

# Capstone Project

Get hyped, Datanauts! The next 3 weeks of your life 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 be available on Slack (taking healthy breaks) until 6:00
- Participate in stand-ups, whiteboarding, and check-ins
- Finish acquiring your data and have a baseline model and dataset (**MVP**) by the established due date: **4/10**
- Adhere to the project decisions made by project managers
- Communicate issues with your project manager
- 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.

## 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. It should be obvious that 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.

## Project Requirements

The listed requirements below are guidelines that should help you to determine what the complexity of your project should be. They are not hard and fast rules. Final project approval is up to your leads and coaches, who will be acting as project managers.

Your final project is an elevated end-of-mod project . It can be from any of the topics that we've covered over the past 12 weeks and should demonstrate an understanding of data science concepts.

## Things You Can Build

Here are some suggestions of things you can do; you're certainly not limited to these!

- Regression (linear, CART, etc)
- Classification (scikit models, neural nets)
- Time series
- Clustering
- Recommendation engine

## Data Science Concepts/Components You Can Include

- Databases (SQL, MongoDB, etc)
- Natural Language Processing
- Image Processing
- Some frontend

## What Do We Want?

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. Data wrangling/cleaning
  - b. Feature selection/engineering
  - c. Model building/tuning
  - d. Evaluation
  - e. Final product
3. Your project should have a **narrative** that ties the whole thing together. Why are you doing your project? At each step of the way you should be able to justify your decisions in relation to the purpose of your project. It's very important to contextualize!
4. It seems like common sense, but it's very important that you **understand the models and the concepts** that you're incorporating into your project. If you use Singular Value Decomposition in your recommendation system, you should be able to answer the question: How does SVD work? These project-related questions are guaranteed to come up in job interviews! As a **bonus**, be able to explain these concepts at different levels. i.e. explaining XGBoost to a technical/non-technical person.

5. A **final presentation**. On Day 3/Day 4 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 4.5 minutes long. There will be a **hard stop at 5 minutes!**

## Schedule

For much of the final project, 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 sleep in late or leave early or take four hour lunch breaks. Think of this as a work environment. Below is the schedule for a typical day as well as a list of particularly important dates.

### Typical Day:

- 9:30 Stand-ups (initially this will be random groups until you have been assigned a PM)
- 1:30 Whiteboarding + Mid-day stand-ups
- 3:00 Student led Study group / Optional Lecture (Varies based on interest)
- 5:30 Daily log deliverables
- \*\*\*\* Scheduled PM check in (See note below)

\*\*\*\* **PM check ins** - Once projects have been chosen you will each be assigned an instructor as your Project Manager. For the two following weeks you will have an assigned time to check in with your PM and discuss your project in greater detail. Expect to discuss and justify the decisions you've made for the project and to be asked probing questions about your understanding of the models and concepts being used.

**Whiteboarding** - We will be working through interview questions in small groups on a whiteboard to simulate a common interview practice.

**Study groups** - Instructors will facilitate study groups at the request of students. This is an opportunity to collaborate with your classmates and discuss the cool libraries/pipelines/models you've discovered in the course of working through your final project and to brainstorm ways to push forward.

**Daily log** - In place of stand downs every student will be delivering to the instructors over Slack a daily log of their progress. This is an important and **mandatory** part of the day as it will be used to determine attendance. The daily log should contain at least 3 specific objectives completed over that day, e.g:

- Visualized my target class to understand its distribution
- Grouped features by target class to compare their distributions
- Researched different scaling/normalization techniques for different distributions

*\*\* Please let us know in advance if circumstances complicate your adherence to any part of the above schedule. \*\**

## Week 1

### Day 1 - Initial Project Brainstorming

This is when you get to pitch whatever pie in the sky things you want. We're not yet worried about what machine learning models you will use, having MVP (minimum viable product), or whether your idea is practical - come up with a couple ideas. You won't be starting on your project until Day 3, so think of this as your research and development phase. Look into interesting datasets you may want to use, APIs and machine learning libraries. Your coaches will give you an idea of what is and isn't practical, and some guidance on the technologies you're looking at.

### Day 3 - Project Proposals

You will have projects (always have some backups) ready to pitch to a panel including a lead and your coaches. You'll have your data sources, models, and libraries you would like to use, along with evidence that you're able to use the APIs and libraries that you're going to implement. You'll also have a proposal for what your machine learning models you will use and MVP will look like. You're expected to have your data, a baseline model (or MVP complete), **along with a timeline** for features you would like to add a week from this day. Together you and your instructors will determine the best project to move forward with.

## Week 2

### Day 3 - EDA Deliverables

By this point you will have a cleaned dataset and have performed EDA to understand the underlying structure of your data. Over slack you will deliver 5 distinct visualizations and specific insights to your project manager and a lead instructor to display this understanding.

### Day 5 - MVP Presentations

You will present your MVP to the instructors with rudimentary slides (no need to make these fancy, just functional) along with a timeline for additional features or additions you

would like to make over the following week. Presentations should last no more than 5 minutes to give instructors time for followup questions.

## Week 3

### Day 3 - Projects Complete

By week three, day three, you should have your project done. There'll still be time to work on it before the science fair, but this should be cleaning notebooks, the readme, and working on the presentation - you should be done with core modeling and data wrangling. We will have a dry run of presentations with the entire cohort and instructors to give us time to ask questions and provide feedback before the Science Fair.

### Day 4 - Science Fair

This is the day that you'll be showing off your projects to the rest of the school, friends and family, and prospective employers. Science fair lasts from around 4:30 to around 6:00. The structure will look like past projects where you will take turns presenting to the audience with time afterward for follow-up questions. Make sure to put your best foot forward, dress a little nicer than you normally would, and get that elevator pitch ready! **The elevator pitch for the science fair should be a short overview/demonstration of your project around 5 minutes.**

Really think about the main aspects of your project you want your audience to take away. **What** did you build? **Why** did you build it? **How** does it contribute to society? Often, unless you did something revolutionary, people won't care about your data cleaning process. (Sad. I know.) Imagine you want to tell your non-technical friends about your project. Can you do that in under 5 minutes? Of course, be prepared to discuss in technical detail as there will be time after your presentation for questions from the audience.

### Day 5 - Graduation

You made it! Today you get your final survey and you get to take part in the graduation ceremony. Graduation will be over by noon. (Timing subject to change)

# Helpful Tools

## Kanban/Scrum Board

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. 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.

## Pomodoro Timer

If you don't take breaks, you'll end up hurting your eyes, getting an RSI or burning yourself out. The Pomodoro 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.

# Code Resources

[Postman](#) - Test API calls

[Heroku](#) - Simple, free 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

**Read the documentation!!!!!!**