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](https://github.com/new) (if you have yet to do so)
5. [Add your instructor as a collaborator](https://docs.github.com/en/account-and-profile/setting-up-and-managing-your-personal-account-on-github/managing-access-to-your-personal-repositories/inviting-collaborators-to-a-personal-repository) (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](https://my.learn.co/courses/586/quizzes/6353?module_item_id=79223)

**[project name]**

**Business Understanding**

* What problem are you trying to solve, or what question are you trying to answer?
  + This project will seek to solve the problem of customer churn in the banking world. The predictive models constructed would serve to identify factors that lead to customer churn and in turn reduce customer churn for banks and financial institutions. The question being answered is- Will the given customer churn (leave the bank) or not?
* What industry/realm/domain does this apply to?
  + Banking industry
* What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)
  + The motivation behind the project is to apply my learnings from AI academy to develop useful predictive models to minimize banking customer churn.

**Data Understanding**

* What data will you collect?
  + A dataset consisting of customer data as well as data fields related to the customer’s history at the given bank.
* Is there a plan for how to get the data (API request, direct download, etc.)?
  + Direct download from Kaggle
* What are the features you’ll be using in your model?
  + Name, Credit score, geography, gender, age, tenure, balance, salary, credit card status, exited, number of bank products used, member status

**Data Preparation**

* What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?
  + One hot encoding certain categorical variables
* What are some of the cleaning/pre-processing challenges for this data?
  + Missing values, preventing data leakage

**Modeling**

* What modeling techniques are most appropriate for your problem?
  + Regression techniques including logistic 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)
  + Customer churn
* Is this a regression or classification problem?
  + Regression problem

**Evaluation**

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

**Tools/Methodologies**

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