**API 222 Analysis plan**

**Research Question 1** - What kind of engagement can upcoming songs draw, based on song characteristics like danceability, energy, loudness, instrumentalness, tempo, valence, etc.

* Methods (predictive modeling) -

Linear regression, random forests (Rohini)

LASSO, decision trees (Pranav)

* classification (hits v/s non hits)

**Research Question 2** - How can songs be grouped into genres using the song characteristics to improve user experience or for targeted marketing by music companies?

* Methods - Clustering (K means or hierarchical) (cluster analysis)

**Flow of the presentation**

1. **Problem statement** 
   1. Why is it important? - done
   2. Who is it relevant for? - done
2. **Our approach**
   1. What is the prediction problem in this?
   2. Outcome variables of interest
   3. Predictors used
   4. Datasource
3. **Models chosen**
   1. Models chosen
   2. Rationale behind the models
4. **Results**
   1. Training MSE
   2. Test MSE
   3. Comparison between different models
   4. Interpretation
5. **Challenges and Limitations**
6. **References**