

Teaching Case

Managing Data Analytics Group Projects in Large Online Courses

Prof. Frederic Bien
fvbien@gatech.edu
Scheller College of Business
Georgia Institute of Technology
Atlanta, GA 30308, USA

Abstract

In business management courses, it is common practice to ask students to do projects by groups of 3 to 5 students. Students must present their projects in front of the class and submit a written report with graphics and code. Students learn a lot from doing these projects and hone their skills for doing business presentations. With the advent of online education, a key challenge is, how can we assign group projects in large online courses? How can one guide and grade lots of projects, and offer students a chance to present their projects in front of the class to get meaningful feedback? We present here a method we have used successfully in courses of up to 1,000 students per semester in an Online Masters in Analytics program. At such class sizes, one gets 150 to 200 group projects, which is not humanly possible to manage in a traditional way. Working with teaching assistants trained to interact in an efficient way, using online *peer* grading for *certain* aspects of the deliverables, and relying on multiple online learning platforms, we have obtained repeated success in managing this many student projects. We share this method to provide ideas and boost confidence for other online courses in data analytics.

Keywords: DataAnalytics, OnlineCourse, GroupProject, BusinessCases, LargeClasses, OnlinePresentations, ActiveLearning, ExperientialLearning

1. INTRODUCTION

"Doing is a great way of learning," especially when one can *do* things that one is interested in. *Doing group projects* is known - and has been shown in pedagogy research - to be an efficient method of learning. This applies also to data analytics courses where statistical methods can often seem abstract - too easy or too hard - to students who don't have a chance to actually "put their hands" in the data.

In on-campus business management courses, it is a common and useful practice to ask students to do group projects by teams of 3 to 5 students. This type of learning activity is an essential part of any business education throughout the world.

Students must present their projects in front of the class and submit a written report with code and graphics. Students can learn a lot from doing such projects and they hone their skills for business presentations.

With the advent of online education, the challenge is now, how can we offer "progressive education" in large online courses? That is, a type of education that is more than just asking students multiple choice questions to check if they have read the reading material or watched the videos.

Offering students to do group projects on a large scale is a key step toward this goal. How can one guide and grade students properly, and offer them a chance to present their projects with appropriate feedback?

We present here a method we have used and tested successfully in courses of up to 1,000 students per semester, in an Online Masters in Analytics program. At such class sizes, one gets 150 to 200 group projects. These numbers make it humanly impossible to guide and grade in a traditional way with the instructor looking at every project in one semester.

Doing group projects online turns out to be a big challenge because with the current surge of interest in data analytics degrees and jobs, online versions of business analytics courses attract a lot more students than one sees in on-campus programs. In our experience we have had class sizes of 600 to 750 students per semester, even when a course is offered three times a year.

To offer group project in online courses, one must overcome the following obstacles:

1. Solve the issue of group formation among students,
2. Figure out how to read enormous amounts of student reports and grades their paper,
3. Provide guidance like an instructor would in an on-campus schedule
4. Sometimes be the "counselor" of problematic teams from an interpersonal viewpoint.

We found, perfected and tested an efficient way of running group projects in business analytics. We present it here, so that others can also use it.

We applied this method to organize and grade 130 to 150 original and independent group projects per semester, more than it is humanly possible for one instructor to just review.

We found that students for the most part collaborated very well with each other, and many created excellent quality projects. Only about 2 percent of teams had interpersonal organization issues.

Results expressed in course/instructor opinion surveys (CIOS) were slightly superior or similar to assigning multiple choice question (MCQ) tests to students, with the proviso that projects demand more time to students to work on the course material as compared to MCQ tests.

This teaching case presents the methods we have used to manage business analytics group projects on a large scale. The outline is: a review of the movement of progressive and experiential education methods, benefits of this approach, the need in business analytics, methodology used to

manage group projects in large online courses, results, discussion/limitations, and conclusion.

2. PROGRESSIVE, ACTIVE AND EXPERIENTIAL EDUCATION

Progressive Education is a pedagogical movement that began in the late 19th century and continues today. The term *progressive* was used to distinguish this education method from the traditional curricula of the 19th and early 20th century, which was rooted in classical preparation for university courses. By contrast, progressive education finds its roots in modern experience.

Progressive Education is based on the principle that education must prepare students for active participation in a democratic, global society.

Thus, the focus is on raising critical thinkers and inquirers who are active rather than passive learners. The teacher's job is to nurture and support children's innate curiosity and desire to learn, fostering internal motivation rather than reliance on external rewards. A key goal is that the teacher's role is to assist in removing the obstacles that might be blocking the road of students to learning new material and skills.

Most progressive education programs have several qualities in common:

- Emphasis on learning by doing : hands-on projects, experiential learning
- emphasis on problem solving and critical thinking
- Understanding and action as the goals of learning as opposed to rote knowledge
- De-emphasis on textbooks and multiple choice questions in favor of varied learning resources
- Group work and development of social skills
- Assessment by evaluation of student projects and productions

With the advent of online education, the challenge is now, how can we offer progressive education in large online courses? Offering students to do group projects on a large scale is a key step toward this goal. How can one guide and grade students properly, and offer them a chance to present their projects with appropriate feedback?

Progressive education principles were explored and embraced by John Dewey, Kurt Lewin, Jean Piaget, Carter Woodson and others who studied education of children and adults.

Within the framework of Progressive Education, one finds the concept of "Active Learning" in classrooms, meaning that students participate, collaborate with others, and apply learned concepts to the real world. It can require hard mental effort but leads to better retention and an understanding of the material that can be transferred to other situations.

Active learning was written about by Bonwell and Eison (1991) in their influential article "Active Learning: Creating Excitement in the Classroom." Other professors such as Carl Wieman, a Nobel Prize winning physicist at Stanford University, researched and championed this approach. (Wieman, 2021)

Experiential learning is the process of learning through experience, and specifically "learning through reflection on doing". The concept of learning through experience is ancient. Around 350 BC, Aristotle wrote in the *Nicomachean Ethics*: "for the things we have to learn before we can do them, we learn by doing them." In the 1970s, David Kolb helped develop the modern theory of experiential learning.

Experiential learning has significant teaching advantages. Peter Senge in his book *The Fifth Discipline* (1990) states that teaching is of utmost importance to motivate people. Learning only has good effects when learners have the desire to absorb the knowledge. Therefore, experiential learning requires the showing of directions for learners. To go further, when developing a group project with a topic chosen by the students, one should ask students to reflect on their product.

Experiential learning is distinct from rote or didactic learning, in which the learner plays a comparatively passive role. It is related to, but not synonymous with, other forms of active learning such as action learning, free-choice learning, cooperative learning service learning and situated learning.

It is worth noting that in Senge's perspective, the basic 5 disciplines of learning and highly performing organizations are: 1) Personal Mastery, 2) Mental Models, 3) Building Shared Vision, 4) Team Learning, and 5) Systems Thinking, which is the integrative (fifth) discipline that fuses the other four into a coherent body of theory and practice. This fusion is what is sought after in the assignment and development of a group project with self-reflection by the team.

3. BENEFITS OF GROUP PROJECTS IN EDUCATION

Benefits for Students

Data analytics are a natural field where to apply student group projects. When structured properly, group projects can promote important intellectual and social skills and help to prepare students for a work world in which teamwork and collaboration are increasingly the norm, (Caruso & Woolley, 2008; Mannix & Neale, 2005), and (CMU #1), (CMU #2).

Positive group experiences have been shown to contribute to student learning, retention and overall college success (Astin, 1997; Tinto, 1994; NSSE, 2006).

Properly structured, group projects teach and reinforce skills that are relevant to both group work and individual work. These skills are important in future career development.

Here are some examples of skills acquired through group projects:

- Break complex tasks into parts and steps
- Plan and manage time
- Delegate and collaborate on task completion
- Refine understanding through discussion and explanation
- Challenge group assumptions
- Develop stronger communication skills
- Give and receive feedback on performance

Group projects can also help students develop skills specific to collaborative efforts in small teams. These skills serve well future workers.

"There are significant *potential* learning benefits of group work. Note the 'potential' because simply assigning group work is no guarantee that these goals will be achieved." (CMU #2)

Group projects can create negative experiences when they are not designed, monitored and assessed in a way that encourages meaningful teamwork and rewards real collaboration.

When well-run, group projects provide oftentimes the one aspect of the course that students will remember the best from their course, years later.

It is easy to observe this, even with our own memory, students quickly forget what they learned in which course. Yet their group project

remains generally more firmly implanted in their brains than other material they have learned.

Benefits for Instructors

Self-chosen group assignments are useful when there are a limited number of viable project topics to distribute among students. Group project can reduce the number of final products instructors have to grade, although they are a lot more complex to grade than tests with multiple choice questions, since the work product is open-ended and requires reading/viewing by a human.

Complex, authentic problems can be challenging to find for faculty, although they can assign more challenging problems to group than to individuals.

Group work allows potential competition between different groups who can focus on different tasks and solve problems in novel and interesting ways. This is helpful for instructors and students who like to go beyond textbook-based instruction.

Note also that group project instructions should assign group work tasks "that truly fulfill the learning objectives of the course and lend themselves to collaboration." (CMU #2)

4. THE NEED FOR GROUP PROJECTS IN BUSINESS ANALYTICS EDUCATION

Group projects in business analytics education help address an important issue observed worldwide in the current business landscape.

Many business organizations have already built some of the foundation needed to utilize data analytical thinking in their business process. They've begun collecting data and running descriptive analytics tools like dashboards. What some companies still need, however, are the supporting elements to conduct predictive and prescriptive analytics processes, then make decisions and take actions based on these analytics.

