

# 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](#)

## Spotify Unwrapped

### Business Understanding

- What problem are you trying to solve, or what question are you trying to answer?  
[Predicts of next year Spotify Wrapped of users \( so they can compare with their Wrapped later next year \). Cannot predict top song \( because it is not out yet \) but can still predict the top genres or top artists \( with data of new album releases next year for example\). Can we also predict the number of minutes on Spotify also?](#)
- What industry/realm/domain does this apply to?  
[Digital Music/Music](#)
- What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)  
[The idea behind Spotify Wrapped 2022 being everywhere on the internet and I was wondering if I we can predict the next year Spotify Wrapped of User.](#)

### Data Understanding

- What data will you collect?  
[Music data from Spotify](#)  
[Spotify 2022 Wrapped from Spotify](#)
- Is there a plan for how to get the data (API request, direct download, etc.)?

Import Spotify API

- Are the features that will be used described clearly  
Yes. Correlation coefficient will be calculated

### Data Preparation

- What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?

Classify the variables in the data into types, encoding to transform data to machine-readable features, maybe sentiment analysis

- What are some of the cleaning/pre-processing challenges for this data?

The data would have various attributes which were not relevant like BPM, Liveness etc. Also dropping duplicate songs

### Modeling

- What modeling techniques are most appropriate for your problem? [Clustering](#)
- What is your target variable? (remember - we require that you answer/solve a supervised problem for the capstone, thus you will need a target)  
[Top artist, top genres](#)
- Is this a regression or classification problem? [Classification](#)

### Evaluation

- What metrics will you use to determine success (MAE, RMSE, etc.)?  
[Accuracy, Recall. Precision, ROC AUC](#)

### Tools/Methodologies

- What modeling algorithms are you planning to use (i.e., decision trees, random forests, etc.)? [Random Forest, Logistic Regression, K-Means](#)