ASEE 2022 ANNUAL CONFERENCE
Excellence Through Diversity
MINNEAPOLIS, MINNESOTA, JUNE 26TH-29TH, 2022
Paper ID #37094

Introducing Data Analytics into Mechanical Engineering Curriculum

Xiyuan Liu (Assistant Teaching Professor)

Xiyuan Liu is currently an assistant teaching professor in the Department of Mechanical and Aerospace Engineering in College of Engineering. She received B.S. in Electrical Engineering in China in 2009 and then completed her M.S. in Mechanical Engineering at Clemson University. She received her PhD degree in Mechanical Engineering at Michigan State University. Her PhD work mainly focused on developing biosensing, lab-on-a-chip systems for the emerging applications in clinical diagnosis, wearable sensing and mobile heath (mHeath) technology. In 2017, she joined Syracuse University as an assistant teaching professor for a joint position between the Department of Biomedical and Chemical Engineering and the Department of Mechanical and Aerospace Engineering. Since 2020, she becomes a full-time assistant teaching professor in the Department Mechanical and Aerospace Engineering. As an instructor, she teaches courses at different levels, from first-year undergraduate engineering programming course to graduate level technical elective courses. She particularly interests in improving engineering education through enhancing students learning experience, cultivating an active learning environment and promoting diversity, equity and inclusion (DEI).

© American Society for Engineering Education, 2022
Powered by www.slayte.com

Introducing Data Analytics into Mechanical Engineering Curriculum

Xiyuan Liu
Department of Mechanical & Aerospace Engineering, Syracuse University

Abstract

The advancements in information technology, computing power, data mining and artificial intelligence have enabled all the engineering disciplines to take the advantages of large datasets to model, classify, and make proper predictions for numerous engineering applications. To educate next-generation mechanical engineers in the new era of data science and artificial intelligence, engineering educators have been urged to integrate these new technological advancements into existing curriculum to adapt to the fast-changing needs from the future workplace. My effort primarily focuses on implementing an interdisciplinary approach to introduce the concepts and principles of data science to the undergraduate students of mechanical engineering. I re-designed the class of Statistics for Engineering as Data Analytics for Engineering, in which the students can practice new tools used in data analytics applications while they are still learning the basic statistical principles behind these techniques.

In this class, the students are exposed to the real-world examples of how data analytics has been applied in the field of mechanical engineering. The course content arrangement is based on the data analytics lifecycle: problem discovery – data understanding – data preparation – data visualization – model building – conclusion/decision making. Statistical concepts related to each stage are introduced to the students along with the corresponding programming basics in *R-studio*. Parallelly, the semester-long project is assigned to the student groups from the first day of lecture. Each group is required to select a real-world dataset and complete the data analysis using data cleaning, data preparation, data visualization, regressions, and several machine learning algorithms. To help the student better complete the project, I develop interactive activities at different development stages, including project proposal, proposal peer review, project interview, preliminary report, and final report. The assessment of the effectiveness of this new class was conducted by comparing exams and feedback of students by the end of the semester. This class provides students with sufficient knowledge of both fundamental statistics and practical data analytical techniques for engineering fields, comprehensive experience in data analytic workflow, and the opportunity to exercise their data analytical skills in engineering applications.

Introduction and Background

Data science is an emerging field based on statistical methods and machine learning techniques to convert extensive dataset into useful insights that can lead to effective decisions. Due to the rapid growth in the size, speed, and diversity of data streams, data analytics technologies have been widely used by many businesses such as insurance, healthcare, and manufacturing as key strategies for decision making. More recently, data analytics skills are increasingly gaining attention in the engineering disciplines, thus many engineering curricula have implemented data analytics as a new course. [1]–[3] However, there is still a lack of design and practice to effectively perform data science instruction in the engineering disciplines, successfully implement a data-centered teaching approach, and develop relevant data science expertise to the engineering undergraduate students.

Data science focuses on information technology and mathematical models to visualize and extract useful information from raw data. Teaching and learning data analytics can be streamlined well in the mechanical engineering curriculum, since the students in mechanical engineering are required to have extensive mathematical knowledge and programming training. At junior or senior year, all the students in the mechanical engineering program at Syracuse University are required to take Data Analytics for Engineers (MAE 333) as an advanced statistical and data analytics course. Aerospace engineering students can take this course as a technical elective. Historically, MAE 333 followed the traditional statistics class structure with the first half of the semester focusing on probabilities and the second half focusing on statistics, assessed by classic approaches of homework and exams.