Business analyst Gartner observed in their 2019 IT Top Data and Analytics Predictions that only 20% of analytics initiatives in the USA will deliver business outcomes based on objectives, while the other 80% will fail to deliver tangible results. Gartner even said that "through 2020, 80% of AI projects will remain alchemy, run by wizards whose talents will not scale in the organization." (Gartner 2019)

It's easy to fall into a technology-first approach when thinking about analytics. A common cause of failed analytics projects is not the wrong tools

or a lack of technical proficiency. Often, the main issues are business process problems. Here some examples; one could add or find others:

- Asking the wrong research questions
- Failing to define a clear purpose for collecting data
- Selecting the wrong uses for analytical methods
- Lack of supporting business culture

It takes technical expertise and business acumen to establish the conditions for analytics to thrive. Business leaders must know which technologies to use, when to deploy them and how to optimize the outcomes. These are challenges for business organization that deploy analytics projects.

Quantitative thinking has become an indispensable aspect of leadership. Managers have to be trained to look for and see the big picture of how metrics can be connected to a business problem. "Then one needs to be able to drill down deeper to see what are the sub-problems that need to be solved. One does not have to be an expert at coding every aspect of the problem, but one has to be able to see how these organization pieces fit together." (CMU #3)

5. ONLINE GROUP PROJECT SITUATION

Online group projects provide valuable opportunities for in-depth learning with business and social skills development beyond 'traditional online education'. This sounds like an oxymoron "traditional online", yet it is our new reality post-2020 Covid pandemic.

Online students interface with course materials via learning management systems, such as Canvas, edX, Blackboard, TopHat, Coursera, 2U, GetSmarter, etc. There are some very good ones.

In our experience, we had to run group projects concurrently in Canvas.com and in edX.org. Each cohort had several hundred students in it. Having two LMS complicated things for us, yet it forced us to adopt an approach that is flexible and can fit multiple LMS, besides the ones we used.

Other online tools used were:

- Piazza
- Google Docs
- GitHub
- Vocareum
- PeerFeedback

Online Discussion Forums

Piazza is a convenient platform for running online discussion boards. This is where all students ask their questions about the course. This way all TAs see the questions and can respond. It is a lot easier to manage than email, which otherwise would get out of hand quickly.

Our courses typically average over 1,000 posts (or questions) per semester in Piazza, per cohort or LMS, and 6 to 7,000 contributions in total. Online forums are an efficient platform for the TA team to collaborate in answering student questions. We work with an "army of TAs". Typically, we get one TA for every 40 to 50 students. Therefore, a course teaching team might have 20 TAs, (who must be hired, trained and coached. It's another topic...)

Students can also ask private questions that only they see, and all TAs see. Piazza is also capable of handling group communications, so that a small group of students (3 to 5) would see a message that has been posted and all TAs could answer it for the benefit only of that group.

Team Formation

This initial step was one of our main fears in attempting to do group project for large classes: 'How will students meet other students to work with?' The answer is a demonstration of the ability of student groups to self-organize when requested to do so by a deadline. It might be a characteristic of biological systems, namely human beings, who naturally do well and can be hugely productive in small groups.

Piazza is our main platform for helping students form groups. At the beginning of the semester, most of our students do not know each other. They have to form a group. How can they do this when there are several hundred teammates to choose from?

After about 3 weeks in the course, we run a structured post in our online forums. Students respond to this post at their own pace by explaining their own background and interests in data analytics in a subpost.

Other students can comment further on these initial responses. Conversations form, and remarkably quickly, small conversations form, discussing their strengths, and hopes for complementarity. They may indicate their time zone or geographic location to ensure the team can communicate well. The discussions go on for a few comments/responses. These discussions are seen by all students in the class.

Surprisingly quickly students team gel together: students agree to work together and groups form.

Once a group of students has formed, they can close the discussion following the initial subpost, so as to indicate to other students that this team is now complete.

Team formation might be one of the most remarkable human factor elements that we've observed in this process. Students have similar motivations; they have a goal and a deadline. They quickly find others with a similar level of commitment and interests. It is great to watch, and it feels like a sort of "speed-dating" exercise.

We set up two deadlines for team formation. The first deadline (soft) provides an incentive for people to decide on their group by a certain date. They still have 1 week after this to change their mind if they choose to. During that week, they can read more posts in Piazza and decide whether they chose well or poorly and would prefer to join another team.

By the second deadline a week later, teams must be formed. Students enter their own names and email addresses in a big online spreadsheet. We request that one team member be named the main communicator of the group, so that we can talk to one person when we must make an appointment with the team.

For the few students who have not found a group by the second deadline, our TA team mixes and matches them. We introduce them to each other and let them talk with each other. Then we 'freeze' the team selection spreadsheet. Changes are not allowed without special permission from course instructors.

Believing in the self-organization of small human groups is important for all this. When properly instructed, with a clear goal and appropriate time frame, small groups of students will find their own way of working together and will generally be pleased with the results for the duration of one semester.

Each semester we get a few teams that do not work well. Typically, this number is in the order of 1 to 2% of students at most. This means that in a cohort of 500 students, we'll have 5 to 10 students are who "floating" and not in a team. We must manage these manually.

Besides the challenge of creating teams, one has the challenge of making sure that the student personalities involved get along on a team. At

times, subgroups of students get formed and some students find themselves excluded of the rest of the team. TAs are coached by the course instructor to depersonalize negative comments and to work on conflict resolution within groups.

We do not allow individual projects. This is a big point. The goal is to learn to collaborate with another data analyst therefore, we do not want "singletons" or teams of 1 student.

TA Assignments for Group Mentoring

Two TAs get assigned to each group. A primary TA who will act as the group's mentor, and a secondary TA who is there as a backup for the primary TA in case the primary TA gets sick or is not available to continue mentoring. The secondary TA also provides a validation or check-up factor to avoid special favors being handed between teams and their mentors.

A primary TA will typically have 6 to 9 teams to mentor. 6 to 7 is good. 8 to 9 can be a lot. It's a function of economics and what your program can offer.

The secondary TA will be primary for another group of 6 to 9 teams. So every TA is primary on 6 to 9 projects and secondary on 6 to 9 projects.

One can stagger the primary and secondary TA assignments to avoid closed circle of information, or let a pair of TAs work on double the number of teams, which makes it easier to communicate for the TAs. The goal is to have good openness and fairness in the process for the students, to ensure that all projects are reviewed and graded, and to not create extra or confusing work for the TAs.

6. GROUP PROJECT DELIVERABLES

There are five deliverables to our group projects:

1. Project Plan Proposal
2. Project Plan Presentation Video
3. Progress Report
4. Final Video Presentation
5. Final Report with code and slides

These are due over an about 8-week period. (See project instructions in Appendix A for details) We've found it is best to have these 5 deliverables – not more and not less.

Project Plan

This is a short document the primarily focuses on a) what is the data students will use, b) what is the main research question (RQ) students want to answer, and c) why is this interesting from a business viewpoint? We are not focused at this

stage on the models that students will use to analyze this data. The main questions are a), b) and c) above.

All 3 aspects of the proposal are important. The business justification c) is key. We want students to explain and justify from a business viewpoint, why is this research question important?

Project Plan Presentation Video

This is a short video, usually 3 minutes only, where a couple (or all) members of the team present in video format their project. It is the "trailer of the movie", the ad for the product. It is a test of whether the team has figured out how to do videos, and how to present them nicely.

All videos for the course are posted as unlisted videos in YouTube. This means that one can see them with a link and they are not searchable in YouTube. Since YouTube is a permanent archive, students now have a permanent link to a work they have done.

Progress Report

This is an intermediate deadline, usually asked after 4-5 weeks to ensure that students are working on their projects and are not leaving all the work to be done in the last 2 weeks of the semester.

Final Presentation Video

This is a KEY deliverable, maybe the most important and most memorable one. 12 and 15 minutes long, with at most 3 minutes of speaking on camera by each team member. Every team member must speak on camera. (Some still hide behind an avatar or image, but we ask them.) This final video is like the presentation students would have been done in class if students had a chance to be all together in a classroom.

This video is housed in YouTube and remain the most vivid memory of their project. If well done, this is the video students will want to show to prospective job interviewer. Since it is in video, it is easy to watch and does not require executing code later on.

Final presentation video are peer-graded online. Peer assignment is done by the platform Peer Feedback. Every student must grade 3 videos, according to a rubric; (Appendix C). With 4 students per team on average, every video will get 12 grades on average and some may get 15.

We've tried non-anonymous grading and found that it led to "grade inflation" due to students being nicer to avoid being criticized. Using

anonymous grading generally yields more authentic grades.

Final Report, Code and Slides

The Final Report is a document turned in the LMS file submission system for group assignments (in Canvas, or using the account of the team communicator for edX). Final reports are 10 to 15 pages long, including graphics. Slide decks usually have 15 to 20 slides. We asked that students turn in their programming code.

Assigning a Github account to each group is a great tool for version tracking and for resolving conflicts when a team member is accused by other team members to have contributed less or not at all. The trail of "GitHub commits" gives a good indication of who checked in code or added new code to the project and who added text to the final report, also stored in GitHub.

Ideally, students code would be placed in a Jupyter notebook or cloud lab environment, with data attached to it. We want a cloud notebook environment that integrates with the gradebook of our LMS. One cloud platform does this: Vocareum. However, at this time, there are limitations on how it can be used, and hence we have not forced student groups to use it.

Ideally all groups would submit project links with their code that work smoothly from the LMS, like playing an unlisted video in YouTube. Anyone with the link or URL would be able to run their code. We are still working on finding the proper environment for this. Also the project confidentiality should be controlled by the team.

7. COURSE OPINION SURVEYS

Course instructor opinion surveys have confirmed that many students appreciated this method of learning. Ratings were generally positive (4.1 out of 5.0). Many comments were positive of the type: "I think the project was helpful to give students a chance to practice on real data the methods of this course."

Some comments were negative due to the fact that projects require a significant amount of work. It is recommended to give only a light final exam with a big group project due. Else you might get comments such: "The project detracted from learning the course material IMO. I would have liked to have internalized more of the concepts and had more time to become intimately familiar with them vs. being distracted by a project that overloaded the end of the semester. Other classes that have projects normally do not have

exams. It's either one or the other for time purposes."

