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 (if you have yet to do so)
5. Add your instructor as a collaborator (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 for Cohort 3b

How to make a hit: Spotify Song Data Analysis

Business Understanding

- What problem are you trying to solve, or what question are you trying to answer?
    - How do you make a hit song?
        - More specifically, which features go into making a hit song
- What industry/realm/domain does this apply to?
    - Entertainment/music
- What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)
    - I like music. And I want to use what I learn from this later to analyze music in different ways

Data Understanding

- What data will you collect?
    - Spotify API
    - Billboard Webscraping (already done)
- Is there a plan for how to get the data (API request, direct download, etc.)?
    - Webscraping and API
- What are the features you'll be using in your model?

    - Tempo, genre, Spotify created metrics (dancability, energy, wordiness, etc)

Data Preparation

- What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?
    - Scrapping and formatting data from billboard
    - Getting data from API
    - Maybe some hotoneencoding
- What are some of the cleaning/pre-processing challenges for this data?
    - The biggest challenge will be getting the data for top songs and getting enough data

Modeling

- What modeling techniques are most appropriate for your problem?
    - Regression, EDA stuff,
- What is your target variable? (remember - we require that you answer/solve a supervised problem for the capstone, thus you will need a target),
    - Number of streams (and a cut off to see if it's a hit)
- Is this a regression or classification problem?
    - classificatoin

Evaluation

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

Tools/Methodologies

- What modeling algorithms are you planning to use (i.e., decision trees, random forests, etc.)?
    - Regression for EDA
    - Decision tree / random forrest
    - Perhaps Logistic regression
    - NN
    - CNN (waveform analysis)