To meet the current professional needs on data science in engineering disciplines and enhance the students' understanding on data analytics, the primary goal of my class redesign focuses on implementing an interdisciplinary approach to introduce the concepts and principles of data science to the undergraduate students of mechanical engineering. I redesign the MAE 333 in which the students can practice new tools used in data analytics applications while they are still learning the basic statistical principles behind these techniques. I also develop a semester-long project that provides students opportunities to practice and demonstrate their understanding on the course materials.

In this paper, I demonstrate how I structure the course and lecture content based on data analytics lifecycle [4], [5] and how I design the class activities and project based on R programming. I also present my first experience with project design addressing teaching, implementation and assessment of data analytics learning modules in a mechanical engineering undergraduate class. Our course guided the students through emerging experience with real-world large-scale dataset and data analytics workflow by having them perform a series of data preparation, data visualization and machine learning models developed specifically to apply new data analytics techniques in engineering applications.

Lecture Structure and Methodology

The learning goal of this course is to educate students with both statistical fundamentals and analytical skill set to process, analyze and comprehend the data with a data professional's perspective. To achieve this goal, I develop new learning outcomes for this redesigned course, through which the students will be able to:

1. Properly collect and prepare data for analysis;
2. Use a variety of methods to describe and explain data;
3. Identify and utilize commonly used probability and sampling distributions;
4. Understand how data collection dictates the choice of statistical method and perform statistical inferences;
5. Interpret and communicate the outcomes of estimation and hypothesis tests in the context of a problem
6. Understand the scope of inference for a given dataset
7. Understand scripting/code development for data management using R and R-Studio
8. Perform basic computational scripting using R and other optional tools

Based on these learning outcomes, I structure my course content, shown as a weekly syllabus in Table 1, with three major learning focuses: understanding data analytics lifecycle, understanding fundamental probability and statistical concepts, and mastering a data analytics tool.

Table 1: Data Analytics for Engineers Weekly Syllabus

Week	Topics
1	Introduction, raw data
2	Sampling and gather useful data, Probability
3	R – introduction, R – data frames
4	Random variables: discrete, continuous
5	Joint distribution, R – data munging
6	R – data visualization, Midterm
7	Simple linear regression
8	Multiple regression, Logistic regression

9	Sampling distribution, Confidence interval: 1 sample
10	Confidence interval: 2 samples, Hypothesis testing – classic method
11	Hypothesis testing – modern method, Hypothesis testing – two samples
12	Thanksgiving Break
13	Final, R – association rule
14	R – k means clustering, R – decision tree

I arrange the entire course structure based on the forward order of data analytics lifecycle that consists of six basic stages to define how information is created, gathered, processed, used, and analyzed for the final engineering goals. This lifecycle takes place in my class throughout the entire semester in an order of problem discovery, data understanding, data preparation, data visualization, model building, and conclusion/decision making. To be specific, at the beginning of this class, I introduce raw data for problem discovery with different descriptive methods for data understanding. Although students are not required to perform data collection in this lecture, data preparation is also introduced with details in survey design, sampling methods and bias, which aim to help the students be better aware of potential problems with the existing datasets. For the later practice on data visualization, model building and decision making, I couple these contents with R programming for project development and completion.

Statistics is not only the fundamental concept and knowledge base for this class, but also a must-have skillset for the students. It provides guidelines for data preparation, data processing, and data description in quantitative measures. I believe that the students should master the data analytical tools with a thorough understanding of their statistical basis. Therefore, in Week 2, I introduce fundamental probability rules and explain how these rules are associated with the statistical principles. In Week 4 and 5, I introduce the concepts of random variables, including discrete random variables, continuous random variables, and joint random variables. Particularly, the knowledge of joint random variables is one of the key concepts for the next topic of data regression, including linear regression, multiple regression, and logistic regression, which are introduced via lectures and labs (*R* programming) simultaneously. Next, I cover the other key statistical concepts, including sampling distributions, confidence interval, hypothesis testing, which can show the students how to make proper inference between population information and sampled data.

