The Neural Network: Katelyn Fu, Rahi Dasgupta, Edwin Lin

From our PM meeting, Ella gave us insights on why our code may not be running properly for the visualizations for last week's milestones. She hypothesized that our venv wasn't running properly so our imports weren't being recognized. She helped us with setting our venvs up and we fixed errors that our code had as well.

Because we weren't able to get our visualization done last week, we worked on that this week plus our linear regression and we weren't able to get to our web scraping. We will do that for next week's milestone though.
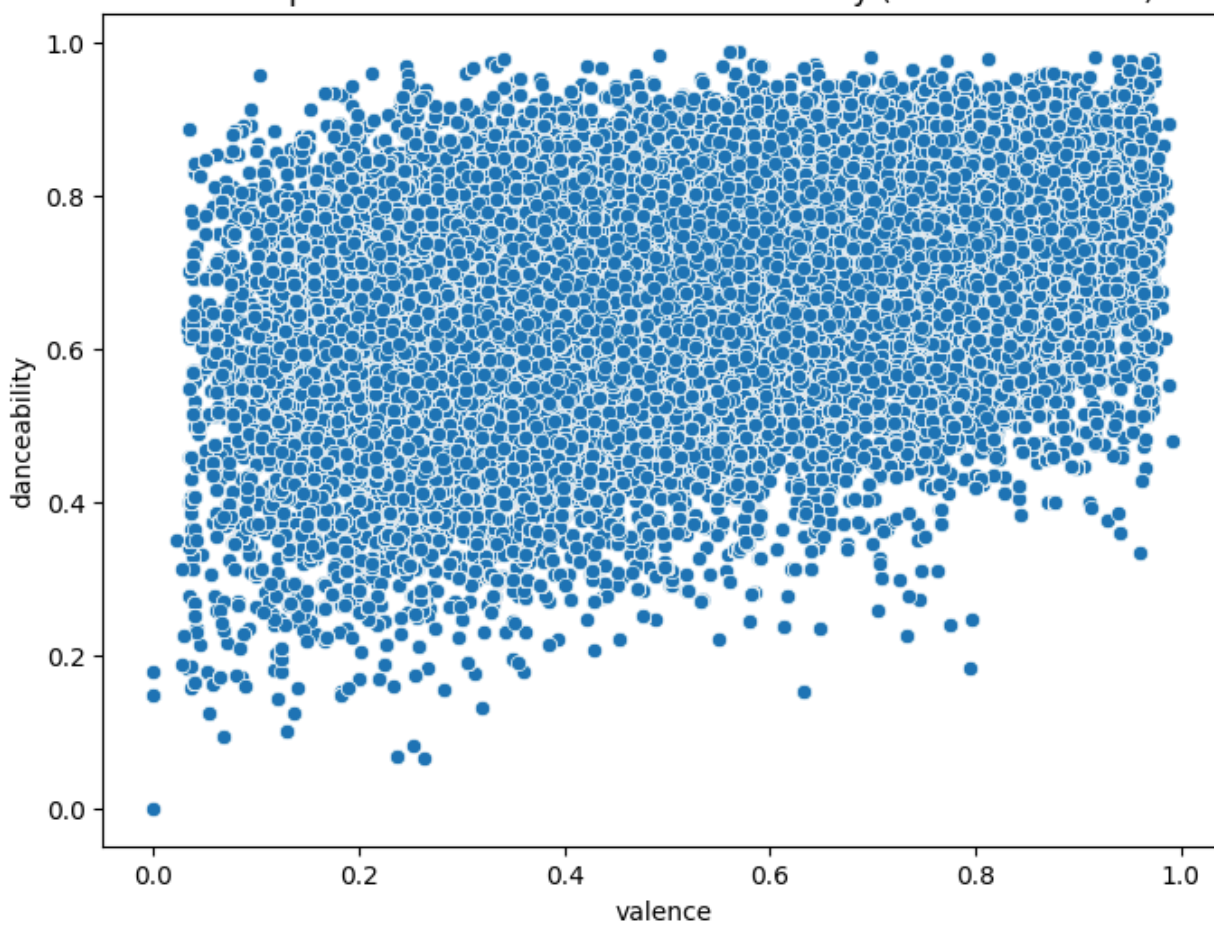
Correlation Matrix for popularity
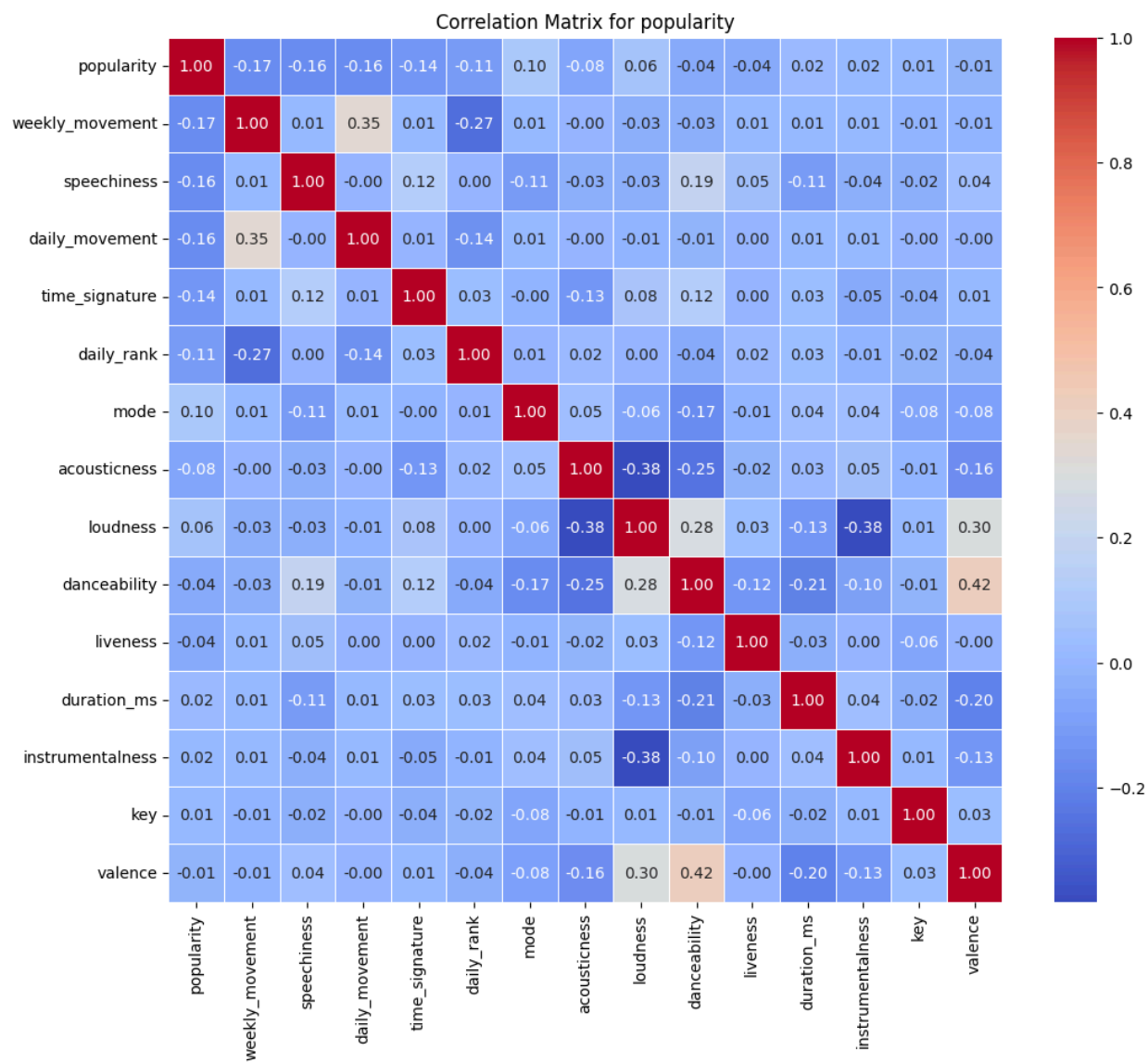
