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 (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

[project name]

Business Understanding

- What problem are you trying to solve, or what question are you trying to answer?
  My Capstone project will attempt to create a working music recommendation system by checking songs already liked by the similar users and

- What industry/realm/domain does this apply to?
  This project will apply to the Music/Streaming industry.

- What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)
  As streaming libraries grow bigger and finding music has become difficult algorithms that can recommend music that the user is most inclined to like will keep users happy and coming back

Data Understanding

- What data will you collect?
  The Million Song dataset and the Spotify dataset off Kaggle.

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

- Are the features that will be used described clearly?
  Yes, the features that are used are described clearly.

Data Preparation

- What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?
  So far in the project I only see simple encoding as a preprocessing step I have yet to dive far enough into the data to.

- What are some of the cleaning/pre-processing challenges for this data?
  The data is large and has many missing values that will need to be dealt with.

Modeling

- What modeling techniques are most appropriate for your problem?
  A Relational Model will be most appropriate for my problem.

- What is your target variable? (remember - we require that you answer/solve a supervised problem for the capstone, thus you will need a target)
  The target variable will be whether there are recurring listening event(s) triggered within a month after the user's very first charted listening event.

- Is this a regression or classification problem?
  This will be a regression problem.

Evaluation

- What metrics will you use to determine success (MAE, RMSE, etc.)?
  I will use mean absolute error to evaluate my project.

Tools/Methodologies

- What modeling algorithms are you planning to use (i.e., decision trees, random forests, etc.)?
  I'm planning to use decision trees, collaborative filtering, and logistic regression.