Moral of the story: if you ask students to do a group project that requires a significant amount of effort, lighten up the final exam. I like to maintain a brief (proctored) theoretical final exam and use the group project as the practical or computation part of the final assessment.

Many students understood that the project gave them a chance to show their learning better than by answering multiple-choice questions on tests, and gives students a foundation to do consulting data analytics projects later on.

8. HIGHLIGHTING PROJECTS, EXPO, AWARD CEREMONY

Each TA selects the 2 best projects they have reviewed by comparison with other projects they've seen. Recall that each TA is a primary mentor to about 6 to 9 projects, and a secondary reviewer to a similar number of projects. The course instructor runs an online meeting with all TAs where all projects are discussed. This meeting can be long, yet it is so worthwhile.

The goal is to hear about the best projects and to decide on possible awards or special recognition given to these projects. Think as if one was at a Capstone project expo for on-campus courses. We replicate the project judging dynamics online.

The best projects are highlighted, and their final video is shared with the whole class. Ideally the best projects would be selected to be showcased in online data analytics platforms, such as Kaggle, where the public could see the projects and provide recognition and feedback directly to the students. Some students may not wish to have their projects showcased publicly; (check FERPA).

Having the final presentation video in a public platform such as YouTube is a real benefit for the students who can refer to this piece of work later in job interviews, when asked: 'Do you have examples of your past work in business analytics?' They control its visibility (using an unlisted link). And it is also free marketing for your course among hiring managers and human resources specialists who could refer future students.

9. CONCLUSIONS

This method of handling group projects has proven to be effective and efficient. I hope you

can use it and give it a try in your own courses.

Students perform well on these group projects. It helps them retain more material from the course than if they were asked only multiple-choice tests to assess their mastery of the concepts. Students also gain self-confidence in using the material taught in the course in real situations.

Using the documents provided in the appendices, and with access to the proper LMS and sidekick platforms, course instructors can have a lot of students ready to work on group project in a short time.

One thousand students in one class is a large number, yet it happens in online courses in data analytics, partly because it is a universal language like mathematics, hence there is global demand for such courses. In the future, some online course environments could be structured to handle a million students in an online program or at an online university. (Amazon Web Services training is already at that level...) Several thousand students could take a course at one time and could do group projects this way.

It would be nice to do a longitudinal study of student retention of the material. Typically, student's recall is much better for the project one has done than for any other specific chapter of course, though students may forget years later the details of how they actually got their projects done.

10. ACKNOWLEDGEMENTS

It's a pleasure to acknowledge the collaboration of my teaching assistants who helped conceptualize, write and proofread the students' instructions and rubrics in the appendices, especially Ajinkya Panaskar, Ronak Patel, Evan Jones, Maria Romero-Creel, James Bradley Ashworth, Harsha Andey and Akshay Gangavalli.

The advice of Dr. Polo Chau in the Department of Computer Science at GeorgiaTech (Chau 2018) was a key influence in the development of this technique for business analytics group projects. Dr. Chau teaches a course on Data Visualization for the Online Master in Analytics. His guidance on handling large numbers of students doing computer science projects was a big boost in my confidence to be able to implement a similar teaching style for business analytics projects.

11. REFERENCES

- Astin, A. (1993). What matters in college? Four critical years revisited. San Francisco: Jossey-Bass.
- Bonwell, Charles C.; Eison, James A. (1991) Active Learning: Creating Excitement in the Classroom. *ASHE-ERIC Higher Education Reports*.
- Caruso, H.M., Wooley, A.W. (2008). Harnessing the power of emergent interdependence to promote diverse team collaboration. *Diversity and Groups*. 11, 245-266.
- Chau, Polo (2018 and after). Polo Club of Data Science. <https://poloclub.github.io/>
- CMU#1: Using Group Projects Effectively. Carnegie-Mellon University, Eberly Center, Teaching Excellence & Educational Innovation. <https://www.cmu.edu/teaching/design/teach/design/instructionalstrategies/groupprojects/index.html>
- CMU#2: What are the benefits of group work? Carnegie-Mellon University, Eberly Center, Teaching Excellence & Educational Innovation. <https://www.cmu.edu/teaching/design/teach/design/instructionalstrategies/groupprojects/benefits.html>
- CMU#3: The Benefits of Learning Data Analytics From a Business School. Tepper School of Business, Online MS in Business Analytics. <https://analytics.tepper.cmu.edu/articles/benefits-learning-data-analytics-from-business-school/>
- Experiential Learning article and references therein. https://en.wikipedia.org/wiki/Experiential_learning
- Gartner: White, Andrew (2019) Our Top Data and Analytics Predicts for 2019. https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-2019/
- Hartzel, Kathleen S. & Ozturk, Pinar (2022). Analytics in the Business School: Insights from the Literature, *Journal of Information Systems Education*, Vol. 33 (2), Spring 2022. <https://jise.org/Volume33/n2/JISE2022v33n2pp169-181.pdf>
- Kang, J. W., Holden, E. P., & Yu, Q. (2015). Pillars of Analytics Applied in MS Degree in Information Sciences and Technologies.

- Proceedings of the 16th Annual Conference on Information Technology Education*, Sept 2015, 83-88.
<https://dl.acm.org/doi/pdf/10.1145/2808006.2808028>
- Mannix, E., & Neale, M.A. (2005). What differences make a difference? The promise and reality of diverse teams in organizations. *Psychological Science in the Public Interest* 6(2), 31-55.
- NSSE (2006) National Survey of Student Engagement Report. http://nsse.iub.edu/NSSE_2006_Annual_Report/docs/NSSE_2006_Annual_Report.pdf
- Progressive Education, Children School information website (2016)
<https://www.thechildrenschool.info/progressive-education>
- Progressive Education article in Wikipedia, https://en.wikipedia.org/wiki/Progressive_education and references therein such as Kennedy, Robert: *How Children Learn*, (2019) Thought Company, May 30, <https://www.thoughtco.com/progressive-education-how-children-learn-today-2774713>
- Radovilsky, Zinovy & Hegde, Vishwanath (2022). Contents and Skills of Data Mining Courses in Analytics Programs, *Journal of Information Systems Education* Vol. 33 (2), Spring 2022. <https://jise.org/Volume33/n2/JISE2022v33n2pp182-194.pdf>
- Senge, Peter (1990), *The Fifth Discipline: The Art and Practice of the Learning Organization*, Doubleday (revised 2006)
- Tackels, Devon. Six Benefits of a Collaborative Approach to Analytics and BI. Content Marketing Manager, Sigma. <https://www.sigmacomputing.com/blog/6-benefits-of-a-collaborative-approach-to-analytics-and-bi/>
- Tinto, V. (1987). *Leaving college: Rethinking the causes and cures of student attrition*. Chicago: University of Chicago Press, 1994.
- Wieman, Carl (2021): Active Learning: Optimize for Retention. Jan 22. Video: <https://www.youtube.com/watch?v=xxVxgQJwV7w> and article from Sprouts School: <https://sproutsschools.com/active-learning/>

Appendices and Annexures

APPENDIX A

Online Group Project Instructions

Introduction

Students are asked to produce a data analytics project over an approximately 3-month period in groups of 4 to 5 students. Due to segmentation between course sections and platforms, students will only be able to form groups with their course section and platform. You can form your own group. We will randomly assign you to one if you do not have a group.

In terms of scope, you can think of this project as a Kaggle competition or a detailed Towards Data Science/ Medium page. To get a better idea of what a good project should look like, the professor may show some past projects as examples during his OH session if time permits.

The topic for your project should be chosen from one of the 4 main topics of this course:

1. Statistics/Data Analytics
2. Finance/Stock investments
3. Digital Marketing or Advertising
4. Logistics/Operations Management.

Other topics are possible but will require special permission and are not guaranteed. However, we are flexible. This project will be worth 25% of the total course grade.

The 25% will be spread as follows:

- 3% project proposal (TA reviewed)
- 5% project progress video presentation and progress report (TA reviewed)
 - 3% progress report (TA reviewed)
 - 2% progress video presentation
- 5% final project video presentation (peer-reviewed)
- 12% final project report with R code and slides (TA reviewed)
 - 3% final project report: choice of topic, business justification
 - 3% final project report: understanding of the data, and data wrangling
 - 3% level of programming, code, and modeling

3% overall presentation, graphics, and visuals

Deliverables Overview

The completion of the group project will be organized according to five primary “phases”. Each phase will have its own unique set of deliverables and requirements.

Phase	Description	Deadline (All in Eastern Time)
1	Team Formation Sheet	(Allow 2 weeks to form teams) Teams formed = Day 0
2	Project Proposal	Deadline 1: 3 weeks after Day 0
3	Project Progress Presentation Video (4-5 min)	Deadline 2 : 3 weeks after D1
	Project Progress Report (4-5 pages)	Deadline 2 : 3 weeks after D1
4	Final Presentation Video (10-12 minutes)	Deadline 3: 2 ½ weeks after D2
	Final Report (8-10 pages)	Deadline 4 : 3 weeks after D2
	Final Video Presentation Slides, Code, Data, etc.	Deadline 4 : 3 weeks after D2
5	Peer Review: Out-of-Group Final Video Presentation	Deadline 5 : ½ week after D4
	Peer Review: Within-Group Performance Evaluation	Deadline 5 : ½ week after D4

Notes:

D = Deadline or Day

Total span of time of the project deliverables: 9.5 weeks

Add 2 weeks at the front of the project for team formation

Therefore, count on spending 11 to 12 weeks thinking about and working on your group project.

Project Submissions

Platform 1: Canvas Students: Masters students will be placed in a Canvas group where the primary contact person for the group will submit and upload all submissions or files into the given Canvas assignment module for grading, feedback, etc. No other group member should submit files to these assignments. Only the group leader will submit deliverables.

Platform 2: EdX Students: MicroMasters students in EdX should have the primary contact submit any submissions and files for grading, feedback, etc. and the corresponding grade will then be assigned to the remaining students grading page.

