Capstone Project Proposal

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

Music Prediction

Business Understanding

- What problem are you trying to solve, or what question are you trying to answer?
    - Which components affect cater of larger audience
- What industry/realm/domain does this apply to?
    - It primary apply to music industry, but it can apply to film industry since music is utilized as a big component.
- What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)
    - I am interested in music and learned music almost for decades. And I am also interested in AI and data science. So, combination of two attract attention.

Data Understanding

- What data will you collect?
    - Data related to billboard top 100 of recent decades. Data related to celebrity
- Is there a plan for how to get the data (API request, direct download, etc.)?
    - Direct download from Kaggle
- What are the features you'll be using in your model?
    - Components of music such as danceability, energy, loudness, tempo, artist, etc.

Data Preparation

- What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?
    - There is no need for encoding nor matrix transformations. There were some columns that has null values when it should not, and some columns were in different types such as string when it is date or integer.
- What are some of the cleaning/pre-processing challenges for this data?
    - There are multiple duplicates for each music since some music were in Billboard top 100 for multiple times. And as merging two data set together, titles were in different format.

Modeling

- What modeling techniques are most appropriate for your problem?
    - Multiple linear regression model as it has multiple features to compare
- What is your target variable? (remember - we require that you answer/solve a supervised problem for the capstone, thus you will need a target)
    - Predictors would be features (components of music) whereas target would be chosen between artist and track
- Is this a regression or classification problem?
    - Regression problem

Evaluation

- What metrics will you use to determine success (MAE, RMSE, etc.)?
    - Because most features are in range between 0 to 1, use MAE

Tools/Methodologies

- What modeling algorithms are you planning to use (i.e., decision trees, random forests, etc.)?
    - As the data has multiple columns (features), use multiple linear regression to find the relationship between predictors and target features.