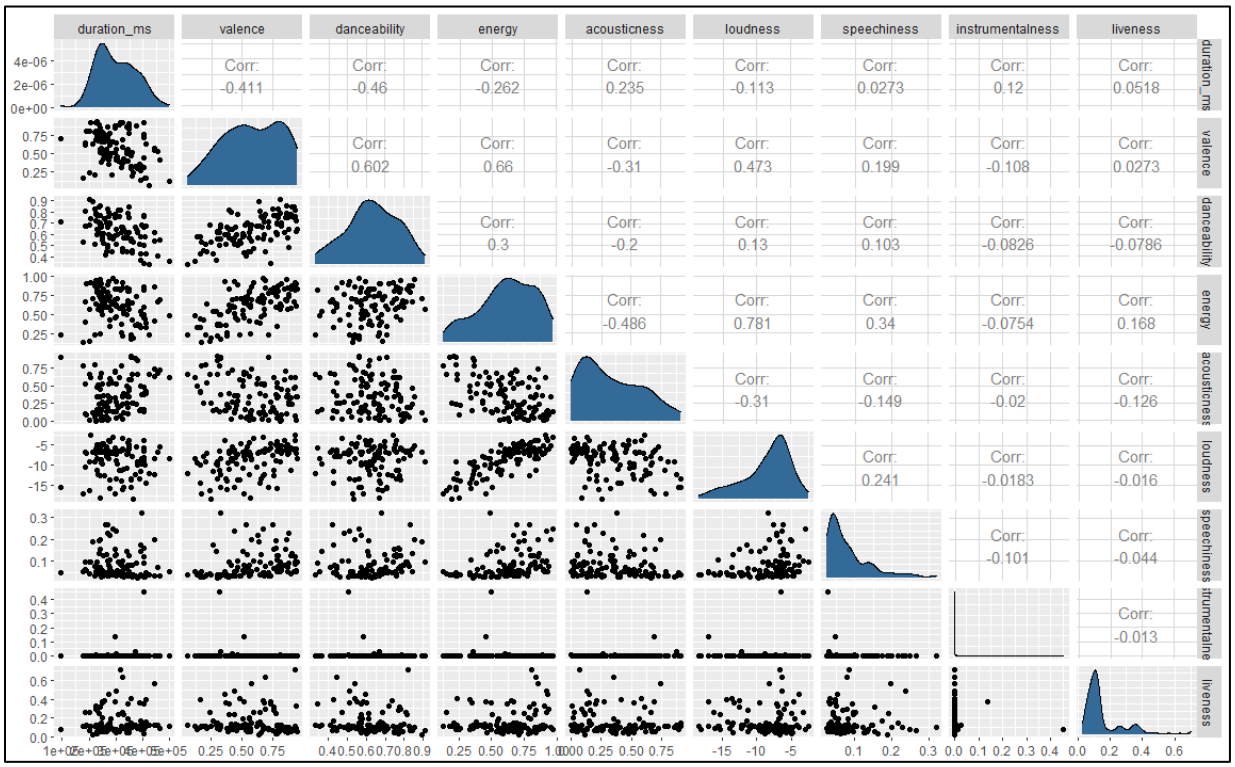
Spotify Artist Analysis

I. Introduction

In this study of Spotify Artists, I am choosing “Avadhoot Gupte” as the study artist because I am a big fan of him & his rock songs. Avadhoot Gupte is an Indian music composer and singer who is popularly known for his work in the Marathi film and music industry. His songs generally tend to be loud, energetic, and high-pitched yet melodious. Hence, I expect the same from the statistical analysis on the data as performed further.