Submitting Data: When submitting the dataset for your project as required in several of our deliverables, you may (depending on the size of your dataset) be unable to directly upload it to Canvas or Edx due to platform limits enforced on the file size of assignment submissions. If you encounter this issue, we will recommend the following alternative acceptable approaches:

- Provide a link to where the dataset is hosted on the web or where you pulled it from.
- Provide a reduced dataset which is based upon a small subset of rows from your full dataset.
- Provide a Dropbox link to download your dataset. Remember that you might have access to free Dropbox accounts from your university, so feel free to take advantage of this.

We also recommend storing these dataset links directly in your group's GitHub repository for easy access. However, we ask that the group DO NOT store datasets on GitHub. Keep these locally and ensure their file endings are added to your .gitignore file.

GitHub

As the semester gets started, you will form or be assigned to a group and each group will be assigned a GitHub repository on `github.your_university.edu` (Canvas) or `github.com` (Edx). You must use this assigned repository for all your project work. We make this requirement for a couple of reasons. Firstly, knowing how to use and be productive with Git/GitHub is an extremely valuable skill in Data Science and Machine Learning. Version control is an exceptionally important technology that makes working on code in groups (or even individually) much more practical. Secondly, GitHub offers the ability to track contributions to a project. Often at the end of a long semester, disputes about group member participation can arise, and to settle these disputes in the most unbiased and fair way possible, we may use GitHub's records of contributions to verify and remedy these claims.

1. **Organization:** When each group's individual GitHub repository is created, you will notice the presence of several folders already in the repository structure. These are to help you stay organized. Although tedious, keeping an organized repository is especially important when working in groups. Do your best to maintain a clear structure for your group's repository. Feel free to alter the directory structure as needed but be sure to keep the deliverable folders at the top level of the repository for clear and easy access by course staff.
2. **README.md:** the "readme" file is the front cover of any GitHub repository, and should contain information about running, installing, and using your project. Be sure to complete this for your project. These readme files are also suitable places to provide information such as your directory structure, project purpose, and goals. For some readme inspiration, see this [repository](#) on GitHub which posts links to some excellent examples. Although we don't expect all of your readmes to be as detailed or creative as soon as these, these examples should give you a good guide as to what to include and how to structure it.

Other Important Considerations

Ownership: Your team's project is work owned by you and your team members, not the course or university. Feel free to share and show off your work, especially to potential employers.

Objectivity: As your team decides on the project topic because this is a business course, it is ok to express business opinions or make recommendations on solutions to the problem you are addressing. However, please base your recommendations on your work of having performed an objective analysis based on facts and science.

Originality: Your project should be unique, and not simply a complete replication of what has been done before. Feel free to stand on the shoulders of past academic work or otherwise, but remember to also add something, take an original approach, or add your own twist on it. Replication is a valuable skill and a necessary part of any research, but for the purposes of this project please avoid direct replication.

Phase 1: Team Formation, Data, & Topic Selection

Team Formation

The guidelines for forming a team are given as follows:

- The work will be carried out in teams of 4 to 5 people. Ideal team size = 4 people. You are welcome to decide on who to team with since each team needs to decide on their topic.
- See **Piazza post @5** for our group project team formation/introduction threads. We highly recommend seeking group members with similar or compatible time zones and work schedules.
- Canvas Section students can **ONLY** sign up for a group with Canvas section students. Similarly, Edx students can **ONLY** sign up for a group with other Edx students. Due to FERPA regulations we are not allowed to have any overlap.
- Some students mistakenly believe that the group projects reduce the amount of work that needs to be graded. Any instructors or TAs know that open-ended questions, and in this case projects with topics chosen by students, need a lot of thinking and time from the graders' end. In fact, an open-topic group project is the hardest thing to scale to large classes. We thought hard about launching a group project in his class. We believe that a group project presents many benefits in a business class about data analytics. Most of all, we hope that several years from now, you will remember this group project as the most memorable assignment you did in this course: Data Analytics in Business.
- Team Formation Signup Deadline: This deadline is the date when the Team Formation Signup sheet will close to responses. This is the deadline by which ALL students should have their group leader record their team information. The sheet should only be filled out by the group leader on behalf of the entire group.
- The only formal delivery in this section is the team leader's completion of entering all group member information into our team signup excel sheet

Choosing Topic for Analysis

You need to justify that the topic is interesting, relevant to the course, and is of suitable difficulty. No formal delivery/submission is associated with this section.

Guiding Questions while choosing a topic:

Here is a list of tasks and questions to think about while you are discussing your project topic as a group internally. These questions and answers do not need to be submitted; they are only provided to help guide your choice of topic:

- What is the problem or objective?
- Why is the problem/objective hard or interesting?
- How is it solved today?
- How do you plan to approach the problem?
- What analysis will you perform on the data?
- What is your (new) technical idea?
- Why can/will this new idea succeed now when it hasn't in the past?
- What is the impact if successful?
- How will your project and your team be organized?
- How will intermediate results be generated?
- How will you measure progress?
- What would it cost if one could implement your idea on a large scale?
- What monetary impact would your project have if its results were implemented widely?

Choosing a Dataset

In the field of data analytics, finding a dataset and settling on a topic are highly related. For this reason, we also mention some specifics here about our dataset requirements which may influence your choice of topic. No formal delivery/submission is associated with this section.

Guiding Questions while choosing a topic:

Dataset selection considerations: Finding a good dataset is a skill we want you to develop through research and investigating repositories and sources online. These considerations and questions should help guide your selections:

1. We recommend utilizing two unique datasets in your project. However, if your single dataset is complex and substantial enough alone, TAs will allow it at their discretion. Finding two compatible different datasets can be challenging, but often creates many more opportunities for unique and detailed analysis.
2. If you need some ideas, please look for data repositories in the document "Dataset Suggestions" provided in the course Canvas or edX site. If the link is not accessible to you, the file will be uploaded in the Project module in Canvas or EdX.
3. It is OK to bring your own data, that you have either collected or that comes from your company. If you have collected data through surveys or your own measurements or record keeping (of stock trading for instance), this is great for your project.
4. If the dataset comes from your company, please make sure you have permission to use it in this school project, where confidentiality may not be guaranteed. And make sure your teammates are equally excited about working on a dataset from your company.
5. Your dataset needs to be substantial enough in size to warrant your analysis and conclusions. We do not have hard limits or requirements when it comes to your dataset's number of features, observations, or sample size. This is because size is very relative to your topic, and we could not enumerate all potential combinations. With that in mind use your best judgment on size, and if you are in doubt, it is likely that your dataset is too small. As a rule of thumb think 1000+ rows.
6. Can Non-trivial analysis/algorithms/computation be reasonably and affordably performed on the dataset given your known limited resources and time (e.g., computing basic statistics, like average, min/max will not be enough)?
7. How "dirty" is the dataset? Does it seem like most of your project will be devoted to cleaning efforts? Although good data cleaning is an extremely important component of any well-rounded project. If your dataset is so unclear that it could completely dominate your workflows and hinder any deep analysis work, then you should probably consider other avenues.
8. Will this dataset be useful/central to answering our project's stated goals?

Phase 2: Project Proposal

Project Proposal

- The project proposal document will serve as the introduction of both your team, topic, and dataset.
- Imagine this document as something you would hand to a team manager as the formal proposal of an internal project your team wishes to pursue.
- The document itself will be a text submission.
- You need to convince us that this project is something interesting and worth working on.

What to Include & What is Required?

- You may use the provided and recommended Proposal template to structure your proposal document
- Include names of group members with brief bios (2-3 sentences each)
- Suggestions for good Research Questions (Ok to suggest 2 or 3)
- Datasets and/or Potential Data Sources (Try to find 2 data sources)
- Description of the dataset and key variables you plan to analyze
- Screenshot of your Dataset and link to the sources where you found it
- Your plan for developing your Models. List of models you will try.
- Your anticipated discoveries or current conclusions

- Answers to the following questions:
 1. What are you trying to do? Articulate your objectives using absolutely no jargon.
 2. What are your datasets? Describe them and the information you plan to extract from them. Include a snapshot of your datasets and a link to the source where you found them.
 3. What are the key variables of the datasets? Which variables will be considered independent and dependent? Are you going to create new variables?
 4. What variables do you hypothesize beforehand to be most important?
 5. What is your approach and plan for training models and optimizing hyper-parameters?
 6. What is your approach to comparing your models?
 7. What tools will be used?
 8. How long do you think it will take?
 9. What are some important milestones/goals to keep your project on track?
 10. What do you think are the expected outcomes of your analysis?
 11. How will progress towards these goals and outcomes be measured?

Submission

- Group Primary Contact will submit your Project Proposal on behalf of the group via the corresponding assignment on Canvas/Edx.
- Please name the submission entry in the following format: **teamXXXproposal.pdf**
- See the "[Submitting Data](#)" section for details on how to submit your dataset.

Grading

- TA feedback /grades on Project Proposal will be available within 10 days (about 1 and a half weeks) after the assignment due date (These may imply substantial changes).

Rubric: See the project proposal template document for a full breakdown of point assignments on the project proposal document.

Phase 3: Proposal Video Presentation & Progress Report

Proposal Video Presentation

What is it?

- A 4–5-minute video presentation (one presentation per team). The video should show your slides (e.g., as pdf/ppt/slides/or similar on your computer screen via screen capture, say using QuickTime, MonoSnap, Screen Recorder etc.) with voice narration; it is up to you whether to show your face. You should be able to create this recording quickly with little effort – no need to do any special video or audio editing although if you feel the need to it is permitted.
- This video presentation is really a primer for students to get familiar and comfortable with creating a proper video presentation of their findings which will be necessary for the final report presentation.
- Think of the presentation as something you would present to a manager who may or may not be familiar with all the details and purpose of the project. Take care to explain everything clearly and provide proper background knowledge.
- It is alright to assume your audience has some familiarity with statistical models.
- The submission itself will be a video presentation 4-5 minutes in length
- Note that this is a PROGRESS video presentation, it is ok if you haven't done everything, yet we don't expect you to show us what you've done so far and what you plan to do.

