Questions management would like you to answer:

1) Consider the 10 song features duration, danceability, energy, loudness, speechiness, acousticness, instrumentalness, liveness, valence and tempo. Are any of these features reasonably distributed normally? If so, which one?

2) Is there a relationship between song length and popularity of a song? If so, is the relationship positive or negative?

3) Are explicitly rated songs more popular than songs that are not explicit?

4) Are songs in major key more popular than songs in minor key?

5) Energy is believed to largely reflect the “loudness” of a song. Can you substantiate (or refute) that this is the case?

6) Which of the 10 individual (single) song features from question 1 predicts popularity best? How good is this “best” model?

7) Building a model that uses *all* of the song features from question 1, how well can you predict popularity now? How much (if at all) is this model improved compared to the best model in question 6). How do you account for this?

8) When considering the 10 song features above, how many meaningful principal components can you extract? What proportion of the variance do these principal components account for?

9) Can you predict whether a song is in major or minor key from valence? If so, how good is this prediction? If not, is there a better predictor?

10) Which is a better predictor of whether a song is classical music – duration or the principal components you extracted in question 8?

Extra credit: Tell us something interesting about this dataset that is not trivial and not already part of an answer (implied or explicitly) to these enumerated questions.