The final report should be roughly 3-5 pages including the following 6 aspects:

1. Introduction/Executive Summary
   * synopsis of problem statement
     1. The goal of our project is to predict the genre and popularity of a song based on its audio features. Popularity prediction could help an artist select songs for an album, event, or a marketing campaign. By knowing which songs are likely to be popular, an artist can focus their attention and resources on the likely popular songs. Genre prediction can also assist musical artists by simply providing the genre classification information. A new artist may not be sure which genre they themselves belong to and could therefore use genre prediction to learn about themselves. On the other hand, an artist may be confident in their genre, but produce a song that is likely to be popular and in another genre. With that information, an artist could market a single song to a new audience.
   * data/benchmark description
   * main approach/algorithms used and result highlight on both accuracy and scalability/runtime
     1. Trained various models
        1. Brief preprocessing description
        2. Brief parameter tuning description
        3. Highlight the accuracy/RMSE and runtime of the best performing models
   * Include a subsection entitled: *Improvements after the last presentation*
     1. Popularity RMSE improvement
     2. Genre accuracy improvement
2. Details on problem statement and data/benchmark description, include references to kaggle problem and open datasets
   1. Problem statement details
   2. Data description
3. Details on the approaches, algorithms and programming tools used, include a description of the source code structure with end-to-end system diagram
   1. Programming Tools
   2. Models used
   3. Approach
      1. Data Exploration
      2. Preprocessing
      3. Train and Tune Models
         1. After the data has been preprocessed, model training can begin. Various models were selected and tested. Once the model is selected, it is improved through tuning. Parameters are selected based on our knowledge of the data. Parameters are hyper tuned with sklearns Grid Search and Random Search.
4. Details on the metrics and evaluation setup
   1. Popularity
      1. RMSE
      2. Baselines
   2. Genre
      1. Accuracy
         1. The accuracy is a ratio of correctly predicted observations. It is a straightforward metric. Accuracy score is easy to interpret but is not a good fit if there is class imbalance. However, our dataset has evenly distributed genres.
      2. Baselines
         1. The first baseline is a random guess. With 114 genres, the accuracy of a random guess is 0.88%. This is quite low as a baseline. Practically any chance of guessing the correct genre will outperform a random guess.
         2. In addition to the random guess sklearn’s Dummy Classifier was used. It also provided a low accuracy of 0.86%.
         3. The final baseline model is a Decision Tree Classifier with the default parameters. This yielded a more practical baseline with an accuracy of 16.8%. The baseline accuracy that we expect to improve upon is 16.8%
5. Detailed analysis on results, comparisons of different approaches
   1. Popularity
   2. Genre
6. Conclusion on challenges and lessons learned
   1. Challenges
   2. Lessons learned