What to Include & What is Required?

1. 4-5 minutes of video/audio content that presents your group's project's current progress.
2. Team members' names should be listed in the presentation.
3. Speaker should mention who they are before speaking
4. Necessary background information/framing of the problem.
5. Include an overview of the problem in general as well as your planned approach (it is ok if this approach changes later in the project as you learn more information)
6. Any initial hypotheses?
7. What type of models do you plan to use?
8. How has your data cleaning progressed?

9. The video should talk about measurable progress that you have made in terms of the project since the submission of the project proposal document.
10. If you encounter any unexpected problems, challenges, or interesting findings please mention these. Discussion of things that didn't work is also encouraged.
11. You could include things like any new datasets that you have found, any analysis that you have done on the datasets, or any other impactful and measurable progress that you have made thus far.
12. The team should review 2-3 research articles/papers/Kaggle kernels/ etc. related to your problem statement and then discuss them in your progress report slides and video. Be sure to cite the reference at the end of the slide show.
13. If you have any graphs or key visuals to include, be sure to add titles, captions, labels, axes, legends, and most importantly context where necessary!
14. Visuals and Slides are necessary; DO NOT just read your written progress report
15. Include audio narration
16. We do not require all group members to be part of the video given the limited time allocated
17. Sources cited on the last slide (no need to read them just include them).
18. 4-5 minutes is a limited time window to cover all these things but do your best to manage this time wisely and play to your strengths in what you have accomplished so far.

Submission

1. Group Primary Contact will submit your Project Progress Video Presentation on behalf of the group via the corresponding assignment on Canvas/Edx.
2. Upload the video as an [unlisted](#) YouTube video (NOT "private" or "public"). Unlisted videos can be viewed by anyone with the link to your unlisted video.
3. Name your YouTube video with an informative title being sure to include your Team number and "Progress Video Presentation" ex: "[Project name] Team XXX Progress Video Presentation"
4. Add a description to the video that lists a brief overview of the video, your team, team number, and project.
5. Double-check that the URL works, visit that URL using a separate web browser that has been fully logged out of Google services (e.g., all cache cleared, use "Incognito" mode in Chrome, etc.)
6. The submission will be in the form of an unlisted YouTube video URL. Entering that URL will be the only requirement for the assignment submission.

Grading

TA feedback /grades on Project Progress Report Video Presentation will be available within 10 days (about 1 and a half weeks) after the assignment due date (These may imply substantial changes).

Project Progress Report

What is it?

- It will be a detailed description of what you have done so far, preliminary results you obtained, adjustments to be made if any, and the work that lies ahead.
- The document itself will be a text submission similar in format to an academic paper
- Think of the report as a formal progress report to send to a manager who may or may not be totally familiar with all the details and purpose of the project. Take care to explain everything clearly and provide proper background knowledge when necessary.
- It is alright to assume your audience has some familiarity with statistical models.
- Note that this is a PROGRESS report, it is ok if you haven't done everything, yet we don't expect you to show us what you've done so far and what you plan to do.

What to Include & What is Required?

- "Academic-like" paper with dense text inline figures, no direct dense code cells within the paper. Code files related to the project are included in the repository files but are self-contained. Detailed instructions are provided accordingly. No need to submit these files with the progress report, but they will be needed during phase 4.
- Describe in depth the novelties of your approach and your initial discoveries/insights/experiments, etc. and the analysis that is still to be done to conclude your analysis
- Necessary background information/framing of the problem
- Include an overview of the problem in general as well as your planned approach (it is ok if this approach changes later in the project as you learn more information)
- Any initial hypotheses?
- What type of models do you plan to use?

- How has your data cleaning progressed?
- If you encounter any unexpected problems, challenges, or interesting findings please mention these. Discussion of things that didn't work is also encouraged.
- You could include things like any new datasets that you have found, any analysis that you have done on the datasets, or any other impactful and measurable progress that you have made thus far.
- Literature survey of at least 2-3 sources (can be the same ones from video presentation)
- Works cited section.
- Include key visuals in line with text, but always be sure to include labels, axes, captions, legends, and most importantly context!

Submission

- Group Primary Contact will submit your Project Progress Report on behalf of the group via the corresponding assignment on Canvas/Edx.
- Please name the submission entry in the following format: **teamXXXprogressreport.pdf**.

Grading

- TA feedback /grades on Project Progress Report will be available within 10 days (about 1 and a half weeks) after the assignment due date (These may imply substantial changes)
- Rubric: See Appendix B

Phase 4: Final Project Submission

Project Final Video Presentation

What is it?

- The submission itself will be a video presentation 10-12 minutes in length that explains your entire project.
- The video should show your slides (e.g., as pdf/ppt/slides/or similar on your computer screen via screen capture, say using QuickTime, MonoSnap, Screen Recorder etc.) with voice narration; it is up to you whether to show your face. You should be able to create this recording quickly with little effort – no need to do any special video or audio editing although if you feel the need to it is permitted.
- It will cover from start to finish all the key highlights and work you've done for the past few months. It will also effectively summarize your findings and conclusions.
- Imagine that this video would be the result you present to a manager at the culmination of an internal project who may or may not be familiar with all the details and purpose of the project. Take care to explain everything clearly and provide proper background knowledge.
- It is alright to assume your audience has some familiarity with statistical models.

What to Include & What is Required?

- All group members are required to appear in the presentation both by audio and by visual.
- Group members should include their names and introduce themselves before speaking
- Each group member should be featured for a similar amount of time
- Visuals and Slides are necessary; DO NOT just read your final paper/results
- Include audio narration
- We would like each group member to cover a similar length of the video. Obviously exactly equal time would be impractical to the flow of the presentation but try to keep everyone's presentation time similar
- **Overview of Project:**
 - Team members' names listed in the presentation.
 - Necessary background information/framing of the problem.
 - Include an overview of the problem in general as well as your general approach.
 - State initial hypotheses.
- **Overview of Data:**
 - How involved was the cleaning process?
 - What were your key variables?
 - Any interesting insights from EDA?
 - If you used feature engineering how and was it successful?

- Where did the dataset come from?
- Super quick overview of the data.
- **Overview of Modeling:**
 - What type of models did you use?
 - How do they compare?
 - How did you perform model selection?
 - How did you perform hyperparameter optimization?
 - How did the models perform generally speaking?
 - Are they useful and in what ways?
 - Why did you choose those models in particular?
- Include a couple of key visualizations and be sure to include captions, labels, legends, and most importantly context where needed!
- If you encounter any unexpected problems, challenges, or interesting findings please mention these. Discussion of things that didn't work is also encouraged.
- Is there any unfinished business or areas which if given more time or resources you would deem promising or interesting to further pursue?
- Overall conclusion and key takeaways from your project as a closing message.
- Sources cited on the last slide (no need to read them just include them).

Submission

- Group Primary Contact will submit your Project Final Video Presentation on behalf of the group via the corresponding assignment on Canvas/Edx.
- Upload the video as an [unlisted](#) YouTube video (NOT "private" or "public"). Unlisted videos can be viewed by anyone with the link to your unlisted video.
- Name your YouTube video with an informative title being sure to include your Team number and "Final Video Presentation" ex: "[Project name] Team XXX Final Video Presentation"
- Add a description to the video that lists a brief overview of the video, your team, team number, and project.
- Double-check that the URL works, visit that URL using a separate web browser that has been fully logged out of Google services (e.g., all cache cleared, use "Incognito" mode in Chrome, etc.).
- The submission will be in the form of an unlisted YouTube video URL a. Entering that URL will be the only requirement of the assignment submission.

Grading

- The Project Final Video Presentation will be graded by peers outside of your specific group anonymously.
- Each person in the group will review 3 other group submissions and thus receive 3 peer reviews on their group final video presentation. The group will receive $3*n$ total presentation peer reviews (where n is the number of members in the group).
- These scores will then be aggregated by taking the median score across each rubric item, and then summing the result. This score will be the final video presentation score assigned to all your group members.
- All students are required to peer grade their assigned submission and are individually responsible for completing their own reviews. There will be a 30% penalty if you do not complete peer-grading
- Peer feedback on the final video presentations will be made available on completion, but grades will be made available a few days following the peer review deadline.
- No TA feedback/grade will be added for this deliverable
- If we have 4 people per group on average and 600 students. We get 150 project teams, and 150 final videos to peer review. If we assign 3 videos to review for each student, we will have 1,800 "reviews" available. Therefore, each video will be reviewed 12 times ($= 1800/150$), which should provide a good amount of feedback for each video, to ensure some fairness.

Project Final Report, Data, Slides, & Code

What is it?

- It will be a detailed description of what you did, what results you obtained, clear interpretations of the results and what you have learned and/or can conclude from your work.

- All deliverables and work created throughout your project including code, notebooks, reports, slides, etc.
- Imagine this deliverable as the official and final project in its entirety. This would be what you deliver to your client/manager on the completion of a project.

What to Include & What is Required?

