Capstone Project Proposal Template

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 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

Beau Home House Prices Prediction

Business Understanding

- **What problem are you trying to solve, or what question are you trying to answer?**
    - Beau Home Properties is a new Real Estate Company that is just starting out in the State of California. They will be entering the business of investing in properties by buying and selling. They reached out to our Data Science team to build a value estimation system that can automatically deduce the value of the houses. This project will focus on building a machine learning model trained on existing housing features to predict house prices.

- **What industry/realm/domain does this apply to?**
    - Real Estate Industry

- **What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)**
    - I have a personal interest in Real Estate, and I would like to explore the different features that are involved in predicting housing prices.

Data Understanding

- **What data will you collect?**
    - Data on the housing market in California with different features on the houses.

- **Is there a plan for how to get the data (API request, direct download, etc.)?**
    - I downloaded the data directly from Kaggle.

- **Are the features that will be used described clearly?**

    - Yes, the features and the columns are clearly described with proper column names.

Data Preparation

- **What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?**
    - Since some of the housing features are categorical, one-hot encoding will be used in order to utilize those data points.

- **What are some of the cleaning/pre-processing challenges for this data?**
    - The data will require cleaning such as removing duplicates, missing values, one-hot encoding categorical variables.

Modeling

- **What modeling techniques are most appropriate for your problem?**
    - Linear 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)**
    - House Prices predictions

- **Is this a regression or classification problem?**
    - Regression problem

Evaluation

- **What metrics will you use to determine success (MAE, RMSE, etc.)?**
    - This will be identified as the project goes forward.

Tools/Methodologies

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