**
Report identifies relevant variables needed for the text mining task.	8 points	6 points	4 points	2 points	1 point
Report: Choose and apply tools and methodologies to solve data science tasks.	**Excellent**	**Strong**	**Competent**	**Needs Improvement**	**Needs Significant Revision**
Appropriate Text Mining tools are applied. Report provides adequate explanation of how the analysis was carried out.	16 points	12 points	8 points	4 points	2 points
Report: Help non-technical professionals visualize, explore, and act on data science findings.	**Excellent**	**Strong**	**Competent**	**Needs Improvement**	**Needs Significant Revision**
The findings are explained with carefully crafted visualizations and clear (and detailed) explanations that are easy for any non-technical person to understand as well.	24 points	18 points	12 points	6 points	3 points
R Code	**Present and very well commented**	**Present and well commented**	**Present and commented**	**Present but not commented or poorly commented**	**Not Present**
Submission includes R code file and proper comments.	40 points	32 points	24 points	16 points	0 points
Overall Score	**A** 144 or more	**B** 128 or more	**C** 96 or more	**F** 0 or more	