With understanding both data analytics and statistics, students also need to master a programming tool in order to solve a practical project in the class. In this course, I choose *R* as the programming language [6], since it is an open-resource software that can be used for data analysis, data visualization, and statistical modeling. All the students in the mechanical engineering program are required to take a computational tool course based on *MATLAB* programming in their freshman year, which builds a strong foundation for the students to continue learning *R* as an additional programming language. I introduce *R* programming as early as the third week of the semester, focusing on *R/R-studio* interface, *R* programming basics, data frames and variable types. These programming basics greatly echo with the fundamentals in statistics and data analytics that are introduced in the first two weeks of the class to make connections between mathematics and coding practices. In the middle of the semester, data munging and visualization are discussed using *R* programming, which provides students practical insights and technical skills in preparing and processing the raw data. Towards the end of the semester, I introduce three machine learning (ML) algorithms, including association rules, K-means clustering, and decision tree. This short exposure to artificial intelligence provides students with different perspectives on data analytics. More importantly, I also encourage the students to practice these ML techniques in their projects.

Interactive Class Activities

I design and introduce interactive activities during the lectures to improve the student engagement. One activity that I practice in lectures is activity sheets. I design and organize my lecture into 10 - 15 minute segments. After each segment with a specific topic, students are asked to complete the activity sheets as in-class practice. These hands-on practice can reinforce understanding and memorizing the key concepts and formulae, keep students engaged with class activities, and promote students asking questions. For example, in Week 2, we mainly focus on discussing the bias in data collection and processing. After introducing different types of bias, students are asked to complete the activity sheet and discuss with peers on what kind of bias/biases are introduced in the given problems (Fig.1).

Lecture 3 Sampling, Surveys and Gathering Useful Data
Supplemental Material
MAE 333 Data Analytics for Engineers

Collecting and Using Sample Data Wisely

Try It! Fundamental Rule?
For each situation explain whether or not the Fundamental Rule holds.

a. **Research Question:** Do a majority of adults in state support lowering the drinking age to 19?
Available Data: Opinions on whether or not the legal drinking age should be lowered to 19 years old, collected from a random sample of parents of high school students in the state.

b. **Available Data:** Pulse rates for smokers and nonsmokers in a large stats class at a major university.
Research Question: Do college-age smokers have higher pulse rates than college-age nonsmokers?

Bias: How Surveys Can Go Wrong
Type of Bias

Which type of bias **do you think would be introduced if …**

a. A magazine sends a survey to a random sample of its subscribers asking them if they would like the frequency of publication reduced from biweekly to monthly, or would prefer that it remain the same.

b. A random sample of registered voters is contacted by phone and asked whether or not they are going to vote in the upcoming presidential election.

Figure 1. Activity sheets (partial) in Week 2 to encourage students discussing different biases.

I believe that by fully understanding the concepts, students should be able to not only solve the homework problems, but also apply what they have learned in class to real-world problems. Therefore, at the end of each topic, I design in-class discussion with practical problems that people may encounter in daily life, aiming to encourage students to think critically how the concepts taught during the lectures can be applied in real life. For example, after introducing the concept of sampling distributions, I initiated a discussion about how health insurance companies change their policies based on sampling distribution and sample sizes (Fig. 2). Every other week, I also spend 10 - 15 minutes facilitating an open-end topic discussion to encourage students to think from a data professional's perspective on the application of data analytics in real-world problems. Discussion topics include data collection and analysis for streaming services, smart use of data from Coca Cola Freestyle Machine, and how an engineering facility analyzes user data to predict and provide preventive maintenance, etc.

