

70-469 End to End Business Analytics

Time & Place: MW 11.50-1.10 pm, TPR 2612

Semester: Fall 2021

Units: 9, **Section:** A

Instructor information

Name	R. Ravi
Contact Info	ravi@cmu.edu
Office location	TEP 4212
Office hours	11-noon Fridays (Or email for appointment)

TA Information

TA name	Daniel Deroux Uribe
TA Contact Info	dderoux@andrew.cmu.edu
Office hour location	TBD
Office hours	TBD (Or email for appointment)

Course Description

Deriving value from business data is an integrative process. It requires putting together the virtuous pipeline of using the data to derive descriptive and diagnostic inferences, building explainable predictive models, and incorporating them in prescriptive decision making. This course charts this process end-to-end by describing and integrating common tools from statistics, machine learning, and optimization in the context of common applications from finance, marketing, and operations. The homework is based on a data-driven case that will have students analyzing data from real business applications to derive their own insights, predictions and decisions and communicating them effectively. The course will prepare students for careers in consulting and any form of business data analysis in any functional area.

Since the course is focused on integrating various techniques from Statistics and Optimization, the foundational courses on building such models are prerequisites: Statistics (70-207 or 36-200 or 36-220) & Optimization (70-257 or 21-257 or 21-292).

Learning Objectives

On completing the course, you should be able to

- ✓ Translate a data-driven business problem to analytical models based on their purpose, available data, solution methods, evaluation metrics and deployment strategies.
- ✓ Make judicious modeling choices in the different stages of the translation using your knowledge of the techniques used in different analytical models.
- ✓ Communicate the results of technical business data analysis models to all stakeholders to convince them of its value, validity, effectiveness, and risks in eventual deployment.
- ✓ Prototype business analytics models using R and Excel and use them in all the steps above.

Learning Resources

The course will require coming to class every week, attending classes ready to participate and work in the in-class labs, completing the weekly quiz online over the weekend to enhance revision and practice of the material, and turning homework updates as part of a team on a project that extends throughout the semester. The course will introduce and build models in Excel and R, a free programming language used ubiquitously in modern analytics ecosystems. Prior programming experience is not required. Class sessions will introduce new ideas every week that will typically be followed up by walking through a prototype model in R or Excel.

The following books are recommended (but not required). If you would like to buy one, I would suggest Abbott since it is the broadest in coverage. BDS and ISLR are more technical.

1. [Applied Predictive Analytics: Principles and Techniques for the Professional Data Analyst](#) by Dean Abbott. Unlimited number of online copies available on course eReserve at the CMU library [here](#).
2. [Business Data Science](#) by Matt Taddy. Unlimited number of online copies available on course eReserve at the CMU library [here](#)¹.
3. [An Introduction to Statistical Learning with Applications in R](#) by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani (available online).

For those with no programming experience, the following two books will serve as good references for basic introduction to R.

1. [Programming Skills for Data Science](#) by Freedman and Ross. Unlimited number of online copies available on course eReserve at the CMU library [here](#)².
2. [R for Everyone](#) by Lander. Unlimited number of online copies available on course eReserve at the CMU library [here](#)³.

Assessments

There will be an online quiz to be completed individually every other week. The homework should be done in teams of two students that are formed at the beginning of the semester and retained throughout: The homework is to work on a case study that involves the end-to-end analysis of a real data set. Each

¹ Click on the O'Reilly for Education link in the "View Online" section of the [library catalog record for the book](#). You will get a "Select your institution" dropdown menu prompt. Select the "Not listed? Click here" option in that dropdown. Enter your Andrew email address and click "Let's Go." You will then get full text access to the book through the platform after signing in through CMU Andrew SSO.

² Same instructions as for the Abbott reference.

³ Same instructions as for the Abbott reference.

homework submission will involve submitting a short update report (PDF file of 3-5 pages) and a code workbook in R or Excel – To aid your analysis, simple scripts or workbooks will be provided as a template for each such assignment. In your report, I would like a careful description of the main findings - plan to formulate your charts and tables to highlight your most interesting findings that also build up the story for your final presentation and report.

Software

We will use R and Excel to develop prototype models for this class. We will use R studio for model development in R. All in-class labs and the homework related to the class project will involve starting from scripts provided, understanding them, interpreting the results, and modifying them to extend and deepen the analysis.

