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 dodgy719) 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

[Musical Recommendation System)

Business Understanding

- **What problem are you trying to solve, or what question are you trying to answer?**
    - Many music related businesses and services, including record labels (Rough Trade, EMI, Matador, etc.), music reviewing publications (Pitchfork, Stereogum, Consequence of sound, etc.), radio stations (KEXP, NPR, The Current, KCRW, etc.), rely on music curation and playlist creation, as a means to establish their presence or artists presence in their respective music scene. A music related company that delivers popular playlists with songs that appeal to a broad range of listeners is more likely to succeed than a music related company that does not. I am looking to see if it is possible to create a music recommendation system, to help these businesses choose the best songs/sounds/genres to apply to their playlists to achieve maximum visibility and exposure, either to their publications. Radio stations, and record labels/artists.
- **What industry/realm/domain does this apply to?**
    - Music/Entertainment Industry
- **What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)**

    - I am personally curious to see what results my project yields as I work on music related endeavors outside of Deloitte employment.

Data Understanding

- **What data will you collect?**
    - I will be collecting data from Spotify, as it is the music streaming service that I use regularly.
- **Is there a plan for how to get the data (API request, direct download, etc.)?**
    - I plan on collecting data as a direct download from a Spotify data archive site. I also may use the Spotify Web API for extra data as well.
- **What are the features you'll be using in your model?**
    - Assuming this means data features, I will be taking a look at an overall dataset of Spotify music genres, data by artists, genres, and years.

Data Preparation

- **What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?**
    - I will be using sklearn.preprocessing import StandardScaler for preprocessing data.
- **What are some of the cleaning/pre-processing challenges for this data?**
    - Since I am pulling data straight from an old Spotify archive, I do not believe data cleaning will be necessary, since Spotify will not allow the upload of songs without certain data categories not being filled out

Modeling

- **What modeling techniques are most appropriate for your problem?**
    - Linear Regression will probably be the most valuable
- **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 would be enjoyability. Basing my recommendations off of Valence/Arousal, meaning how much a track can be enjoyed based on mood.
- **Is this a regression or classification problem?**
    - This is a regression problem

Evaluation

- **What metrics will you use to determine success (MAE, RMSE, Accuracy, Precision etc.)?**
    - Valence/Arousal

Tools/Methodologies

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