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

Movie Analysis - What makes a movie successful

Business Understanding

- What problem are you trying to solve, or what question are you trying to answer?
    - Trying to understand what contributes to a successful film and what are some of the metrics to expect.
- What industry/realm/domain does this apply to?
    - The entertainment industry
- What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)
    - A new company is trying to get into the entertainment industry and they want to know what sort of a return on investment they should expect and who are some of the people to watch out for and keep in touch with.

Data Understanding

- What data will you collect?
    - Information about the movie gross and all the details about cast and crew as well as genre and runtime of the film.

- Is there a plan for how to get the data (API request, direct download, etc.)?
    - I will get the data from a dataset on Kaggle- direct download
- What are the features you'll be using in your model?
    - Genre of the film, release date, the cast and crew (actors and directors)

Data Preparation

- What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?
    - One hot encoding for the names, probably extracting some information from the release date including the release month
- What are some of the cleaning/pre-processing challenges for this data?
    - Dealing with missing values, making sure there is no data leakage, test train split

Modeling

- What modeling techniques are most appropriate for your problem?
    - Regression techniques including multiple 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)
    - My target variable will be box office gross
- Is this a regression or classification problem?
    - Regression problem

Evaluation

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

Tools/Methodologies

- What modeling algorithms are you planning to use (i.e., decision trees, random forests, etc.)?
    - Will start with simple and multiple regression and try to include decision trees.