

# Capstone Project Proposal Template

## Notes:

- This should take no more than one hour to complete – the clearer you are about the business problem you're working to solve with your ML-driven solution, the easier your proposal will be to complete
- This will be uploaded to your repo, which will be a part of your final submission
- Due date for submission is 1/16

## Instructions:

1. Download this document as a Word Doc
2. Answer each question using a few sentences, at most
3. Save your completed proposal as a PDF
4. [Create a project GitHub repo](#) (if you have yet to do so)
5. [Add your instructor as a collaborator](#) (username `dodgy719`) to your project repo
6. Add your mentor as a collaborator
7. Push your proposal PDF (created in Step 3) up to your repo
8. Copy the URL corresponding to the location of the PDF in your repo
9. Submit the copied URL using [this link](#)

## [project name]

### Business Understanding

- What problem are you trying to solve, or what question are you trying to answer?

Attempting to understand which video games are successful based on the genre/ type.

- What industry/realm/domain does this apply to?

### Video Gaming

- What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)

A new gaming company is tasked with creating a video game but wants to advertise to the largest crowd possible while sustaining income in the long term.

### Data Understanding

- What data will you collect?

Information about which video games have been the most successful and the genre which they belong too.

- Is there a plan for how to get the data (API request, direct download, etc.)?

Direct download from open-source information on Google.

- What are the features you'll be using in your model?

Genre of video game, release date

### **Data Preparation**

- What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?

Encoding for the names, release date, and genre.

- What are some of the cleaning/pre-processing challenges for this data?

Shouldn't have any cleaning problems as all the scalable data will be extracted from reliable sources. There is limited data required, missing values MAY occur.

### **Modeling**

- What modeling techniques are most appropriate for your problem?

Regression

- What is your target variable? (remember - we require that you answer/solve a supervised problem for the capstone, thus you will need a target)

Popularity and long-term retention.

- Is this a regression or classification problem?

Regression

### **Evaluation**

- What metrics will you use to determine success (MAE, RMSE, Accuracy, Precision etc.)?

RMSE

### **Tools/Methodologies**

- What modeling algorithms are you planning to use (i.e., decision trees, random forests, etc.)?

Simple and linear regression