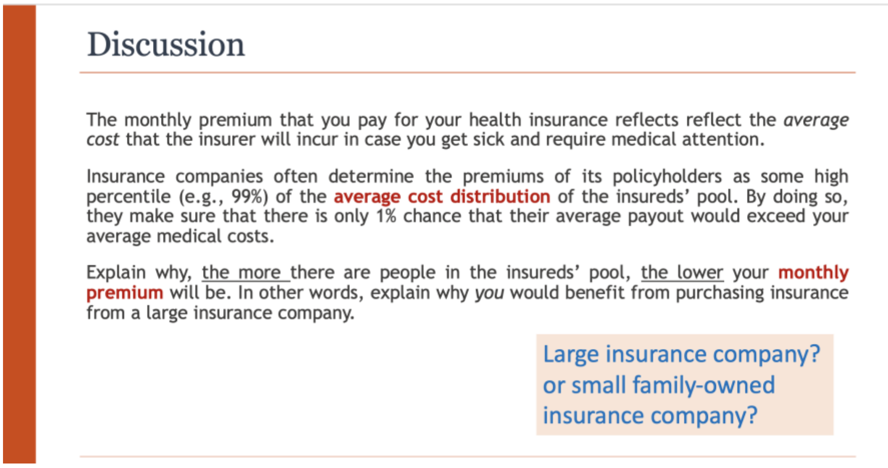

Figure 2. The example of discussion topics related to sampling distribution.

Lecture/Lab Design on R Programming

All the enrolled students have some programming experience in MATLAB and Python, thus introducing R programming can further enrich their programming skill sets. Four weeks of classes are dedicated to R programming that allows students to be more proficient in the homework and class project. The main topics in R programming lectures include:

1. R basic introduction
2. R data frames
3. Data cleaning using R
4. Data visualization using ggplot2 package
5. Perform different regression analysis using R
6. Machine learning algorithms (Association rule, k means clustering and decision tree)

Each topic is introduced via an 80-minute lab session along the semester. For each lab session, I introduce all the concepts based on a problem-driven method. For example, when introducing the ggplot2 package for R visualization, I use a real dataset "Titanic", which contains the data for 887 Titanic passengers. With the data visualization package, I demonstrate how to use different types of graphs to reveal the information hidden behind the data. Students are able to understand the survivors' profile based on gender, race, age and other factors using data visualization. Class plan for data visualization is provided in Table 3.

Table 3. Sample of class plan for data visualization in the lab session.

Format	Class Plan	Estimated Time
Powerpoint Slides	Introduction in graph visualization, recap on different graph types, how graphs go on in real life data	10 min
	ggplot2 package	5 min
Website, EXCEL	R dataset introduction, brief discussion on different attributes	10 min
R studio	Data import, data cleaning and preparation, attributes checking	5 min
	Data visualization demonstration and practice, Questions to answer: 1. What was the survival rate?	40 min

	2. What was the survival rate by gender? 3. What was the survival rate by class of ticket? 4. What was the survival rate by class of ticket and gender? 5. What is the distribution of passenger ages? 6. What are the survival rates by age? 7. What are the survival rates by age in histogram? 8. What are the survival rates by age when segmented by gender and class of ticket?	
-	Free time for questions or other circumstances	10 min

Project Design
To strengthen students' understanding of statistics and skills on data analytics, hand-on projects are assigned to the student groups from the first day of the class. Through the semester-long projects, the students can gain experiences with large-scale raw data sets, apply statistical fundamentals learned from the class to data preparation, processing and analysis, and practice R programming for data munging, data visualization, and model building. In addition, through the collaborative effort in a group project, the students can enhance their communicative skills to effectively discuss their analyzed results with audiences in both oral and written formats. To assure successful completion of the projects, multiple checkpoints throughout the semester are designed and listed in Table 4.

Table 4: Project Checkpoints

Week	Topics
1	Complete CATME
2	
3	Assign group, select project data, project proposal
4	
5	
6	Peer proposal review
7	Receive feedbacks from instructor and peers, group interview, preliminary report
8	
9	
10	
11	Receive feedbacks from instructor, final report, team evaluation
12	
13	
14	

At the beginning of the semester, each student is asked to complete a CATME survey for group assignment (2 – 3 students for one group) [7]. The purpose of this survey is to collect customized student information, such as gender, race, schedule, software skill, writing skill, leadership role style, etc. The group assignment is performed by weighting three major criteria: working schedule, leadership preference and current GPA. Balancing the working schedule can ensure student participation, and distributing the leadership preference can encourage all the group members to contribute to the project. How to group female and minorities students is also considered as some research suggests minorities should not be outnumbered on the student groups [8].