Grades The final course grade will be calculated using the following categories:

Assessment	Percentage of Final Grade
Weekly Quizzes (individual – best 5)	5 X 8% = 40%
Five homework project updates + Final folio (Teams of 2)	5 X 8% + 10 = 50%
Class participation and team effort	10 %

- **Homework (5 pages pdf total):** Use the guiding questions in each of the first five homework assignments to construct a narrative for presenting the update of your analysis. Aim for a smart business audience without making the report overly technical. (Please add supporting material such as a script that is well annotated as one additional supplementary file - zip into one if you have many).
 - ✓ Use at most 5 pages including any graphs, charts, and model outputs to explain your reasoning.
 - ✓ Do not regurgitate facts but use the report to highlight what new discoveries you made from your analysis (on which data set) and how that might be useful or important
 - ✓ Craft the report to highlight the main findings and how they might be useful, then explain how the findings were arrived at, followed by assumptions and risks in the analysis, ending with a way to take the findings forward in the eventual pipeline of your work.
- **Final presentation (10 minutes, 6 slides):** The goal of the final presentation is to provide a convincing case for the deployment of analytical modeling in a business situation and making a strong convincing call to action to deploy the recommendations based on your analysis.
 - ✓ Make the presentation goal oriented and a call to specific action.
 - ✓ Use the recommendation to drive the structure of your arguments logically
 - ✓ Articulate the end-to-end pipeline of your data analysis and its component parts
 - ✓ Anticipate audience concerns in your delivery
 - ✓ Arrange checkpoints in your presentation to make sure that the audience is following or invite them to contribute an idea or suggestion to keep them engaged
 - ✓ Begin and end with powerful visuals and delivery
- **Final report (7 pages pdf total):** In addition to summarizing the key points from your presentation, think of how a report can communicate some key details and arguments that back up your presentation. Rather that repeat the arguments in the presentation, you must either present them in a different perspective or provide more supporting arguments for your presentation points. Here are some tips for writing an effective report:

- ✓ The writing overall should be crisp: state the main problem addressed in the first paragraph (or a summary paragraph), and the resulting benefit (increase in effectiveness) from following the recommendations of your analysis, as much as possible in quantitative business terms relevant to the context.
- ✓ Outline the argument of how your model's benefit is achieved and how it works, as well as any new insights it unearthed and leveraged to drive your recommendation that you could not detail in your presentation
- ✓ Use the title of the report, sections, and the various subsections/paragraphs to reflect the main message and not something generic.
- ✓ Separate your new ideas clearly from your references.
- ✓ Make specific deductions or recommendations.
- ✓ Anticipate shortcomings/risks of recommendations/approaches used towards the end of the paper.
- ✓ Close with a recap paragraph restating the main message, recommendation, and a call to action.

General Policies

- **Academic Integrity & Collaboration:** Academic Integrity is a core CMU value, and as a member of the CMU community, it is important that the work you turn in for this class is wholly your own. As your instructor, I will strive to ensure that you develop the necessary knowledge and skills to meet the learning objectives for this class, just as it is your task to put in the effort to complete the work and ask for help if you need it. In this uncertain environment for Fall 2021, you might have questions about what is and is not acceptable. The weekly online quizzes are strictly individual work; while you may study with class mates to prepare for each quiz, it is forbidden to consult anyone while you are taking those timed quizzes. The homework assignments are done in groups and it is important that each group develops its own approach to answering the key issues in each homework. While you can discuss with other groups about their approaches, you will need to implement your groups approach yourself and write independently about it in your reports. The [University's policies on academic integrity](#) govern the class, so please make yourself aware of them especially when in doubt.
- **Accommodations for students with disabilities:** If you have a disability and have an accommodations letter from the Disability Resources office, I encourage you to discuss your accommodations and needs with me as early in the semester as possible. I will work with you to ensure that accommodations are provided as appropriate. If you suspect that you may have a disability and would benefit from accommodations but are not yet registered with the Office of Disability Resources, I encourage you to contact them at access@andrew.cmu.edu.
- **Technology Use During Class:** This semester involves regular use of technology during class – both for in-person and remote students. Research has shown that divided attention is detrimental to learning, so I encourage you to close any windows not directly related to what we are doing while you are in class. Please turn off your phone notifications and limit other likely sources of technology disruption, so you can fully engage with the material, each other, and me. This will create a better learning environment for everyone.
- **Late Work:** Remember: If you registered for this class, you have until November 9th to change your grade in this course from a letter grade to a Pass/Fail grade. All quizzes and assignments have due dates indicated on the syllabus. In general, submitting assignments on time lets the instructional team provide feedback in a more timely and efficient manner. Assignments build on each other, so timely submissions are crucial to your progress in the class. However, sometimes life happens. If you cannot submit an assignment on time, the default will be that you will be eligible for 90% of the grade the first 48 hours that the assignment is late. If you have to submit beyond 48 hours past the due date, please contact me as soon as possible so we can make arrangements.