- Slide deck used in the final video presentation.
- Text document with a link to the final video presentation.
- "Academic-like" paper with dense text inline figures, no direct code within the paper. Code files related to the project are included in the repository files but are self-contained. Detailed instructions are provided accordingly.
- **Overview of Project:**
 - Team members names listed in heading
 - Necessary background information/framing of the problem
 - Include an overview of the problem in general as well as your general approach
 - State initial hypotheses
- **Overview of Data:**
 - How involved was the cleaning process?
 - What were your key variables?
 - Any interesting insights from EDA?
 - If you used feature engineering how and was it successful?
 - Where did the dataset come from?
 - Super quick overview of the data
- **Overview of Modeling**
 - What type of models did you use and how do they compare?
 - How did you perform model selection and hyperparameter optimization?
 - How did the models perform generally speaking?
 - Are they useful and in what ways?
- Include a couple of key visualizations inline and be sure to include captions, labels, legends, and most importantly context where needed!
- If you encounter any unexpected problems, challenges, or interesting findings please mention these. Discussion of things that didn't work is also encouraged.
- Is there any unfinished business or areas which if given more time or resources you would deem promising or interesting to further pursue?
- Detailed discussion of methodology.
- Detailed discussion and evaluation of results.
- Overall conclusion and key takeaways from your project as a closing message.
- Describe in depth the novelties of your approach and your discoveries/insights/experiments.
- Works cited section.
- **Code Files**
 - Code used to generate the project and accompanying figures.
 - Include a Readme with installation and running instructions as well as an overview of your directory structure and any other important elements.
 - Requirements.txt/.yml file if you are using libraries outside of those used in the course.
 - Any extra documentation/explanation for a user/TA to run the code to see the results or ideally a notebook that shows all the results in a logical sequence
- **Data**
 - See the "[Submitting Data](#)" section for details on how to submit your dataset

Submission

Group Primary Contact will submit your Project Final Report, Data, Slides, & Code on behalf of the group via the corresponding assignment on Canvas/Edx. The easiest way to do this is to download your entire

GitHub repository and upload that resulting zip file as your submission. This cloned repository should contain all the necessary deliverables mentioned above.

Grading

TA feedback/grades on Project Final Report, Data, Slides, & Code will be available prior to the conclusion of the course.

Rubric: See Appendix C

Phase 5: Peer Review & Evaluation

Out-of-Group Final Video Presentation Peer Reviews

What is it?

All group members will individually review up to three other students' final video presentations. Each of these students will be outside your group and should all be from different groups so that all students get to view a variety of unique projects.

Submission

- You will use the Peer Feedback Platform (Canvas) or Microsoft Form (Edx) to score each of your peers' videos. A rubric will be provided for clarity.
- Details on how to use and complete these assignments will be forthcoming when we get closer to the end of the project schedule.

Grading

- Peer reviews are anonymous, yet please be courteous, constructive, and respectful to your reviewees.
- For each of the rubric items, we will take the median score achieved across all the groups scores for that given section. These scores will then be summed to arrive at your final rubric score. This score's percentage will be used for the final grade received for this assignment in the course.

Within-Group Peer Performance Reviews

What is it?

- This is an assignment which requires each student to evaluate the performance of all their group members
- All group members will individually review each of the group members they have worked with this semester.
- We provide these surveys to identify potential problems with work allocation and contribution that may arise throughout the project.
- These reviews are anonymous as we understand that when working closely with someone you may wish to remain anonymous to provide a full and truthful description of their behavior, conduct, and contributions.

Submission

- You will use the Peer Feedback Platform (Canvas) or survey (Edx) to score each of your group member's collective efforts throughout the course of the project.
- Details on how to use and complete these assignments will be forthcoming when we get closer to the end of the project schedule.

Grading

- These peer evaluations will be used to determine individual variations from the team project grade – upward or downward for highly or poorly performing team members, respectively.
- Survey results and anomalies will be identified by TAs manually and compared with GitHub contributions to determine their validity and severity.
- The Professor will receive a report of each of these cases and has sole discretion to apply and enforce deductions or additions in scores where he deems necessary.

Project File Naming Conventions

File names listed below are to be used for the pieces submitted in Canvas.

Phase 1:

- No Required Files.

Phase 2:

- Project Proposal: teamXXXproposal.pdf

Phase 3:

- Project Progress Presentation Video: YouTube Link with the video titled: "[Project name] Team XXX Progress Video Presentation"
- Project Progress Report: teamXXXprogressreport.pdf

Phase 4:

- Final Presentation Video: YouTube Link with the video titled: "[Project name] Team XXX Final Video Presentation"
- Final Presentation Slides: teamXXXfinalslides.pptx
- Final Report/Code: teamXXXfinalreport.pdf

Phase 5:

- No Required Files.

Frequently Asked Questions (FAQs)

Dataset FAQs:

1. Is my dataset big enough?

- If you feel the need to ask, it is probably too small.

2. Is my dataset too big?

- If the physical memory limitations of your local store or teammates are a serious concern, then you should probably downsize. Consider using a subset of the full dataset or doing more extensive cleaning. Please DO NOT spend your hard-earned money on cloud data storage solutions, spark instances, etc.

3. Does our dataset have to be in ".csv" or any other specific format?

- No, data is stored in a variety of ways for a variety of reasons. You do not need to convert data to any specific format on our behalf as long as it is usable in that form for your code.

4. Can we scrape data?

- Yes, but remember that scraping can be tricky and is often time-consuming to do so in a responsible manner. Scraping data could take valuable time away from more detailed modeling and analysis efforts which are the primary areas of focus for the project. Scraped data can also present more challenges for subsequent cleaning efforts so keep these unique challenges in mind.
- If you are scraping remember to be a "courteous" scraper and not overload the servers with requests. Doing this can sometimes be as simple as a "sleep" statement in the loop. Some websites even offer dedicated REST APIs to responsibly handle such requests. These dedicated APIs should always be the preferred scraping method if available so be sure to look out for them and their documentation. Being "courteous" is for your own protection as well as for the well-being of the website. Your university, our course, and teaching staff will not be responsible for any rogue scrapers; you have been duly warned.

Formatting FAQs:

1. What is included in "text format"?

- acceptable text formats are .pdf, .md, and .docx; uncompiled latex (.tex) or plain-text files .txt are not acceptable.

2. Are the time/page ranges hard limits or recommendations?

- Time/page ranges in this document are hard limits. We will often provide the time or page limits for a deliverable usually specified in the following form "(10-12 minutes)". These are all hard limits both minimum and maximum which if exceeded can incur a penalty so be mindful of your length.
- 3. What kind of page formatting is expected?**
- The size of such a page is assumed to be a standard 8.5" x 11" piece of letter paper. The page should have no less than 0.5" margin on all sides and no more than 2" margin on all sides. Line spacing should be in the range of a minimum of single line space to a maximum of double line spacing.
- 4. What counts towards the page limit?**
- For the purposes of the project, we consider all pages containing meaningful and informative content to be included in the limit. As an example, this includes visualizations, but would not include the works cited section or title page. Make good use of your document space. Displaying raw code is rarely if ever helpful in a formal paper or report. If an algorithm is necessary for the paper's content, we recommend using a figure containing the pseudo-code routine.
- 5. What font sizes and styles can we use?**
- We will leave font style up to students, but we recommend it be one that is easily legible. On size restrictions, we ask that students use no less than an 11-point font.
- 6. What citation format should we use?**
- We do not mandate a specific citation format for you to use. Any of the common formats: APA, Chicago, IEEE, MLA, etc will all be fine, just stay consistent in your choice.
- 7. Does an appendix count towards the page count?**
- Typically, an appendix would not count towards the page count as it includes non-dense content which isn't absolutely essential to the paper itself. However, some students try to throw all of their visualizations including ones central to discussion into the appendix hoping to avoid a page limit. This will not be allowed. If you use an appendix properly it will not count towards your page count.
- 8. What is considered "essential" vs. "supplementary" when it comes to appendix content? How do I tell the difference?**
- "Essential" content/figures are anything that would, if omitted, cause serious flow or discussion problems/gaps in your paper.
 - "Supplementary" content/figures are anything that adds greater detail to your findings and report but isn't strictly necessary to the flow and discussion in the paper. If you can remove a figure without impacting the document contents and meaning, then it is probably supplementary material.

Grading FAQs:

1. I received my TA feedback for Phase Y deliverable X, but I still have some questions.

How do I seek clarification

- TA feedback will be marked by the TA that posted it. Feel free to send all inquiries to your primary and secondary TAs via email. We may not see or be able to reply to in-submission comments so please handle these over email or piazza. Be sure to mention who posted the feedback and how it can be clarified, and include a descriptive subject title to ensure it will be read.
- 2. My group collectively does not agree with our grade received on Phase X deliverable Y. How do we go about discussing these doubts?**
- All assignments (except peer-graded ones) are graded by your group's primary and secondary TAs independently. These scores are then averaged together to arrive at your overall score for a given deliverable. It is unlikely that we will make additional changes to a grade given the redundancy embedded already in the grading process. However, if your group is set on challenging the grade you may appeal to the lead TA of your platform for a regrade. In this regrade request which should be made via email, you should include a detailed explanation for why and where the previous grading is wrong. Also, include a descriptive subject title to ensure your email will be

read. Your group's arguments will then be evaluated by the lead TAs with consultation from the professor if needed. They will then adjust the grades as necessary. Be aware that the lead TA may conduct a full regrade in order to resolve the dispute. Similar to homework regrade requests, this may or may not be in your best interest. Please consider any and all regrade requests carefully. Email addresses for the lead platform TAs are as follows:...

3. How should I seek advice from my primary/secondary TA?

- When seeking advice about some aspect of your project from your assigned primary and secondary TAs, try and keep your questions as concise and direct as possible. They will do their best to answer them appropriately, but they will not "pre-grade" your work. In the past some students have emailed our TAs entire reports and asked for feedback, this will not be acceptable, and will likely not receive a reply. However, if you do have genuine questions do not hesitate to ask. We are able and ready to help.

4. I see that some of the project deliverables have been broken down further into more specific grade % allocations. Can we expect to see this additional breakdown on our end when grades are posted?

- No, we will not provide a component-specific breakdown. This point allocation is for our internal uses but is provided to you in this document, so you have a better understanding of where and by how much each individual deliverable in a given phase is weighted.

Modeling FAQs:

1. Are we restricted to using only modeling approaches covered in the course content?

- Although we encourage students to apply the modeling concepts we've covered so far in the course, there may be many better modeling approaches available for your given project task. Using these other modeling approaches is fine, but we recommend not venturing too far out into the unknown. We typically recommend confining your approaches to ones covered in the course or within the ISLR or ESL books. However, if one or more of your group members has a lot of experience and knowledge of a framework outside of this scope feel free to leverage that experience. Just be sure to explain it a bit more than you would a model covered within the course or its resources.