In the next phase, assigned student groups choose their project-of-interest with corresponding datasets from the given project pool with appropriate chosen dataset, which is critical to ensure that students can successfully apply learned class contents and *R* programming to the project and complete the entire data analytic lifecycle. The pre-selected pool generally contains 12 projects, including engineering (wearable stress and affect detection dataset, etc.), social life (adult profile and salary dataset, apartment for rent, etc.), biomedical research (heart failure clinical records dataset, Breast cancer Wisconsin data, etc.) and popular issues (forest fires data, COVID-19, etc.). By the end of this second phase, students are asked to submit their first written document – project proposal. The project proposal aims to facilitate the students to comprehend the chosen dataset and form at least five questions based on this dataset to be addressed throughout the semester. The students within each group can also get to know each other and initiate discussion surrounding their project in the group activities.

In the third phase, the submitted project proposals are distributed among different student groups for peer review. Each team is randomly and anonymously assigned with a group proposal from another group with a different dataset. Based on the grading criteria provided by me, each group is required to provide a grade and associated feedback to the assigned proposal within a week. Meanwhile, the teaching assistant and I work together to provide our grades and feedback to all the proposals. This activity gives the students an opportunity to read other students' work and learn from their peers. The students can also compare the differences or similarities between the peer grading and instructor grading. Fig. 3 and Fig. 4 are grading instructions provided to students to complete the peer review and peer review form returned to the proposal team.

MAE 333 Group project project proposal - 36 points

What to include in project proposal (total: 30 pts):

1. Cover Page (1').
   Include project title, team number, group members, and date.
2. Motivation (4' total).
   Which data did you choose (1'), why you are interested (1'), what is this dataset about? (2')
3. Background introduction (4' total).
   Explain why you pick this data set (2'), any information you collected could help you better understand the data (2').
4. Dataset Description (10' total). You are expected to look at your dataset in R. Use the basic functions that you have learned to get a general understanding about your data.
   a. (6') Look into your data, describe each attribute of the selected data (for example, what does each attribute mean and how they are collected, variable type, missing values, etc.). For straightforward attributes, like gender, year, etc., you can briefly describe them; however, for attributes that are involved with professional terms/understanding, you should provide more detailed explanations.
   b. (4') Provide an in-depth understanding of the data structure (relationships between attributes, predictors, factors, etc.)
5. Questions Generation (10' total, 5 best questions are used for grading, 2' each).
   Read your data carefully and generate questions you want to answer through this project, for each question, state the potential attributes that could possibly answer the question. You should have at least one broad question that you can answer after all of the analysis is finished. You also want to come up with at least 5 questions that could help understand the broad question(s) you generated. Each question should provide a different perspective to your project.
6. Reference (1').
   Project original data and related documents are properly referenced. Provide the original data link to the website.

Other requirements (6 pts):

1. Formatting (2'). Use Times New Roman style with 12-pt font, 1.5 line space, and 1 inch margin around.

2. Language (4'). Your work should be written in English fluently, no typos or misspellings. Your work is expected to be in a technical report fashion and reader friendly.

What to submit on Blackboard:

Project proposal in Microsoft word document. **Do not mention** group # or team members' names at any other place in the proposal other than the cover page. Failing to do so will result in 20% point deduction.

For grading, points for each category will be broken down to 3 levels: full mark (excellent job), 50% mark (great but need more details, some of the justifications are wrong (minor mistakes)), 0 (not correct, or no related information is provided)

Figure 3 Project proposal grading criteria.

MAE 333 Group Project Proposal – Peer Review Form

Which dataset are your team reviewing? ____________________

Point Distribution	Total Points	Points Earned	Comments
1. Cover page	1′	1′	You will not see the cover page as it contains the group information.
2. Motivation	1′		
	1′		
	2′		
3. Background introduction	2′		
	2′		
4. Dataset Description	a. 6′		
	b. 4′		
5. Question Generation (pick top 5 questions to grade, each worth 2′)	2′		
	2′		
	2′		
	2′		
	2′		
6. Reference	1′		
7. Formatting	2′		
8. Language	4′		
TOTAL GRADE	36′		

- The point distribution for each part is clearly marked in Stage 1 Project proposal file. Use the file along with this form for grading purposes.
- For grading, points for each row under "total points" can be broken down to 3 levels: full mark (excellent job), 50% mark (great but need more details, some of the justifications are wrong (minor mistakes)), 0 (not correct, or no related information is provided)
- For example, for 4b, the group can earn 4′ if you think their description is clear, reasonable based on reading the dataset and without any error. If there are errors (they made assumptions that they shouldn't, or the description is not very clear after reading), they will earn 2′. If they didn't provide any information regarding this session, you can give 0′. There is no point in between.
- If you deduct any points, you should justify why under "comments", you may also provide suggestions.

