# DAT 205 Data Science Capstone Course

Fall 2020

**Case Assignment Instructions**

Student must form teams (2-3 students per team) as no individual work is permitted. Working as a collaborative team promotes both academic achievement and collaborative abilities, therefore students should learn to inquire, share ideas, clarify differences, problem-solve, and construct new understandings with the course content.

**Case Overview**

For this semester, we have created a partnership with Riipen (www.riipen.com) affirm that connects businesses and other organizations with students who can assist them with their real challenges. We have selected a number of projects, which have been submitted by several organizations. Each team will select a topic of interest from the curated list provided.

Each team has the option to pick a case topic. The individual cases will differ, and may involve design, implementation, and analysis, but it should represent some tangible insights/recommendation for the organization. For example, developing novel predictive analyses and reporting capability from historical outcomes using an existing data warehouse, would represent a significant development that is in scope of this project.

All of the deliverables outlined below must be submitted regardless which option you choose.

**Case Deliverables:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Evaluation Item** | **Weight** | **Due Date** | **Submission** |
| **Team Case Proposal** | **20%** | **Week 4** | **Online - A2L** |
| **Final Plan** | **10%** | **Week 6** | **Online - A2L** |
| **Interim Presentation** | **10%** | **Week 7** | **Online – A2L** |
| **Final Report** | **20%** | **Week 11** | **Online - A2L** |
| **Final "Boardroom" Presentation** | **20%** | **Week 13** | **Online – A2L** |

**EVALUATION AND GRADING**:

The following criteria will apply to the grading of many aspects of your case.

**Grade A -** Work that demonstrate a strong understanding of the various analytics approaches used to gain actionable insights. Students demonstrate their ability to effectively communicate results of data analyses. The student’s work goes beyond the task and contains additional, unexpected or outstanding features.

**Grade B -** Work that demonstrate a reasonable understanding of the various analytics approaches used to gain actionable insights. Students meets minimal criteria for report and presentation, with few errors or omissions.

**Grade C -** Work that fails to demonstrate project requirements and work may not address one or more criteria or may not accomplish what was stated in the proposal.

**Grade F -** Work that is incomplete, inappropriate and/or shows little or no comprehension of the project requirements.

1. Team Formation

By the end of class on Week 1 we should have all teams formed.

A team charter is recommended to outline ways of communicating/working, but a charter is not required to be submitted.

Student names and e-mails must be submitted to your instructor.

1. Team Proposal One Pager (1-2 pages MAX)

Format: *1-page, 11-point font, 1-inch margins, 1 line space, Word format.*

This proposal should reflect the business and Data/Business Analytics goals. Consider the items below that will be due in future weeks. Each group will discuss directly with the instructor in class (Week 2) to get feedback on topics. Try your best to find your data source.

Hint - A good starting point is typically with looking at data options. During a project proposal you will face the challenge of limited understanding of your data, but the most challenging part is finding a dataset, therefore please checkout the below websites. As discussed in class, Kaggle is a free service that has a bunch of large datasets that you can leverage to form your proposal. Any Open Data website or the Teradata website is another good source, but they do not have as many options. Regardless of where you find your data (private or public), please source the website/source and provide a link as backup. Raw data must be used for your project. You are not to download a completed dataset that has already been analysed and reported out on. Plagiarism will be taken seriously and we will check all case deliverables to ensure there is no academic dishonesty.

**Recommended Websites to Consider for Data Downloads**:

<https://www.kaggle.com/>

<https://www.toronto.ca/city-government/data-research-maps/open-data/>

https://www.data.gov

https://www.healthdata.gov

http://www.dol.gov/open/data.htm

https://nycopendata.socrata .com/

http://www.gsa.gov/portal/content/181595

http://open.canada.ca/en

http://www.statcan.gc.ca/eng/rdc/data

http://climate.weather.gc.ca/

http://archive.ics.uci.edu/ml/

http://githubarchive.org

<http://www.crowdflower.com/data-for-everyone>

MIT ICU Open Source Database: MIMIC- https://mimic.physionet.org/

1. Team Proposal Report (4 pages MAX)

Format: *4-pages, 11-point font, 1-inch margins, 1 line space, Word format.*

Submit proposal for review and approval by Instructor. If you are trying to decide between multiple proposals, you may submit more than one abstract (limited to 3 abstracts). Please work in the feedback from week two from the Instructor. The proposal should include some of the following components, but please do not use this as a checklist/table of contents…get creative:

1. Suitable Title - Should reflect the business and data mining goals
2. Business Goal:

* Brief description of the company/context
* Who is the stakeholder/client?
* A description of the business challenge/opportunity. What are the business benefits of implementing any of your proposed recommendations? What opportunity is it creating? What shortcoming does it address? Social/human/environmental implications?
* What would be considered success?

