Module 7: Capstone

Get hyped. The last 3 weeks of the program are all about your final project! This is the capstone project that you'll be showing off to demonstrate all the things that you've learned over the last few months. This is awesome!

But, before we dive into what your project is going to be about we need to establish some ground rules.

Ground Rules

- Show up by 9:00 and leave no sooner than 6:00
- Participate in stand-ups, retros, and check-ins
- Finish acquiring your data and have a baseline model and dataset (**MVP**) by the established due date
- Meet with your scrum group in the morning and at the end of the day
- Adhere to the project decisions made with project managers (coaches)
- Communicate issues with your project managers (coaches)
- Be courteous to your classmates; Work together to solve problems
- Research and debug before escalating to your project manager
- Don't proceed beyond basic data exploration until you have pitched your project and it has been approved by your instructors.

Schedule

For much of Mod 7, your time is your own. There will be a few lectures and other activities, but most of the time will be dedicated to your project. This does ***not*** mean you get to come in late or leave early or take four hour lunch breaks. Think of this as a work environment. Below is the schedule for your few finals weeks here at Flatiron.

Capstone Project Timeline

Daily Stand-up/Stand-down

Scrum group - Students will be assigned groups of 5-6. All group members will become intimately familiar with each others' projects. Responsibility of scrum group members are:

- Morning stand ups
- Provide each other with advice, brainstorming, and resources
- End of day stand downs

Stand-up/Stand-down

1. **Accomplishments:** I've completed _____ since our last check-in.
2. **Plans for today:** Today I will plan to accomplish _____.
3. **Blocking Issues:** I need help with _____.

Final Project

This is the time to dive into the data science process using techniques we've learned. This is the time to build the model that you've always wanted to but never could, and really put your skills on display.

Because we're asking you to show off a specific set of skills, we have some requirements. One of the requirements is that you need to implement the knowledge that you learned while at Flatiron. This isn't the time to build a new classification algorithm or do reinforcement learning. You've done a ton of learning already - it's time to **apply** all of that knowledge.

It is entirely up to you to select both the the data you will use and question to answer. When narrowing down the options for your project, please refer to this list of questions. They are intended to get you thinking about project management and how you are going to allocate your time.

Project Guidelines

Final project approval is up to your coaches and leads, who will be acting as project managers. You are **required to** get approval for your project, otherwise you will not be able to move forward on the project and therefore graduate.

Your final project is an elevated end-of-mod project. Your project must include some modeling. It can be from any of the topics that we've covered over the past 12 weeks and should demonstrate an understanding of the data science concepts you learned at Flatiron. Consider if your project will be relying on models or methods that we have covered so far, or going to be incorporating new ones. Try to keep the amount of new material to a minimum to ensure the last few weeks are spent on working on your project, rather than learning many new concepts.

In the coming days, you will be asked for at least two potential ideas for your final project. Your coaches will help pick one idea, and then you will be responsible for writing up a proposal for lead approval. Once you receive approval, you can begin work on your project. Proposals will be submitted via Canvas.

Elements that Must be Included:

- A prediction on a new datapoint
- Input features and an output prediction
- Use of one of the models listed below
- Be able to be completed in less than 3 weeks

The dataset you ultimately select must contain at least 1000 rows of data (for image classification or NLP projects you must have at least 1000 pictures/observations of text).

Models to Choose From

Here are the models you can use. See section on special project permissions if the model you would like to use is outside of this list.

- Regression models (linear, CART, etc)
- Classification models (KNN, CART, logistic, etc)
- Tree Based models (decision tree, random forest, boosting)
- Time series models
- Neural Networks (limit to basic NN, CNN and RNN)
- Recommendation systems

Data Science Concepts/Components You Can Include (Think wisely about how many of these you are using!)

- Databases (SQL, MongoDB, etc)
- API Interaction
- Clustering
- Hadoop/Spark components
- Natural Language Processing
- Image Processing
- A webapp to showcase your project

Special Project Permission:

If you would like to use a model outside of the above list you must notify instructors before your initial coach pitch. You will need to show proficiency in this model and provide evidence that you can utilize it for the purpose intended in your project early within the project proposal process. In addition, staff support may be extremely limited in assisting you with any roadblocks regarding your model.

Project Essentials

1. A **clean Github**! Try to make it as reader-friendly as possible. How would you organize your project repo to best demonstrate your project to a recruiter?
    a. Readme, clean notebook, gitignore, etc.

2. Within your Github, we really want you to **demonstrate the data science process** in your code/notebooks. As we've learned over the past 12 weeks, the process can be broken down into:
    a. Problem identification
    b. Data wrangling/cleaning
    c. Feature selection/engineering
    d. Model building/tuning
    e. Evaluation
    f. Final product

3. **The elevator pitch** should be a short overview of your project **around 3 minutes** that's compelling to non-technical people, **not a 15 minute technical presentation.** Your project should have a **narrative** that ties the whole thing together. It's very important to contextualize! *Why are you doing your project?* This pitch will be utilized during the Final Project Showcase.

4. A **final presentation**. On Day 5 of Week 3, we'll be doing project presentations. This is the time for you to present your hard work to your peers! Presentations should be within 4 to 5 minutes long. There will be a **hard stop at 5 minutes**, including questions! Take a look at the rubric to know how you will be graded.

Helpful Tools

Just because you're working solo doesn't mean you don't need to stay organized. In fact, because this will be the most complex project you've made at Flatiron, you'll *need* something to keep you organized.

Kanban/Scrum Board

We recommend Trello or a Github Project Board. Use this to track what you're doing and what you need to work on. It's also a great idea to keep track of bugs that you're not going to immediately fix.

Pomdoro Timer

If you don't take breaks, you'll end up hurting your eyes, getting an RSI, or burning yourself out. The Pomdoro Timer method lets you put in solid chunks of work while also giving you regular breaks. We like Marinara Timer, since it's nicely customizable.

Project Resources

Postman - Test API calls

Heroku - Simple, free web hosting for flask or django in python

DB Browser - SQLite interface for making calls to a database

AWS - Amazon Web Services for running models on the cloud