Figure 4 Project proposal peer review form.

After all the grades and feedback are returned to the students, phase four is initiated with interactive group interviews, during which I discuss with each group about the chosen dataset, their own proposal questions, and peer review feedback. Meanwhile, student groups can start preparing the preliminary report to incorporate received feedback from both peer review and interactive interview. In the preliminary report, the students are required to complete data background study, raw data description, data cleaning and transformation using R programming, and then propose next-step plan for data visualization. By the end of this phase, each group should have a prepared dataset ready for in-depth analysis and model building using the machine learning algorithms.

In the last phase of the group project, each group receives feedback from me and starts working on their final report, which is a continuation of the preliminary report. In the final report, each group is required to perform data visualization, regression and use at least two machine learning models. At the end, the students need to comprehend the results generated from different techniques and provide a thorough discussion and conclusion of their work. In addition to my grades, team evaluation is also collected and weighted into the final grades.

Observations, Reflections and Discussions

The redesigned course has been offered three times in the fall semester of the year 2019 (in person), 2020 (virtual Zoom) and 2021 (in person). Here, I only focus on the course taught in person to reflect the challenges that I have experienced in redesigning and delivering this course, since online teaching adds significant complexity to make direct reflection on teaching effectiveness and student learning experience.

1. The relatively large student enrollment limits the faculty – student interaction, and a similar issue has been widely reported [9]. In 2019 and 2021, the class enrollments were 69 and 58 respectively. This challenge could be resolved by dividing the current session into two sessions or more.
2. Implementing the in-class activities in a traditional lecture-based classroom environment is challenging. Some of the class activities are designed to initiate discussions. The seating in the current class room is arranged in the row settings, which makes it difficult to talk to

students who sit far from each other. Using the classroom that is designed for active-learning can resolve this challenge.
3. Implementing R programming during lectures is challenging. Currently, the number of student enrollment exceeds the capacity of computer labs on campus, thus students are asked to bring their own laptop to install R software to practice programming activities with the instructor. This challenge could be resolved by dividing the current session into two sessions, or leasing new laptops from IT services that have been installed with R software.
4. Students are using MATLAB as the main programming language throughout their coursework, thus it is not an easy smooth transition between languages at the early stage of learning R programming. It takes students some time to make adjustments between different logics embedded in these two programming languages. Early introduction of R programming at lower-level courses might mitigate this problem.

In light of the challenges noted above, the class activities and project were positively received by the students.

1. The in-class discussions were designed to spark students' interest and critical thinking in data analytics. Most of the students were engaged and participated in the class discussion. Through the discussions, some of the students were able to apply the key knowledge learned from the lectures and make the connections between course concept and their experiential learnings.
2. The class activity sheets were appreciated by the students. I have received positive students' feedback from course evaluation surveys commenting on the use of activity sheets, like " *keep them focused on the lecture materials, provide more opportunities to calculate problems, and help them better understand the class topics*". In addition, some students also commented that they "*would like to have more class activity sheets*".
3. Although occasionally I received feedback from students who were struggling with R programming or thought that R programming was not necessary for engineering students. Most students indeed appreciated the introduction of R as a new programming language. As 60% - 70% of the students in this class were seniors, they "*found the R Studio component of this class very helpful because there are some jobs that look for that skill*". I also received emails from students by the end of the semester expressing their interests in data analytics and willingness to continue learning these topics in depth.
4. Students found that peer proposal review and group interview were very helpful in guiding them to work through the project at different stages along the entire semester. Peer review allows students to look into project grading rubrics and receive feedback and suggestions from not only the instructor and TA but also their peers. I have observed that peer grades were similar to the TA's grades in most cases, while peer grades tend to be much harsher than the TA's grades for the proposals that were not well written.
5. For a class with a relatively large size, group interview allows the instructor to discuss with the student teams on a one-to-one basis. This was a valuable opportunity for the instructor to check the project progress for each team, identify misinterpretation from the students on the chosen dataset, and provide timely guidance to the teams to ensure they are on the right direction for the next phase of the project.

Student course evaluation surveys were completed at the end of the semester (Fall 2021). There were 58 students enrolled in the class and the response rate was 79.31%. The questions most relevant to this redesign effort are listed below:

Table 5. Course evaluation survey questions that are closely related to the course redesign.