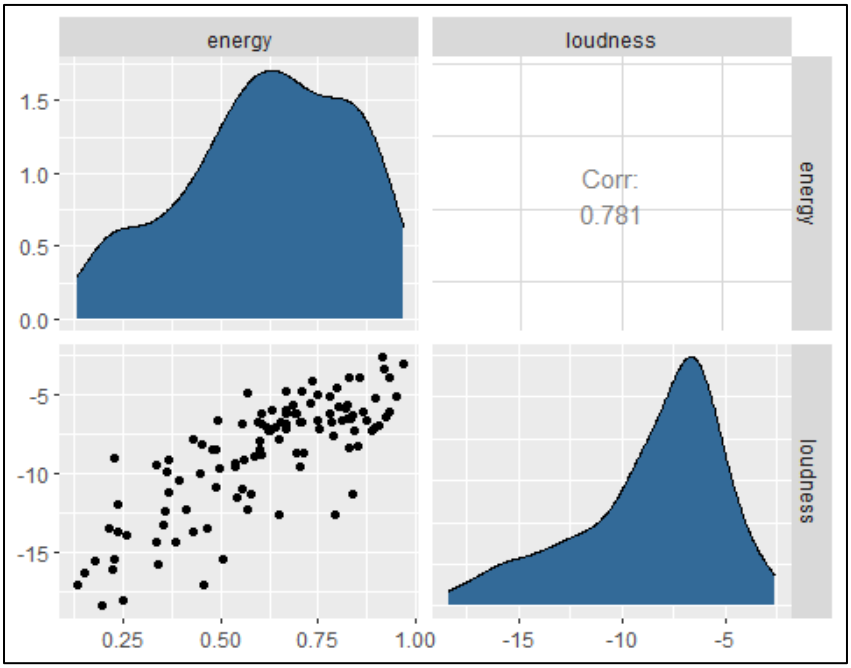
II. Exploratory Data Analysis

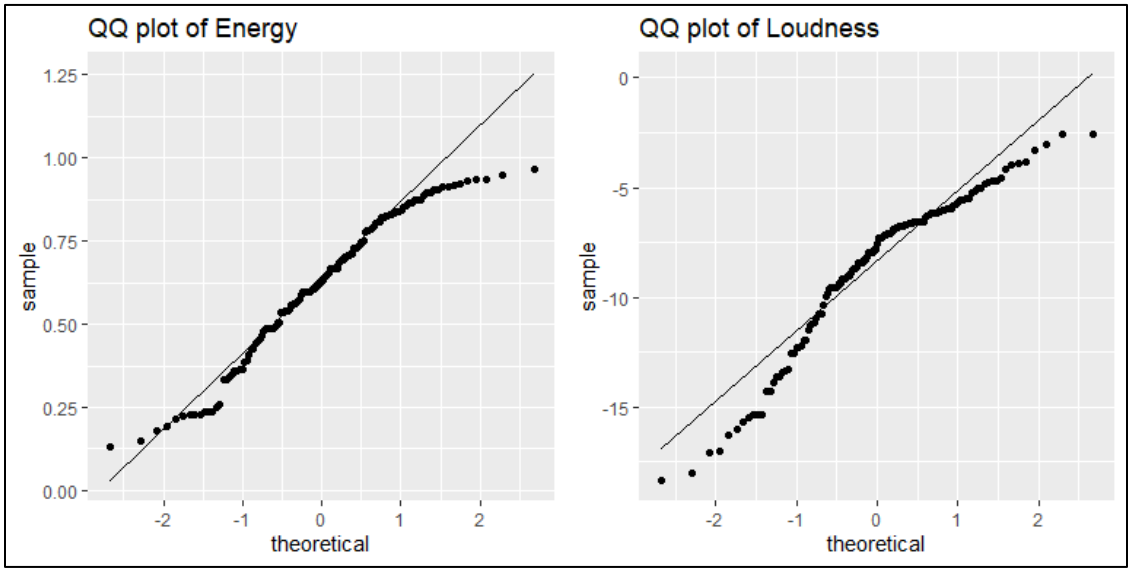



```
##                  duration_ms     valence danceability      energy acousticness
## duration_ms       1.00000000 -0.41087855  -0.46027750 -0.26212269   0.23511936
## valence          -0.41087855  1.00000000   0.60212520  0.65988338  -0.30950750
## danceability     -0.46027750  0.60212520   1.00000000  0.30016741  -0.20040468
## energy           -0.26212269  0.65988338   0.30016741  1.00000000  -0.48575884
## acousticness      0.23511936 -0.30950750  -0.20040468 -0.48575884   1.00000000
## loudness         -0.11318618  0.47296480   0.12987308  0.78132446  -0.31028024
## speechiness       0.02731319  0.19929889   0.10335806  0.34036356  -0.14919606
## instrumentalness  0.11998035 -0.10773912  -0.08255963 -0.07542802  -0.02002188
## liveness          0.05183406  0.02731553  -0.07862501  0.16766006  -0.12576295
```

```
##                     loudness speechiness instrumentalness    liveness
## duration_ms      -0.11318618  0.02731319       0.11998035  0.05183406
## valence           0.47296480  0.19929889      -0.10773912  0.02731553
## danceability      0.12987308  0.10335806      -0.08255963 -0.07862501
## energy            0.78132446  0.34036356      -0.07542802  0.16766006
## acousticness     -0.31028024 -0.14919606      -0.02002188 -0.12576295
## loudness          1.00000000  0.24146570      -0.01830474 -0.01596007
## speechiness       0.24146570  1.00000000      -0.10065282 -0.04401879
## instrumentalness -0.01830474 -0.10065282       1.00000000 -0.01296522
## liveness         -0.01596007 -0.04401879      -0.01296522  1.00000000
```