- **How to Succeed in this Class:** This class is designed to go over a wide range of topics, so attendance is very important. If you are unable to attend, you must plan to review the video of the class or read the appropriate material from one of the recommended sources. The weekly labs should give you a good pulse of whether you are able to understand the material for that week. The weekly quizzes are also designed to make sure that you revise the material at the end of every week before moving on to the next module in the class. Since the material is very technical, I prefer not to lecture for long periods but will do my best to engage you in a discussion during the class using the examples and during the labs, so expect to be cold-called. To the extent possible, I would really appreciate if you are able to keep your camera turned on for most of the class so our discussions can happen more spontaneously. The homework assignments to be done in small teams are designed to ensure that you can apply the materials you are learning in the lectures and labs to a real-life problem and explain your approach and results. I am open at any point for feedback that will improve the class for everyone.
- **Recordings and Videos:** Do not use mobile phones during class. No student may record (in video or audio form) any classroom activity without the express written consent of the instructor and the Tepper administration. Remember that the readings and videos shared with you are copyrighted by the instructor and cannot be shared or posted anywhere without his written permission.
- **Statement on student wellness:** This semester is unusual. We are all under a lot of uncertainty at this time. Make sure to move regularly, eat well, and reach out to your support system or me if you need to. We can all benefit from support in times of stress, and this semester is no exception. As a student, you may experience a range of challenges that can interfere with learning, such as strained relationships, increased anxiety, substance use, and feeling down, difficulty concentrating and/or lack of motivation. These mental health concerns or stressful events may diminish your academic performance and/or reduce your ability to participate in daily activities. CMU services are available, and treatment does work. You can learn more about confidential mental health services available on campus at: <http://www.cmu.edu/counseling/>. Support is always available (24/7) from Counseling and Psychological Services: 412-268-2922.

In case of Remote Only Instruction

- **Zoom Usage:** In case we transition to a remote only posture for instruction, we will be using Zoom for synchronous (same time) sessions. The link will be available on Canvas.
 - Please make sure that your Internet connection and equipment are set up to use Zoom and able to share audio and video during class meetings. (See this page from Computing Resources for information on the technology you are likely to need.) Let me know if there is a gap in your technology set-up as soon as possible, and we can see about finding solutions.
 - Sharing video: In this course, being able to see one another helps to facilitate a better learning environment and promote more engaging discussions. Therefore, our default will be to expect students to have their cameras on during lectures and discussions. However, I also completely understand there may be reasons students would not want to have their cameras on. If you have any concerns about sharing your video, please email me as soon as possible and we can discuss possible adjustments. Note: You may use a background image in your video if you wish; just check in advance that this works with your device(s) and internet bandwidth.
 - During our class meetings, please keep your mic muted unless you are sharing with the class or your breakout group. If you have a question or want to answer a question, please use the “raise hand” feature (available when the participant list is pulled up). I will monitor this from time to time during the lecture, however, please minimize the use of chat during the class and use it only for minor comments among yourselves.
- **Zoom session recordings:** I will make the recordings available on Canvas as soon as possible after each class session (usually within 3 hours of the class meeting). Recordings will live in our Canvas website under the Zoom tab. Please note that you are not allowed to share these recordings. This is to protect your FERPA rights and those of your fellow students.

Tentative Course Schedule

Week	Theme	Topics	Work Due
8/30	Framework	What is End-to-end Analytics?	
9/1	Framework	Value from Data	
9/6	(Labor Day)		
9/8	Framework	CRISP-DM	HW1
9/13	Descriptive Analytics	Dimension Reduction	
9/15	Descriptive Analytics	PCA	Q1
9/20	Descriptive Analytics	Clustering	
9/22	Descriptive Analytics	K-Means	Q2
9/27	Predictive Analytics	Regression review	
9/29	Predictive Analytics	Classification metrics, overfitting	HW2
10/4	Predictive Analytics	Logistic Regression	
10/6	Predictive Analytics	Logistic Regression	Q3
10/11	Predictive Analytics	Decision Trees	
10/13	Predictive Analytics	Tree models	Q4
10/18	Refining Predictions	Regularization, Model Selection	
10/20	Refining Predictions	Regularization	HW3
10/25	Using Predictions	Using clustering for explaining predictions	
10/27	Using Predictions	Personalizing actions based on clustering, ranking using models	Q5
11/1	Prescriptive Analytics	Optimization framework	
11/3	Prescriptive Analytics	Optimization methods	HW4
11/8	Prescriptive Analytics	Sensitivity Analysis	
11/10	Prescriptive Analytics	Communicating policies	Q6
11/15	De-risking	Common Pitfalls	
11/17	Communicating	Effectively communicating technical info	HW5
11/22	Guest Lecture		
11/24	(Thanksgiving - No class)		<Nothing due>
11/29	Presentations	Student presentations	
12/1	Presentations, Review	Student presentations	HW6