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 (if you have yet to do so)
5. Add your instructor as a collaborator (username dodg719) 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

Predicting Hotel Cancellations

Business Understanding

- What problem are you trying to solve, or what question are you trying to answer? Predicting Hotel Cancellations
- What industry/realm/domain does this apply to? Hospitality Industry
- What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation) During the holiday disconnect, I was able to find a room at a hotel at the last minute due to a cancellation. This made me wonder what factors lead to cancellations.

Data Understanding

- What data will you collect? Hotel bookings dataset
- Is there a plan for how to get the data (API request, direct download, etc.)? Direct Download
- What are the features you'll be using in your model? Pandas, numpy, matplotlib, seaborn

Data Preparation

- What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)? Encoding categorical variables
- What are some of the cleaning/pre-processing challenges for this data? Dealing with null values, determining which columns from the dataset are relevant and dropping the irrelevant ones.

Modeling

- What modeling techniques are most appropriate for your problem? Logistic regression, Decision trees, Random forests, Artificial Neural Networks
- 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 Canceled
- Is this a regression or classification problem? Classification (Cancel: Yes or No)

Evaluation

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

Tools/Methodologies

- What modeling algorithms are you planning to use (i.e., decision trees, random forests, etc.)? Logistic regression, Decision trees, Random forests, Artificial Neural Networks