Georgia Institute of Technology

MGT 6203: Data Analytics in Business

Georgia Tech - OMSA Program & EdX MicroMaster

Group Project Guidelines

Georgia Institute of Technology

Table of Contents:

Georgia Institute of Technology

Introduction

Students will be asked to produce a data analytics project over an approximately 2-month period in groups of 4 to 5 students; Due to segmentation between sections of Canvas, students will only be able to form groups with Canvas students. Similarly, Edx students will only be able to create groups with other Edx students. 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)
- 2% project proposal video (TA reviewed)
- 3% progress report (TA reviewed)
- 5% final project video presentation (peer-reviewed, non-anonymous)
- 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

Georgia Institute of Technology

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	June 4, 11:59 pm
2	Project Proposal	June 21, 11:59 pm
	Project Proposal Presentation Video (4-5 min)	July 5, 11:59 pm
3	Project Progress Report (4-5 pages)	July 5, 11:59 pm
4	Final Report (8-10 pages)	July 20, 11:59 pm
	Final Presentation Video (10-12 minutes)	July 23, 11:59 pm
	Final Video Presentation Slides, Code, Data, etc.	July 23, 11:59 pm
5	Peer Review: Out-of-Group Final Video Presentation	July 28th 11:59 pm
	Peer Review: Within-Group Performance Evaluation	July 28th 11:59 pm

Georgia Institute of Technology

Project Submissions

Canvas Students: GT OMSA students will be placed in a Canvas group. These groups will link all member's submissions for the group assignments. The last submission before the deadline received by any group member will be the final submission so be sure to coordinate that the proper version is the last submitted.

EdX Students: MicroMasters students 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. GT students remember that you have access to free Dropbox accounts 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.

Georgia Institute of Technology

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.gatech.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.

Georgia Institute of Technology

Other Important Considerations

Ownership: Your team's project is work owned by you and your team members, not the course or Georgia Tech. 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.

Georgia Institute of Technology

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 unclean 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?

Georgia Institute of Technology

Phase 2: Project Proposal and Proposal Video Presentation

1. Project Proposal

What is it?

- The project proposal document will serve as the introduction of both your team, topic, and dataset.
- This document should be no longer than 2.5 pages
- 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
- Your submission should be no longer than 2.5 pages in length
- 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 7 days 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.

2. Project 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

What to Include & What is Required?

1. 4-5 minutes of video/audio content that presents your group's project proposal
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. If you encounter any unexpected problems, challenges, or interesting findings please mention these. Discussion of things that didn't work is also encouraged.
10. 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.
11. The team should review 2-3 research articles/papers/Kaggle kernels/ etc. related to your problem statement and then discuss them.. Be sure to cite the reference at the end of the slide show or visual medium.
12. 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!
13. Visuals and Slides are necessary; DO NOT just read your written project proposal
14. Include audio narration
15. We do not require all group members to be part of the video given the limited time allocated
16. Sources cited on the last slide (no need to read them just include them).
17. 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 and present a clear idea of your project

Submission

1. Group Primary Contact will submit your Project Proposal 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 Proposal 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 Proposal 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).

Georgia Institute of Technology

Phase 3: 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 the 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 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
- 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 Coming Soon

Georgia Institute of Technology

Phase 4: Final Project Submission

1. 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 at minimum and preferably by visual as well (however visual is still optional).
- 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 non-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 632 students. We get 158 project teams, and 158 final videos to peer review. If we assign 3 videos to review for each student, we will have 1,896 "reviews" available. Therefore, each video will be reviewed 12 times (= 1896/158), which should provide a good amount of feedback for each video, to ensure some fairness…

2. Project Final Report, Data, & 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, 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?

- "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.
- Any relevant code files
- Key visualizations or other supplementary files
- Readme/Documentation for code

- **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 Coming Soon…

Georgia Institute of Technology

Phase 5: Peer Review & Evaluation

1. 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

- Non-anonymous peer reviews so remember to 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.

2. 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.
- The survey is conducted anonymously unless a student compromises their own identity within their feedback comment (ex using their own name)
- 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.

Georgia Institute of Technology

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
- Project Proposal Presentation Video: YouTube Link with the video titled: "[Project name] Team XXX Proposal Video Presentation"

Phase 3:
- Project Progress Report: teamXXXprogressreport.pdf

Phase 4:
- Final Presentation Video: YouTube Link with the video titled: "[Project name] Team XXX Final Video Presentation"
- Final Report:: teamXXXfinalreport.pdf

Phase 5:
- No Required Files.

Georgia Institute of Technology

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. Georgia Tech, 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 insubmission 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:
     i. Evan Jones (Canvas Section): evanjones@gatech.edu
     ii. Maria Romero-Creel (edX Section): mfrc3@gatech.edu

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. Last semester we also successfully piloted a website which publicly displays student's projects for the world to see (with their expressed permission of course). We will likely launch a similar or expanded site this year as well.

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 Georgia Tech developers. 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
     1. Evan Jones (Canvas Section): evanjones@gatech.edu
     2. Ronak Patel (edX Section): mfrc3@gatech.edu

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
  1. Evan Jones (Canvas Section): evanjones@gatech.edu
  2. Ronak Patel (edX Section): mfrc3@gatech.edu

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!