Syllabus for DS 6011: The Capstone Experience

School of Data Science, University of Virginia | Fall Semester 2023

Instructor: Adam Tashman apt4c[@virginia.edu](mailto:jasonw@virginia.edu)

# Course Description

The purpose of the capstone experience is to provide students with an understanding of data science as it may be practiced in the context of real-world problems. It allows the student to collaborate as a member of a team to pull together knowledge from the diverse areas of data science and integrate them in an effective and professional manner. The focus is on addressing an actual client’s need by building a data product that can be shared with the client. The course comprises two semesters, DS 6011 and DS 6013, each devoted to roughly half of the data science pipeline. A simplified version of this pipeline has these four phases: (1) framing the question, problem or value proposition, (2) acquiring, transforming, and exploring the data, (3) analyzing and modeling the data, and interpreting the results, and (4) packaging, communicating, and sharing the results. In general, DS 6011 covers parts 1 and 2 while DS 6013 focuses on parts 3 and 4. During this process, students will be introduced to both technical and professional challenges and will be guided by basic project management protocols in meeting them. The course sequence culminates in the production of a publishable paper and a presentation of results to the client.

# This Document

This document focuses on the first semester (DS 6011) of this experience. A separate syllabus will be provided for the second semester (DS 6013). The following is designed to provide general guidelines for each individual capstone seminar. Capstone advisors may amend as necessary.

# Course Objectives

In general, the course will allow students to gain experience in the following areas:

1. Framing and translating real-world problems into operational terms to produce tangible results
2. Acquiring, managing, and processing data in a variety of forms and conditions
3. Selecting, developing, applying, and evaluating models, tools, and methods to apply to the data in response to the framing problem
4. Interpreting and assessing results of modeling and evaluating the limitations of research findings
5. Working as a team to develop approaches and solutions, allocate the work, and communicate with clients
6. Communicating results in a clear, concise, and compelling manner for the client and the processional community of data scientists

# General Process

Each capstone advisor has the discretion to oversee the students’ learning experience and interaction with the client using tools and methods they find appropriate. In general, each individual capstone group will make use of a task management tool, such as Trello or Smartsheet, to plan, organize, and share activities. In addition, teams will establish a communication protocol for meeting among team members and clients.

## Class meetings

Classes meet once a week for one hour. A typical one-hour course meeting might follow this pattern:

1. Begin with any announcements and course business issues. (5 minutes)
2. Have each capstone team report on progress and issues arising, using a task management tool to guide discussion. Allow other teams to respond and discuss. Teams may also demonstrate data and code at this point. Advisors may suggest that members of each team rotate the duty of reporting. (10 minutes)
3. Have teams break out into their own group and work, making sure to update their task board with new activities. Optionally reconvene and show updated tasks.
4. Steps 2 and 3 may be reversed as well.
5. 15 min discussion on a project management topic.

In general, these meetings may be run as business meetings or stand-up meetings, where the goal is to assess the current state of things and make decisions about moving forward. In addition, time may be taken in these meetings to present information or provide guidance on technical topics as they arise.

If a team member cannot attend a weekly meeting due to a valid excuse, they should notify the instructor and the team members ahead of time. Skipping a meeting without notification may result in a score of 0 for the weekly evaluation.

## Milestones

In addition to the general process for each class meeting, the structure of the semester is loosely structured as a project, with the expectation that the following sequence of milestones be met:

1. Overview meeting. This is the first meeting of the class after capstone projects and teams are selected.
2. **Team contract.** As soon as teams are formed, students should produce a document that establishes expectations for participation in the capstone team.
3. Kick-off meeting with client. This should happen as soon as possible. Students are expected to set up a meeting with their clients.
4. Review of the literature. In addition to discussing project scope and goals with the client, students are expected to review existing literature on the project problem, in order to focus attention on current methods and approaches.
5. Data acquisition. Parallel to the process of reviewing the literature and framing the problem, students should be actively engaging with the client to acquire the data required to complete the project.
6. **Project proposal**. The project proposal defines the scope of the project and its general requirements. It also outlines the process by which the project will be completed.
7. **Progress Presentation**.

Items in bold are associated with assessments. Instructions for these are provide in separate documents.

## Additional Readings

I have posted a reading list that we will use to enhance the project management experience and learnings. We will discuss them on Monday and you will have reflection questions due on the following Wednesday, if a reading assignment was given.

You may download the course pack here à <https://hbsp.harvard.edu/import/968837>

## Communication

Teams will be used as the central channel for course collaboration. Please use Teams to post questions regarding project problems, objective identification, data collection and/or compilation, data processing, data structure and management, and exploratory data analysis. We’ll use this for current updates for dates/times…etc.

# Grading

Participation 40%

Team Contract 10%

Project Proposal 15%

Progress Presentation 35%

# Schedule

Classes meet online each Monday from 3:30-4:30

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Week** | **Date** | **Topic** | **Reading** | **Items Due** |
| 01 | 08/29 | Overview |  |  |
| 02 | 09/05 | What is Success | HBR Guide to PM Chapter 1,2 &3 | Reflection |
| 06 | 09/12 | Fuzzy Front End | HBR PM Chapter 4 | Reflection |
| 03 | 09/19 | Meetings | HBR PM Charter 13 |  |
| 04 | 09/26 | Data Sizing | TBD |  |
| 05 | 10/03 | **FALL BREAK** |  | Team Contract |
| 07 | 10/10 | Overview of Budget |  | Reflection |
| 08 | 10/17 | Failed DS Projects |  | Reflection |
| 09 | 10/24 |  | Failure Article/ HBR Chapter 17&18 | **Budget (In Planner)Due 10/26 at midnight** |
| 10 | 10/31 | What is project success | No Reading before...will give after lecture | Open question on Budgets...submit them on Planner and make sure you add me |
| 11 | 11/07 | More Comms | Data Science and the Art of Persuasion | Reflection |
| 12 | 11/14 | Risk | Is your project turning into a black hole | Reflection |
| 13 | 11/21 | Getting Results | Leadership that Gets Results | Reflection |
| 14 | 12/05 | Progress Presentations |  | Progress Presentation Slides |

# Assessments

Team Contract. The team contract is a document the students in each term create to codify and agree on behavioral and communicative expectations in the project. It is meant to make all team members aware of the professional expectations for participating in the capstone.

Project Proposal. The project proposal defines the scope of the project and its general requirements. It also outlines the process by which the project will be completed. It is a more or less binding document agreed to between the team and the client.

Progress Presentation. At the end of the first semester, teams will present on their progress in the first two phases of the pipeline—-framing the problem and establishing the data. This presentation should contain a summary of work to date including questions, problems and objective identification, data collection or compilation, data processing, data structure and management, and exploratory data analysis. Students are expected to cite key literature where appropriate. After each presentation, the class will offer constructive criticism and feedback to help guide future work.