Question #	Description
1	Student participation and the contribution of ideas, comments, and questions were encouraged.
2	Tests and other assessments accurately measured what I have learned in this course.
3	I received helpful feedback from the instructor to guide my progress in this course.
4	The instructional materials (i.e., books, readings, handouts, study guides, lab manuals, multimedia, software) increased my understanding of the subject matter.
5	Class activities, examples, and discussions helped me to better understand the main concepts in this course.
6	The instructor was engaging and effective when delivering the material.
7	I am able to use a variety of methods to describe and explain data.
8	After taking this class, I have a better understanding of data analytics.

For these 8 questions above, students' responses are based on the scale from strongly disagree (weight as 1 in the score calculation) to strongly agree (weights as 5). I also calculated the corresponding frequency as well as statistics, shown as follows:

Table 6. The distribution of answers to the course evaluation questions and statistics.

Question	Strongly Disagree (1)	Disagree (2)	Neither Agree Nor Disagree (3)	Agree (4)	Strongly Agree (5)	Means	STD
1	0%	0%	8.7%	54.35%	36.96%	4.28	0.62
2	0%	6.52%	8.7%	52.17%	32.61%	4.11	0.82
3	0%	2.17%	4.34%	63.06%	30.43%	4.21	0.91
4	0%	2.22%	13.33%	44.44%	40%	4.22	0.77
5	2.17%	0%	6.52%	45.65%	45.65%	4.33	0.79
6	2.17%	2.17%	2.17%	36.96%	56.52%	4.43	0.83
7	0%	0%	4.44%	42.22%	53.33%	4.49	0.59
8	2.17%	0%	4.34%	54.35%	49.14%	4.28	0.93

Questions 1 - 5 evaluate the students' feedback on course redesign. The mean scores for those questions are around 4.2 out of 5, which indicates that students generally are satisfied with the course content and the course overall. Question 7 and 8 request the students to self-reflect on their abilities of data analysis after taking the course. The average scores of these two questions are 4.49 and 4.28 out of 5, respectively with relatively low standard deviation, which indicates that the overall course objectives are achieved based on students' feedback.

Reference

[1] Z. del Rosario, “An Open-Source Active Learning Curriculum for Data Science in Engineering,” *Journal of Open Source Education*, vol. 5, no. 49, p. 117, Mar. 2022, doi: 10.21105/jose.00117.

[2] D. Otoo-Arthur and T. Van Zyl, “A Systematic Review on Big Data Analytics Frameworks for Higher Education - Tools and Algorithms,” in *Proceedings of the 2019 2nd International Conference on E-Business, Information Management and Computer Science*, New York, NY, USA, Aug. 2019, pp. 1–9. doi: 10.1145/3377817.3377836.

[3] “Opportunities and challenges for big data analytics in US higher education: A conceptual model for implementation - Mohsen Attaran, John Stark, Derek Stotler, 2018.” https://journals.sagepub.com/doi/full/10.1177/0950422218770937?casa_token=oxs3EVRBy-kAAAAA%3APHOgN0LLjQzr-zOlDn3nVHGYWf7Q1rzYBItazbzKxpwM1cbKvZ2ovLYovX4WTfHrVMFAFXWRp8rS (accessed Mar. 27, 2022).

[4] K. Rahul and R. K. Banyal, “Data Life Cycle Management in Big Data Analytics,” *Procedia Computer Science*, vol. 173, pp. 364–371, Jan. 2020, doi: 10.1016/j.procs.2020.06.042.

[5] J. Dean, *Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners*. John Wiley & Sons, 2014.

[6] “R: The R Project for Statistical Computing.” https://www.r-project.org/ (accessed Feb. 09, 2022).

[7] “CATME Project - LOGIN.” https://catme.org/login/index (accessed Mar. 27, 2022).

[8] B. Oakley, “Turning student groups into effective teams,” *Journal of student centered learning*, vol. 2.1, 2004.

[9] “Decreased class size, increased active learning? Intended and enacted teaching strategies in smaller classes - Mary C Wright, Inger Bergom, Tracy Bartholomew, 2019.” https://journals.sagepub.com/doi/full/10.1177/1469787417735607 (accessed Mar. 27, 2022).