Capstone Project Proposal

**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 12/9

**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 nickmccarty) 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/543/quizzes/6212)

**Makeup Sales Predictions**

**Business Understanding**

* What problem are you trying to solve, or what question are you trying to answer?
  + I am trying to answer the question of when makeup companies should release new makeup to optimize their sales.
* What industry/realm/domain does this apply to?
  + This applies to the cosmetics industry
* What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)
  + Makeup is something that I am very interested in and think it will be very interesting and beneficial to see when the best time for companies is to release new products. Especially for newer makeup brands.

**Data Understanding**

* What data will you collect?
  + I will collect data on the date and time that the most people are purchasing makeup.
* Is there a plan for how to get the data (API request, direct download, etc.)?
  + I plan on getting the data from a Kaggle database. I already have a few that I am looking into using.
* Are the features that will be used described clearly?
  + Yes I believe that the features are described clearly.

**Data Preparation**

* What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?
  + I think the preprocessing steps that I am going to have to take are to remove or replace missing data and encoding the data.
* What are some of the cleaning/pre-processing challenges for this data?
  + Some of the cleaning challenges for this data are going to be merging the different months within the dataset into one data set. Also, I will need to get rid of some columns that will not be useful in my model.

**Modeling**

* What modeling techniques are most appropriate for your problem?
  + I think I am going to use linear regression for my model. However, I might change my mind when I do more research into my data.
* What is your target variable? (Remember - we require that you answer/solve a supervised problem for the capstone, thus you will need a target)
  + My target variable is sales.
* Is this a regression or classification problem?
  + Regression

**Evaluation**

* What metrics will you use to determine success (MAE, RMSE, etc.)?
  + I am thinking of doing R-squared.

**Tools/Methodologies**

* What modeling algorithms are you planning to use (i.e., decision trees, random forests, etc.)?
  + I am thinking of using a decision tree however I am not totally sure yet.