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 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)

**Optimal Airbnb Price by Region**

**Business Understanding**

* What is the best price to list a rental property to maximize profit throughout the year?
* This applies to the estate market and the property rental market.
* I have always wanted to rent out properties on Airbnb and this will give me an edge to know the best price range to list my properties.

**Data Understanding**

* I collected my data from Kaggle.
* I download the dataset in a zipped file than I unzipped the file in python. Then I read the data in as a panda frame.
* Yes, it is the same process used in the first capstone. I will describe the process in the final notebook.

**Data Preparation**

* I will be doing some encoding to make the data more readable for the models. I will most likely encode the room type.
* The dataset I found was in a great position before cleaning. There was only a little missing data that I dropped for the most part. The hardest activity will be encoding the data to have less words and more numbers for the model.

**Modeling**

* I believe an XGBoost and Random Forest Regressor would be the best two.
* The target variable will be price.
* This is a regression problem.

**Evaluation**

* I will be using MSE and RMSE to determine success

**Tools/Methodologies**

* What modeling algorithms are you planning to use (i.e., decision trees, random forests, etc.)?
* I am planning on using XGBoost and Random Forest. I will use XGBoost to find the most variables to the model. Then I will use a random forest to find the MSE and RMSE.