2. Can we use auto-ml tools to build our models?

- Although very handy we will not allow students to submit auto-ml created models as it would remove much of the learning experience associated with the project. If you would like to use auto-ml to do some exploratory modeling and hyper-parameter searching in a more centralized automated fashion to get a general idea of what approaches are good candidates for more detailed more manual subsequent modeling efforts that is fine. The use of auto-ml beyond that capacity will not be acceptable. For that reason, we generally discourage its usage.

Other Assorted Common FAQs:

1. Can I enter my project in a Kaggle competition/ Can my project be based on an open Kaggle competition?

- Yes, on both fronts. Kaggle is a great platform to see some interesting projects and work on datasets that regular people would typically not have access to. Kaggle is also a hunting ground for recruiters who often contact top medal-achieving winners. However, that being said, the goals of Kaggle and the goals of our project differ in a subtle, but important way. Kaggle uses overall accuracy or performance metric scores to determine a leaderboard and winner of each competition which causes the platform to focus on achieving often miniscule marginal improvements to accuracy rates. Although there is certainly skill involved in this and it is by no means easy, this aspect of the competition doesn't capture the true nature of a real-world analytics project which is far more focused on solid understanding and analysis as opposed to final accuracy scores. Again, achieving good accuracy is important in modeling, but it certainly isn't everything so if you do decide to put your project on Kaggle, be certain that you still maintain a heavy focus on analysis and explanation in your project deliverables.

2. Will we get a chance to see other completed projects?

- The professor may (as he has in years past) show several examples of past projects during his office hours sessions. These may help your group get a better understanding of what your deliverables should look like and cover from a content perspective.

3. Can you check if my submission went through?

- No, please do this yourself by verifying the submission via the existing tools for that on your given platform (Canvas/Edx)

Peer Feedback FAQs:

1. How picky will graders be about the time limit on the videos? Are we going for an exact time or will a couple of seconds over be alright?

- a couple of seconds over won't cause any problems or incur any penalty, but do try to keep as close as possible to these set time limits.

2. What is PeerFeedback? When will we receive more instructions on how to use it?

- PeerFeedback is a peer review platform maintained by a developers team. The platform was created to fill a gap in existing peer review platforms that allow for Canvas integration and increased functionality beyond Canvas' native capabilities. Instructions on how to use the platform will be provided when we get close to the final project deliverables.

3. Will Edx students use PeerFeedback?

- No, unfortunately, Edx does not integrate with the PeerFeedback platform yet. In lieu of using that platform for peer reviews, Edx students will use a Microsoft Forms survey. Similar to the PeerFeedback platform instructions, they will be announced closer to the delivery of the final report.

4. I received my peer feedback for Phase Y deliverable X and I do not agree with it. What can I do?

- Changing or altering a peerfeedback comment or score would defeat the purpose of peer grading so we will not alter these scores (unless there was a retraction initiated by the peer grader's end or a technical problem that resulted in an inaccurate score). Each student receives multiple peer reviews, and the median (an outlier robust measure) is used to reduce the impact felt by outliers so a single bad review will not tank your score.

5. I accidentally assigned the wrong score during the peer review process. Can I go back and change it?

- Unfortunately, neither our survey platform nor PeerFeedback platform will allow the resubmission or revision of assigned peer scores. For this reason, we encourage students to be exceptionally diligent when entering these grades. If an honest error has been made that sufficiently impacts a peer score assigned, let our course staff know via a private piazza post with just instructors, and we will do our best to correct it.

Proposal Document FAQs:

1. For the 2-3 research questions, could they all be about different things? Or do they have to follow a central theme?

- Research questions should all share a generally common theme, but each should share its own unique perspective in describing the common project theme/question or a perspective/aspect of it. We would not advise straying too far from the central theme, but if you have a good way to tie that question or research in it would be fine. We will caution you here to not try and expand your scope overly far.

2. Do I have to use the proposal template provided?

- No, you are free to deviate from it, but I would HIGHLY recommend using the template as it ensures clear sections and uniform grading by TAs where all rubric items are clearly laid out. Using the template is the best way to ensure your group does not miss out on any points.

Teaming FAQs:

1. Why can't I work alone?

- Most large-scale data analysis projects in the industry are team-based; Many former students found the projects highly beneficial to them. If you strongly desire to do solo projects, this course unfortunately is not a good fit for you.

2. What do I do if my group isn't working out?

- If you encounter any problems with your teammates that are causing your team great difficulty in performing the work at hand, please contact us in a private post with just instructors on Piazza or email your group's primary and secondary TAs as soon as you can so that we can assist and help to resolve the situation. If you fail to inform us of any issues, and they prevent you from delivering work that is wholly complete or on time, we may not be as flexible regarding your grade had you contacted us sooner!

3. What happens if my group members drop the course?

- If after losing a group member your team still remains above the three-person minimum threshold, they will continue in their current form. Teaching staff will make note of this and take it into consideration when grading deliverables, but the deliverable requirements and schedule will remain the same.
- If your team drops below the three-person threshold, alter TA staff immediately, and they will help remedy the situation.

4. If I drop the course what will happen to my group members?

- Your group members will be down a member, but as long as they remain above the three-person threshold, they will continue in their current form. Teaching staff will make note of this and take it into consideration when grading deliverables, but the deliverable requirements and schedule will remain the same.
- If by leaving your team, the number of members will be below the three-person threshold, the TA staff will take special care to remedy the situation.

5. I'm considering dropping the course. What should I do?

- We cannot make any recommendations if you should stay in the course or not; That decision is best left up to you and your academic advisor's council. However, if you are contemplating dropping the course, we highly recommend you tell your teammates this as soon as possible as a professional courtesy to them.

6. If I have a question about my project, who should I contact?

- Once they have been assigned your group should relay all project-specific questions to your assigned Primary and Secondary TAs. For general questions, you may continue to use public piazza threads.
- See our page under the Group project module titled "Role of the Primary & Secondary TAs"

7. Our group was not able to submit our X phase assignment because group mate Y did not complete the things they promised to.

- At the end of the day, groups need to be independently responsible for ensuring that they are functioning properly and working as a team. In meetings, teams should share and verify the work of their other teammates so situations like this don't occur at the last minute. Unless there are valid and documented extenuating circumstances, we will not step in to help here.

8. Group member X has not been in contact with our group for a long period of time. What should we do?

- If you have group members that have "ghosted" the project for periods greater than 2-3 weeks, contact the lead platform TA immediately via email. Include a description of the situation and an informative subject line to ensure the email is read. Be sure to let us know who the individual is, how long they have been inactive, what if any work they have done to date, and how your group typically communicates. Email addresses for the lead platform TAs are as follows: ...

9. One of our group members has a known conflict or period of unavailability. How do we handle this?

- This is an inner-group problem and should be resolved as such. The group has plenty of levers to pull to reallocate workloads at different phases of the project. If such a situation arises where a group member is uncomfortable explaining a known or expected period of absence to the rest of the group, they may reach out to lead platform TAs via email and we will assist in communicating this in a private and respectful manner. Email addresses for the lead platform TAs are as follows:
...

Technology FAQs:

1. Do we really have to commit everything we do to GitHub?

- We highly recommend that you commit everything to GitHub. The reason being is that you want to make sure that you are credited with all contributions that you make to the project. In the event that there is a dispute over group member contributions, we will use the data on member-wise contributions provided by GitHub. So if you want to be sure you get credit for your work, commit to Git often.

2. How do I use GitHub?

- For GitHub learning recommendations see our page under the Group project module.

3. If GitHub is required for the project, why isn't it taught in the course?

- One of the continual challenges for graduate programs, and in the field of data science/computer science in general is the sheer number of platforms, languages, and technologies. Most of these technologies do not require expert-level proficiency to be useful and effective. However, they are still new platforms which all have a learning curve. In this class, we do not cover the use of GitHub strictly because there is not enough time between our four distinct units. Despite this, we believe that students should quickly be able to get up to functional speed in GitHub with only a modest effort. We also believe that being able to quickly adapt and get productive on essential platforms like GitHub quickly is a great skill to develop for your career.

4. Is it ok for us to use multiple branches in our GitHub repository?

- Yes, using multiple branches is a fantastic way to separate workflows temporarily with the ultimate goal of merging these individual branches into one when the time is right. With that in mind, you can use branching all you like, but when it comes time for the final deliverable please try and have all "dead" branches removed or merged into the main branch. This will help keep your repo organized and make it very clear to the grading TA what the final deliverable is.

5. I'm not great at GitHub so is it alright if group member X commits for me?

- No, you will not receive credit for those commits so this is not advisable. We promise that GitHub is not too difficult to get productive in quickly. Please try and learn the basics and commit your own work. If you do not commit your own work we will not be able to help you if disputes arise in the future over contributions and participation.

6. What programming languages can we use?

- Although we cover only R programming in this course, we will allow students to use the language of their choice. However, we HIGHLY recommend students stick with R as it has more statistically focused packages and if your group runs into trouble our TA staff will be more likely/able to provide assistance.

7. What platform should my group use for X, Y or Z?

- For technology platform recommendations see our page under the Group project module titled "Recommended Technologies".
- If your given use/problem case isn't listed there drop us a piazza post and we may be able to help.

Video/Presentation FAQs:

1. Are we allowed to edit our video presentations?

- Although we encourage you to not spend too much time on editing these recordings (as that time is better spent on the analysis) we will not penalize you for going the extra mile.

2. Is the time/page ranges hard limits or recommendations?

- Time/page ranges in this document are hard limits. We will often provide the time or page limits for a deliverable usually specified in the following form "(10-12 minutes)". These are all hard limits both minimum and maximum which if exceeded can incur a penalty so be mindful of your length.

3. Do I need to submit our group's slide deck/presentation with the Phase 3 Project Progress Report Video submission?

- No, we only require the youtube link.

*You've made it to the end of the document.
Congratulations, your dedication is truly commendable!*

APPENDIX B
Managing Data Analytics Group Projects in Large Online Courses

Group Project Proposal Template

Please edit the following template to record your responses and provide details on your project plan.