From the above graph and correlation matrix, we can say that the attributes 'energy' & 'loudness' have a high linear correlation. Thus, we will be going forward with those 2 attributes.





The plot suggests that the two variables are almost normally distributed. Using the QQ plot, it can be concluded that the Energy variable is more normal as compared to the Loudness variable.

Also, as mentioned previously that this artist has high energetic songs, this is reflected in the density plot of the energy where the plot tends to be left-skewed (dense towards the right-hand side where the energy is more). On similar lines, loudness also follows the same trend of being left-skewed. This pattern between energy and loudness is rightly anticipated as more the energy more will be the loudness. Hence, these numerical measurements match up with my opinions about the artist.

III. Statistical Analyses

1. Confidence Intervals of 'Energy' & 'Loudness'

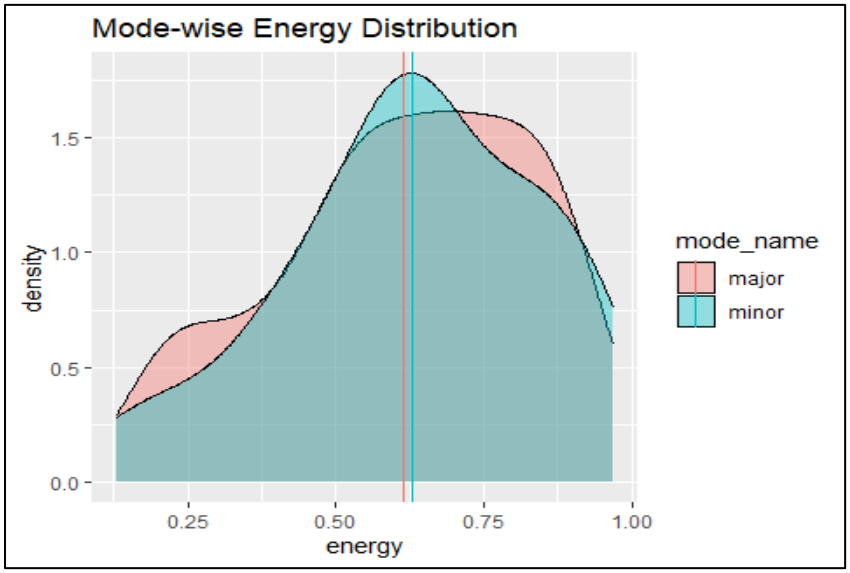
```
## [1] 0.5705923 0.6655841
## attr(,"conf.level")
## [1] 0.99

## [1] -9.362246 -7.803519
## attr(,"conf.level")
## [1] 0.99
```

The confidence intervals of energy lie between 0.5738 to 0.67208. This goes according to the trend about the artist where his songs tend to be more energetic. Hence, the confidence interval lies above 50% of the entire energy range(0 to 1). Also, the confidence interval of loudness ranges from -9.3768 to -7.7566, which suggests that most of the songs are loud enough. Hence, they lie on the upper loudness band on the scale of -60 to 0 dB.

