

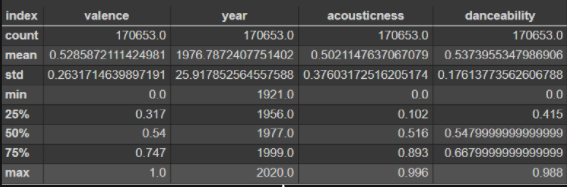
Machine Learning-Music Recommendation

Progress Report

| Roll No. | Name of the Student | Name of the Program |
| --- | --- | --- |
| AU1940142 | Shail | ML |
| AU1940267 | Sahil | ML |
| AU1940034 | Varun | ML |
| AU1940109 | Jap | ML |

**Features reduction and reducing the scale**

Based on the correlation graph and the non linear relation graph we have reduced the highly correlated variables. The threshold that we have kept is 0.75. I.e. if the correlation coefficient between two variables is greater than 0.75 we will be discarding one of the variables. We have also analyzed the non linear relationship between the variables and removed the highly dependent variables. Apart from this we have analyzed the spread of different variables and reduced the variables which are out of scale. We want all the variables to be normally distributed between the range of 0 to 1. For this reason we have used a standard scaler in order to reduce the scale of the variables.



As visible the year variable has a very high value and is out of range as compared to the other variables. We will be performing standard scalers on such variables.

**Future work**

Furthermore we will be working with developing a baseline KNN model. We will be performing dimensionality reduction techniques such as PCA and LDA. Since KNN works better in lower dimensions it is of crucial importance to work with lower dimensions. Thus using PCA and LDA we will be reducing the dimension space to the top 3-4 variables based on the eigenvalues and vector.