1. Analytics/Data Mining Goal: A description of the analytics work and objective.

* Is it predictive or descriptive?
* Is it retrospective or forward-looking?
* What is the main outcome variable(s)?

1. Data: Brief description of available data.
2. Some guidance on the data subset that will be used and the re-processing or preparation that might be needed based on your past experience.
3. Sample of ten rows (records) with ten columns (variables) that will be used, including the outcome column.
4. What are some data mining methods to consider?
5. Which performance measures are appropriate? How do they map to the business goal?

5. Implementation/Production:

* Operational requirements or constraints (who exactly will use your model, data or system and how?
* Will the solution run in real-time?
* Will it require collecting new data?
* One-time analysis or ongoing?

6. Backup

* Submit a draft bibliography/list of websites, data resources, etc.
* Submit list of mentors, or requested mentors

Submission will be due online or via e-mail to the Professor. TBD

1. Final Plan

Format: *1-2-pages, 11-point font, 1-inch margins, 1 line space, Word format.*

During your proposed Project Planning teams may find that their project idea and work process is not in scope of this project, whereas other teams may have multiple project/datasets that they are considering. This final project plan provides a more concrete plan of attack, following your initial investigation.

Submission will be due online or via e-mail to the Professor. TBD

A data professional will possess superior communication skills. During your career you will need to document projects and develop action plans. This report is used as an update and to clearly outline your plan of attack. Envision you are submitting this planning document to your boss to ensure he/she is aligned on your approach.

In general, you are building a data product and you will provide an overview of the overall architecture of your product and the results that you pan to get. Get creative and apply the learning for lectures, but also consider the following:

* Describe the project significance, anticipated outcomes and any limitations. (these might change during the course of your project and if so, update in future documents)
* How the proposed project fits with your program of study (what course you have taken or taking that relates to the project – techniques and applications)? How do you plan to apply these skills/tools?
* Timeline with tentative deliverables.

Submission will be due online or via e-mail to the Professor. TBD

1. Interim Presentation

Format: *10 minute presentation + 5 min Q&A -7 slides max*

Present out to your peers and gather feedback. Ask your peers for help/ideas – make it a discussion vs presentation. Your PowerPoint should include some of the following components, but please do not use this as a checklist/table of contents…get creative:

1. Cover: Informative title, team member intros/roles
2. Provide a brief background of organization/problem addressed.
3. Business problem/opportunity (stakeholder, challenge/opportunity, humanity considerations)
4. Data mining problem & Data description
5. Implications – so what can business/org do with findings/actions
6. Preliminary Recommendations (what should the client be aware of? problems you encountered, suggestions for future data collection or analysis, etc.)

No submission required.

6. Final Report (5 pages)

The project report details the team project, from the business problem through the data mining problem and solution, to recommendations.

Format: *5 pages, 11-point font, 1-inch margins, 1.15-line space, World format.*

The report should be written clearly and professionally and include the following sections:

**Cover page** with informative title, team number and member names

**One-page executive summary**: summarizes the entire report for a non-technical manager (the business problem, data, the analytics solution and recommendations)

**Detailed report:**

* Problem description (business goal and data analysis goal)
* Data description (the data that you end up using: size and dimension, what is a record, list of output and input variables,).
* Data preparation details (how your data were created from the raw data) and key charts. Details can be provided in an Appendix.
* Data analysis solution: Methods applied (with sufficient detail and screenshots; use Appendix if needed) and appropriate performance evaluation (proper choice of measures, benchmarking).
* Conclusions (advantages and limitations) and operational recommendations

7 . Project Presentation (15 minutes presentation + 5 min Q&A -10 slides max)

Each team presents their project to the entire class and mentors are welcome to attend. The project is based on the proposal submission, where the idea is then developed into a data mining solution with final recommendations.

Following slides are a few suggested ideas:

1. Cover: Informative title, team number and member names
2. Provide a brief background of organization/problem addressed.
3. Business problem/opportunity (stakeholder, challenge/opportunity, humanity considerations)
4. Data mining problem (supervised/unsupervised, explanatory/predictive, how to be deployed)
5. Data description
6. Data cleansing, if needed – steps taken – what were the challenges with the data and steps taken to cleanse bring it to usable format (ETL – extract, transform, load steps). Technology/tools used for cleansing/analysis
7. Methods (methods, relevant outputs)
8. Evaluation (metrics of interest, benchmark, comparison)
9. Implications – so what can business/org do with findings/actions
10. Recommendations (what should the client be aware of? problems you encountered, suggestions for future data collection or analysis, etc.)