These numbers are also evident from the above density plot; therefore, these confidence intervals describe the overall nature of the artist's music.

2. Two-sample Hypothesis Test on 'Energy'



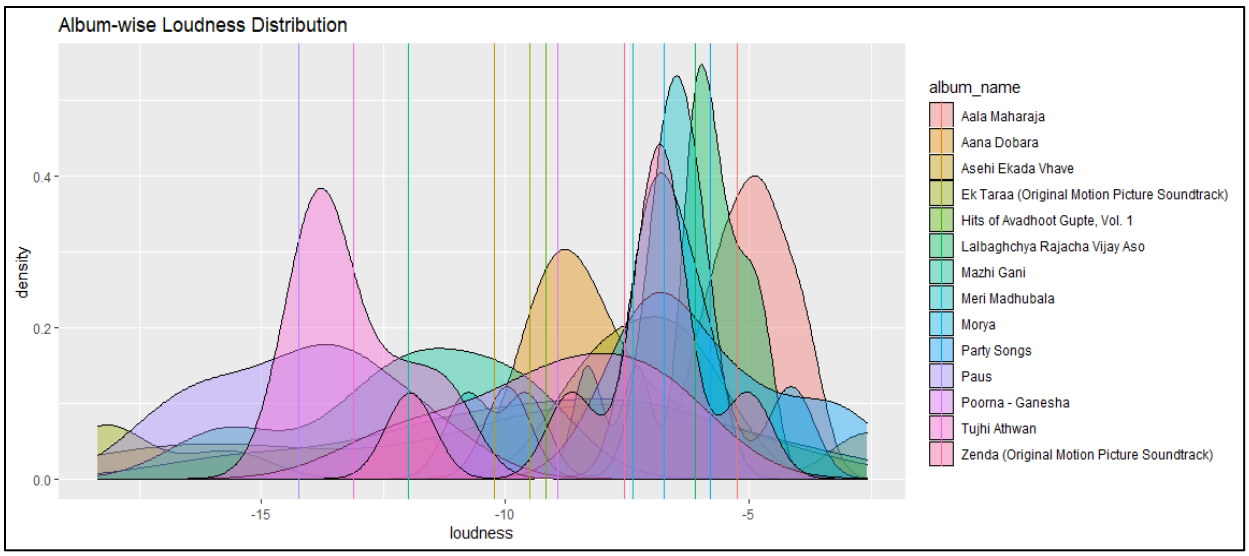
```
##
##  Welch Two Sample t-test
##
## data:  major_data and minor_data
## t = -0.3567, df = 56.137, p-value = 0.7227
## alternative hypothesis: true difference in means is not equal to 0
## 99 percent confidence interval:
##  -0.12795400  0.09775792
## sample estimates:
## mean of x mean of y
## 0.6143137 0.6294118
```

The p-value of the Welch Two Sample t-test states that we do not have enough evidence to reject the null hypothesis. Thus, we accept the null hypothesis where the means of the two groups are the same/similar (Significant level = 0.01). This is also evident from the fact that the energy of the song would not have much impact on the song modality (ie. melodic content). Tracks with Major mode (such as Celebrations, Happy songs) have both high and low energy. This is like the Minor mode songs (Eg. – Sorowful songs) where both energy levels are prevalent. Hence, the means of the two-mode groups almost overlap with each other implying that the energy distributions of the two modes follow the same/similar distribution (irrespective of the mode).

3. ANOVA & Pairwise Hypothesis Test

```
## Analysis of Variance Table
##
## Response: loudness
##             Df Sum Sq Mean Sq F value    Pr(>F)
## album_name  13 855.98  65.845   10.33 2.048e-14 ***
## Residuals  122 777.67   6.374
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA test on the loudness parameter with respect to album name says that there is at least 1 pair of albums which have different sample means ie. not all albums follow the null hypothesis (ie. two means are not equal). This can be concluded by observing the p-value = 1.899e-14. This provides significant evidence that not all means are equal. Hence, we statistically reject the Null hypothesis (Significance Level = 0.01).




