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 end-of-day 3/13 for Cohort 3b

**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 jvntra) 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/615/quizzes/6356?module_item_id=79238)for Cohort 3b

**Predicting Prices of Used Hondas**

**Business Understanding**

* What problem are you trying to solve, or what question are you trying to answer?
* What industry/realm/domain does this apply to?
* What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)

I plan on using data collected on used Hondas to predict the price a car will sell at based on the attributes of the car including its milage, suspension, model, and year. This study will apply to the automotive industry. I am interested in seeing how different characteristics of the cars impact the car’s value.

**Data Understanding**

* What data will you collect?
* Is there a plan for how to get the data (API request, direct download, etc.)?
* What are the features you’ll be using in your model?

I plan on collecting data on Used Honda resale prices along with the vehicles year, model, mileage, suspension, and fuel type. I plan on directly downloading this data from Kaggle. I plan on using NumPy, Pandas, Seaborn, and Matplotlib.

**Data Preparation**

* What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?
* What are some of the cleaning/pre-processing challenges for this data?

This data is coming from India so the currency is in Lakh which I will need to calculate the exchange rate for in USD. Similarly, the data shows the histories of the cars in Kilometers which I will need to convert to Miles. Many of the data types are string which I plan on transforming into integers and booleans.

**Modeling**

* What modeling techniques are most appropriate for your problem?
* What is your target variable? (remember - we require that you answer/solve a supervised problem for the capstone, thus you will need a target)
* Is this a regression or classification problem?

I plan on using supervised modeling techniques. The target variable for my model will be price. (Using data on used Hondas, can we predict the price they will sell at? This will be useful in determining which factors have a large impact on the value of the used Hondas.) This will be a regression problem.

**Evaluation**

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

I will be using R^2, MAE, and RMSE to determine success.

**Tools/Methodologies**

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

I plan on using XG Boost, random forests, and linear regression.