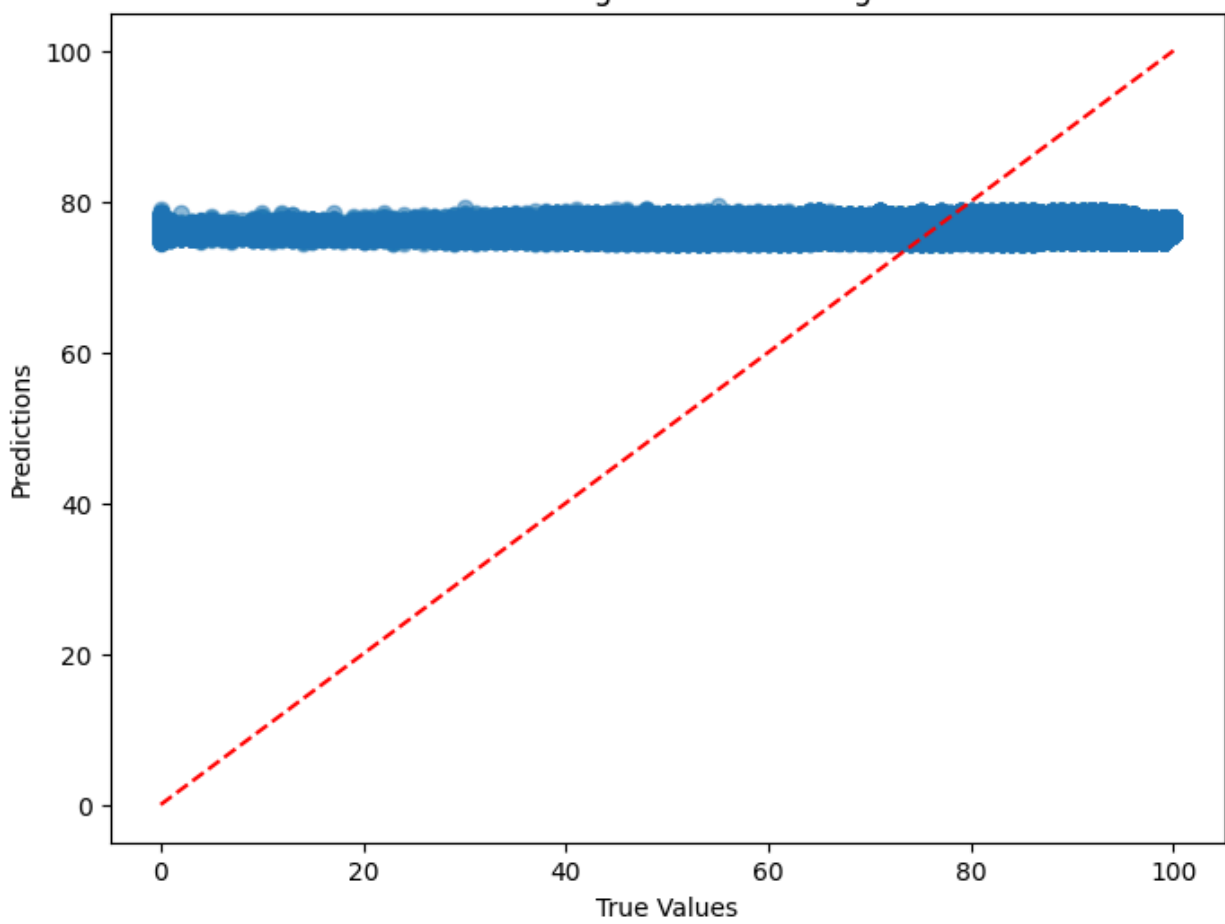Scatter plot between energy and loudness (Correlation: 0.68)
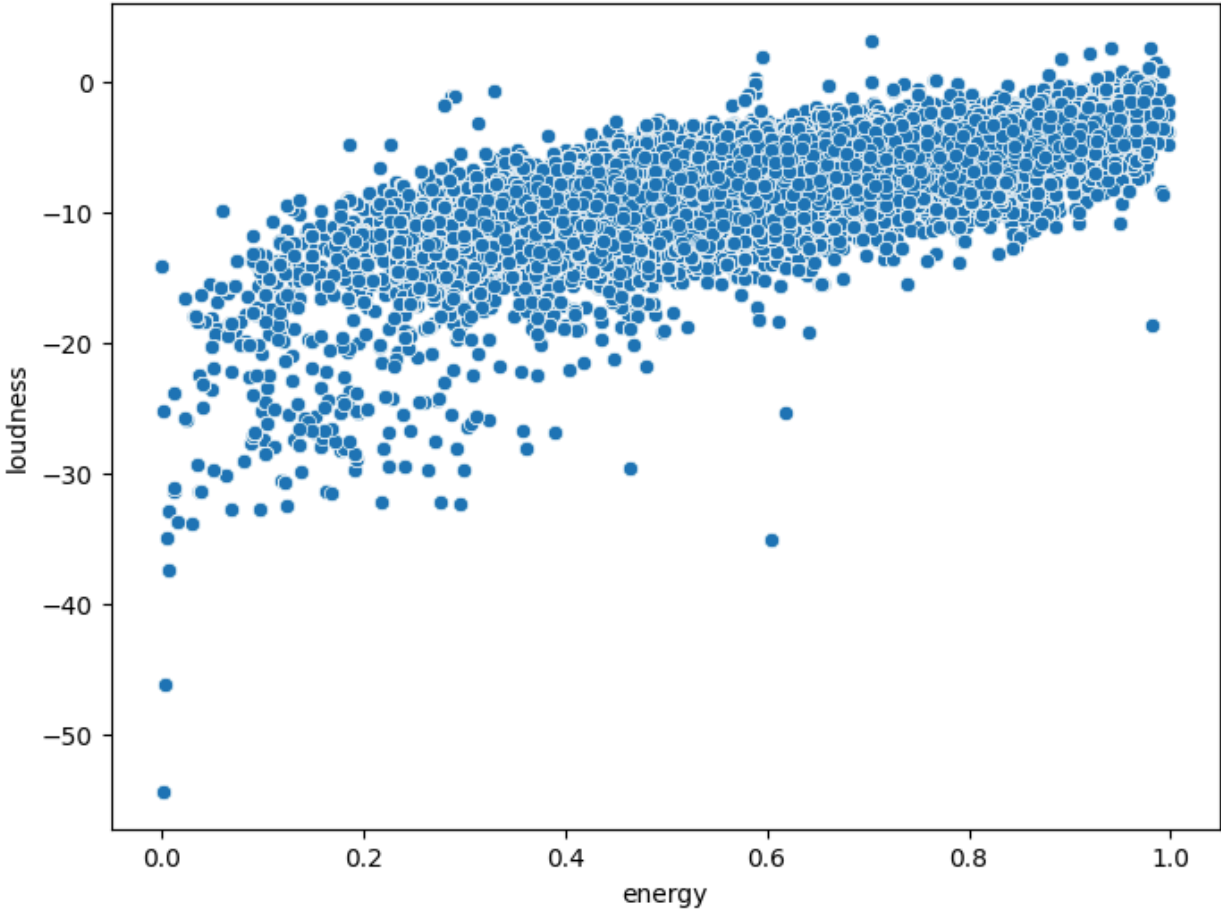
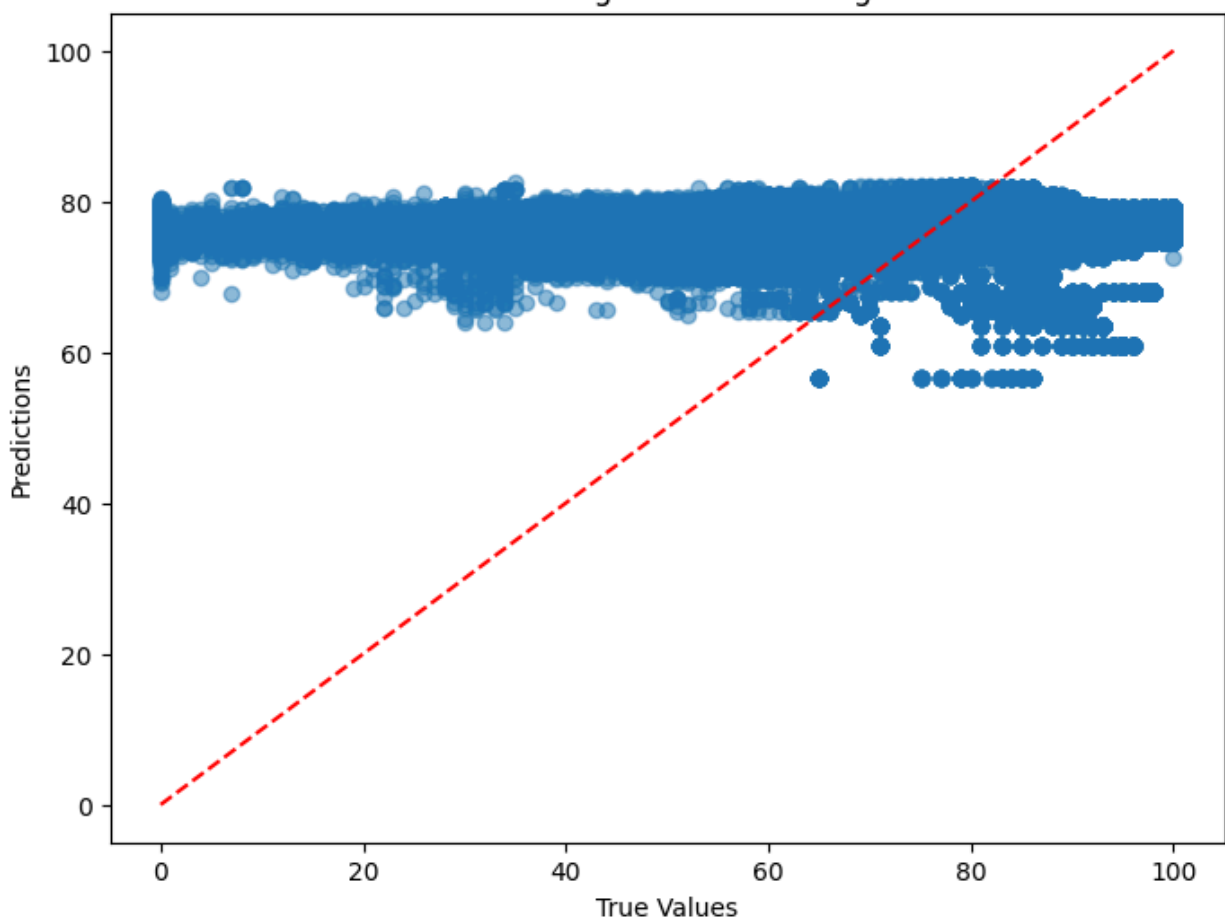Scatter plot between acousticness and energy (Correlation: 0.53)
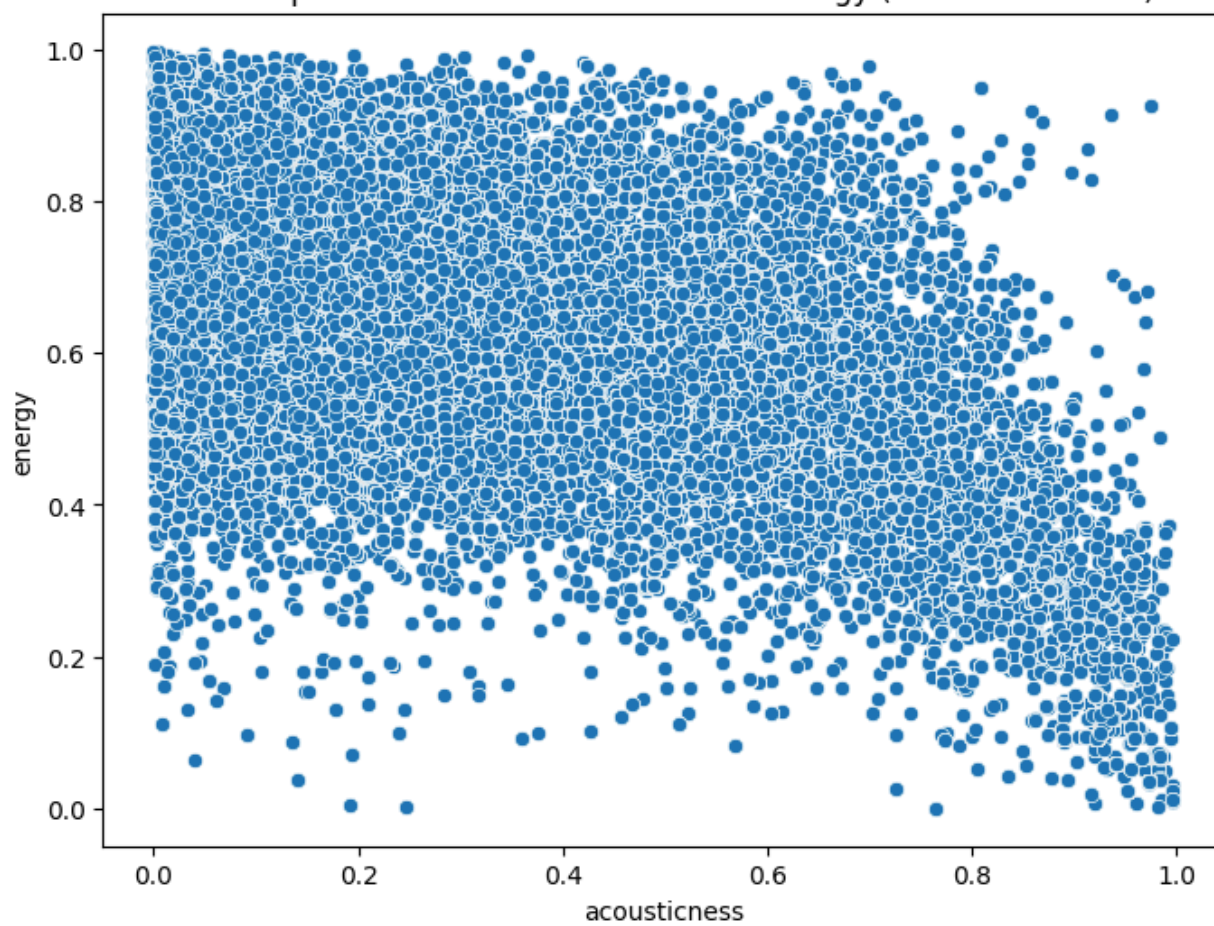
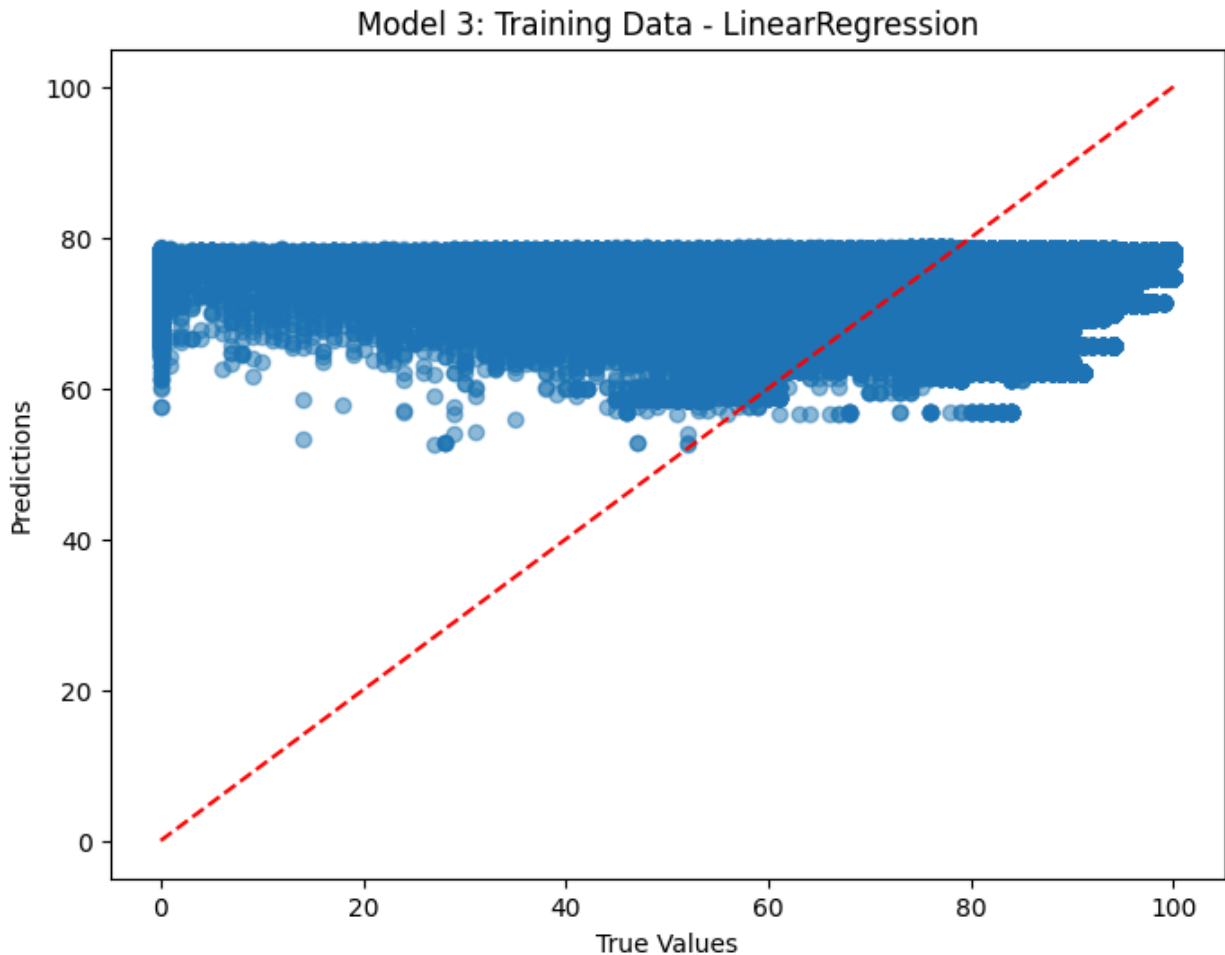Scatter plot between valence and danceability (Correlation: 0.42)

Model 1: Training Data - LinearRegression

Model 2: Training Data - LinearRegression

Model 3: Training Data - LinearRegression

What Did You Learn About Your Dataset?

1. Our correlation matrix showed us that danceability and loudness have the strongest positive correlations at 0.42 and 0.30 respectively. This led us to infer that the danceable songs that has more energy tend to be more popular on spotify.
2. Our scatterplots showed us that energy and loudness have a strong correlation at 0.68.
3. Some features like acousticness and speechiness showed weak negative correlations with popularity with both being at -0.16.

2. What Is the Question You Are Interested in Exploring?

1. The question that we are interested in exploring is which specific features have the highest impact on a song's popularity on Spotify's platform.

3. What Do You Need to Do to Start Answering It?

1. To start answering the question, we decided to do a linear regression and evaluate performance using the R^2 score and mean squared error to see how accurate our regression is.

What challenges arose this week?

1. This week, we were struggling with our code because it was new to all of us. We had to do a lot of research on how to train and test our data and write the code.

What went well?

1. This week, we were able to complete milestone two and get our graphs to see what patterns were in our data.
2. We also were able to properly set up our virtual environments so everything ran accordingly.

Katelyn Fu
Rahi Dasgupta
Edwin Lin