TEAM INFORMATION (1 point)

Team #:

Team Members:

1. Team Member 1 Name; GT Id (OMSA) or EdX username (MM)
2. Team Member 2 Name; GT Id or EdX username
3. Team Member 3 Name; GT Id or EdX username
4. Team Member 4 Name; GT Id or EdX username
5. Team Member 5 Name; GT Id or EdX username

Insert background information: Name, professional background, education background, previous analytics related projects you have worked on, for every team member!

OBJECTIVE/PROBLEM (5 points)

Project Title:

Background Information on chosen project topic:

Problem Statement (clear and concise statement explaining purpose of your analysis and investigation):

State your Primary Research Question (RQ):

Add some possible Supporting Research Questions (2-4 RQs that support problem statement):

- 1.
- 2.

Business Justification: (Why is this problem interesting to solve from a business viewpoint? Try to quantify the financial, marketing or operational aspects and implications of this problem, as if you were running a company, non-profit organization, city or government that is encountering this problem.)

DATASET/PLAN FOR DATA (4 points)

Data Sources (links, attachments, etc.):

Data Description (describe each of your data sources, include screenshots of a few rows of data):

Key Variables: (which ones will be considered independent and dependent? Are you going to create new variables? What variables do you hypothesize beforehand to be most important?)

APPROACH/METHODOLOGY (8 points)

Planned Approach (In paragraph(s), describe the approach you will take and what are the models you will try to use? Mention any data transformations that would need to happen. How do you plan to compare your models? How do you plan to train and optimize your model hyper-parameters?)

Anticipated Conclusions/Hypothesis (What results do you expect? How will your approach lead you to determining the final conclusion of your analysis?) Note: At the end of the project, you do not have to be correct or have acceptable accuracy, the purpose is to walk us through an analysis that gives the reader insight into the conclusion regarding your objective/problem statement

What business decisions will be impacted by the results of your analysis? What could be some benefits?

PROJECT TIMELINE/PLANNING (2 points)

Project Timeline/Mention, key dates you hope to achieve certain milestones by:

Appendix (any preliminary figures or charts that you would like to include):

APPENDIX C
Managing Data Analytics Group Projects in Large Online Courses

Group Project Rubrics : Proposal Video Presentation and Progress Report

Progress Video is worth 2% of your total grade
Progress Presentation Video + Progress Report = 5%

Section	1 point	2 points	3 points	4 points	5 points
<p>Presentation: The video presentation skills of the team. Is the presentation style clear and concise? Does the flow of the whole video make sense?</p>	The presentation is disorganized, not clear, and not easy to follow.	I can see some effort was put in, but overall, it is not good.	Presentation is clear 60% of the time, and concise to follow, yet there are gaps in the analysis that leave uncertain conclusions.	The presentation is good and clear at least 80% of the time.	The presentation has a good pace. The problem statement is clear and concise and the analysis flows and provides good insight into the research questions proposed.
<p>Content: The slides used to present the video. Does the group present ample figures/visualization to make the video more comprehensible? Do the slides do the job of presenting the group's work in a clear and concise manner?</p>	Slides are not easy to read, and visuals are not helpful in understanding the project and analysis.	I can see some effort was put in, but overall, it is insufficient for a project at this level.	Slides and visuals are good 60% of the time and relate to the analysis but no interpretation or walkthrough was provided.	The presentation is good and clear at least 80% of the time.	Slides are easy to follow, and visuals are clear to see and relate to the analysis.

Business Analytics

Progress Report Rubric

**Progress Report is worth 3% of your total grade
 Progress Presentation Video + Progress Report = 5%**

Section	0 point	2 points	4 points	5 points
Choice of Topic, Business Justification, and Problem Statement	<ul style="list-style-type: none"> No progress + not clear 	<ul style="list-style-type: none"> Choice of topic and objective not strong enough to lead to a good flow of analysis 	<ul style="list-style-type: none"> Good understanding and communication of the topic and objective 	<ul style="list-style-type: none"> Excellent understanding and communication of the topic and objective and convincing argument for the business impact
Understanding of the data and data wrangling	<ul style="list-style-type: none"> No progress + not clear 	<ul style="list-style-type: none"> Dataset is decent and basic to extract insight out of 	<ul style="list-style-type: none"> Dataset clearly explained and cleaned. If applicable, data cleaned and transformed for proper analysis. 	<ul style="list-style-type: none"> Multiple complex datasets used and explained in depth. If applicable, data cleaned and transformed for proper analysis.

APPENDIX D
Managing Data Analytics Group Projects in Large Online Courses

Group Project Rubrics : Final Video Presentation and Final Report

Prof. Frederic Bien
fvbien@gatech.edu
Scheller College of Business
Georgia Institute of Technology
Atlanta, GA 30308, USA

Final Video Rubric for Peer Grading

Final Video is worth 5% of your grade
Final Video (5%) + Final Report (12%)= 17%

Section	1 point	2 points	3 points	4 points	5 points
Presentation: The video presentation skills of the team. Is the presentation style clear and concise? Does the flow of the whole video make sense?	The presentation is disorganized, not clear, and not easy to follow.	I can see some effort was put in, but overall, it is not good.	Presentation is clear 60% of the time, and concise to follow, yet there are gaps in the analysis that leave uncertain conclusions.	The presentation is good and clear at least 80% of the time.	The presentation looks professionally done. The problem statement is clear and concise and the analysis flows and provides perfect insight into the research questions proposed.

<p>Content: The slides used to present the video. Does the group present ample figures/visualizations to make the video more comprehensible? Do the slides do the job of presenting the group's work in a clear and concise manner?</p>	<p>Slides are not easy to read, and visuals are not helpful in understanding the project and analysis.</p>	<p>I can see some effort was put in, but overall, it is insufficient for a project at this level.</p>	<p>Slides and visuals are good 60% of the time and relate to the analysis but no interpretation or walkthrough was provided.</p>	<p>The presentation is good and clear at least 80% of the time.</p>	<p>Slides are easy to follow, and visuals are clear to see and relate to the analysis.</p>
<p>Approach: The approach used in this project. Is the approach taken by the group to arrive at conclusions coherent? Is the project missing any significant details or justification for the choices made in the approach?</p>	<p>Approach does not make sense and significant details are left out that make it hard to justify the conclusion.</p>	<p>Approach is ok but it is explained poorly or not justified properly.</p>	<p>Approach explained concisely but proper justification not elaborated upon for methodology.</p>	<p>The presentation is good and clear at least 80% of the time.</p>	<p>Approach explained clearly and concisely. Justification for approach explained and makes sense.</p>
<p>Results and conclusion: Do the results presented make sense? Are the conclusions drawn from the project in line with the objective of the project? Given the work done in the project, are the results justified?</p>	<p>Results and conclusion not provided.</p>	<p>Results and conclusion are mentioned but they do not look/sound very professional.</p>	<p>Results shown but interpretation and conclusion do not make sense or align with the results.</p>	<p>Results and conclusion are good and clear at an 80% quality level.</p>	<p>Results and conclusion are clear to follow and align with the statistical results.</p>

Final Report Rubric for TA Grading

Final Report with programming code and slides are worth 12% of your grade
Final Video (5%) + Final Report (12%) = 17%

- 3% final project report: choice of topic, business justification
- 3% final project report: understanding of the data, and data wrangling
- 3% level of programming, code, and modeling
- 3% overall presentation, graphics, and visuals

Section	0 point	2 points	4 points	5 points
Choice of Topic, Business Justification, and Problem Statement	<ul style="list-style-type: none"> ○ Background information and objective not provided 	<ul style="list-style-type: none"> ○ Background information not provided to give the reader context into the choice of topic ○ Objective and problem statement unclear before diving into the analysis 	<ul style="list-style-type: none"> ○ Background information about the topic given to the reader for context ○ Objective statement is clear 	<ul style="list-style-type: none"> ○ Background information about the topic given to the reader for context ○ Objective and Business Justification clearly stated
Understanding of the data and data wrangling	<ul style="list-style-type: none"> ○ No mention of the data and datasets to provide context before the analysis 	<ul style="list-style-type: none"> ○ Dataset is not complex enough for the rigor of the project ○ Data not cleaned or transformed upon for proper analysis 	<ul style="list-style-type: none"> ○ Sources of the data cited ○ Explanation of the dataset(s) explained but not in enough detail ○ If necessary, proper transformation of the data executed 	<ul style="list-style-type: none"> ○ Sources of the data explained, and datasets are complex enough for the rigor of the project ○ Explanation of the dataset(s) provided ○ If necessary, proper transformation of the data executed
Approach/Methodology, Level of R code & Programming	<ul style="list-style-type: none"> ○ No analysis or approach documented ○ No supplemental code submitted 	<ul style="list-style-type: none"> ○ Approach and methodology are not easy to follow and not documented well ○ Data models do not make sense with the data provided ○ Level of R code is elementary and does not show effort in analysis 	<ul style="list-style-type: none"> ○ Approach and methodology documented but not easy to follow ○ Data models explained and modeled in R but does not flow well with the methodology ○ R code and programming completed with diligent analysis 	<ul style="list-style-type: none"> ○ Approach and methodology make sense and easy to follow ○ Data models explained and modeled in R to help analyze the objective and supporting research questions ○ R code and programming completed at advanced level and runs with no errors

Project timeline, Overall presentation, Graphics, Visuals	○ Results and conclusion not provided	○ Conclusion stated not clear and does not make sense with the results ○ Interpretation not provided well enough for the reader to understand ○ Graphics and visuals are not legible	○ Conclusion not stated clearly and left up to the reader ○ Interpretation of key variables or results not explained ○ Graphics are legible	○ Conclusion and results clearly displayed and stated and align with statistical results ○ Graphics and visuals are appropriate to the flow of the analysis and are legible ○ Graphics and visuals are legible
--	---------------------------------------	--	---	--