```
##                                                                                                                                  p.adj
## Aana Dobara-Aala Maharaja                                                                                                 5.413621e-01
## Asehi Ekada Vhave-Aala Maharaja                                                                                           7.984598e-03
## Ek Taraa (Original Motion Picture Soundtrack)-Aala Maharaja                                                               8.122720e-03
## Hits of Avadhoot Gupte, Vol. 1-Aala Maharaja                                                                              7.929969e-02
## Lalbaghchya Rajacha Vijay Aso-Aala Maharaja                                                                               9.999781e-01
## Mazhi Gani-Aala Maharaja                                                                                                  1.119830e-06
## Meri Madhubala-Aala Maharaja                                                                                              7.035693e-01
## Morya-Aala Maharaja                                                                                                       9.951682e-01
## Party Songs-Aala Maharaja                                                                                                 9.999999e-01
## Paus-Aala Maharaja                                                                                                        1.038624e-09
## Poorna - Ganesha-Aala Maharaja                                                                                            2.310298e-01
## Tujhi Athwan-Aala Maharaja                                                                                                1.620011e-06
## Zenda (Original Motion Picture Soundtrack)-Aala Maharaja                                                                  8.465743e-01
## Asehi Ekada Vhave-Aana Dobara                                                                                             5.122979e-01
## Ek Taraa (Original Motion Picture Soundtrack)-Aana Dobara                                                                 7.034700e-01
## Hits of Avadhoot Gupte, Vol. 1-Aana Dobara                                                                                9.686952e-01
## Lalbaghchya Rajacha Vijay Aso-Aana Dobara                                                                                 9.765351e-01
## Mazhi Gani-Aana Dobara                                                                                                    9.248950e-04
## Meri Madhubala-Aana Dobara                                                                                                1.000000e+00
## Morya-Aana Dobara                                                                                                         9.999754e-01
## Party Songs-Aana Dobara                                                                                                   8.804769e-01
## Paus-Aana Dobara                                                                                                          6.643798e-07
## Poorna - Ganesha-Aana Dobara                                                                                              9.970054e-01
## Tujhi Athwan-Aana Dobara                                                                                                  6.651303e-04
## Zenda (Original Motion Picture Soundtrack)-Aana Dobara                                                                    1.000000e+00
## Ek Taraa (Original Motion Picture Soundtrack)-Asehi Ekada Vhave                                                           9.999965e-01
## Hits of Avadhoot Gupte, Vol. 1-Asehi Ekada Vhave                                                                          9.999109e-01
## Lalbaghchya Rajacha Vijay Aso-Asehi Ekada Vhave                                                                           8.167360e-02
## Mazhi Gani-Asehi Ekada Vhave                                                                                              9.743721e-01
## Meri Madhubala-Asehi Ekada Vhave                                                                                          4.288307e-01
## Morya-Asehi Ekada Vhave                                                                                                   3.645466e-01
## Party Songs-Asehi Ekada Vhave                                                                                             3.305955e-02
## Paus-Asehi Ekada Vhave                                                                                                    1.296792e-01
## Poorna - Ganesha-Asehi Ekada Vhave                                                                                        9.996523e-01
## Tujhi Athwan-Asehi Ekada Vhave                                                                                            7.294959e-01
## Zenda (Original Motion Picture Soundtrack)-Asehi Ekada Vhave                                                              7.796936e-01
## Hits of Avadhoot Gupte, Vol. 1-Ek Taraa (Original Motion Picture Soundtrack)                                              1.000000e+00
## Lalbaghchya Rajacha Vijay Aso-Ek Taraa (Original Motion Picture Soundtrack)                                               1.185960e-01
## Mazhi Gani-Ek Taraa (Original Motion Picture Soundtrack)                                                                  4.765261e-01
## Meri Madhubala-Ek Taraa (Original Motion Picture Soundtrack)                                                              6.053574e-01
## Morya-Ek Taraa (Original Motion Picture Soundtrack)                                                                       5.427775e-01
## Party Songs-Ek Taraa (Original Motion Picture Soundtrack)                                                                 4.199167e-02
```

```
## Paus-Ek Taraa (Original Motion Picture Soundtrack)                                          4.153032e-03
## Poorna - Ganesha-Ek Taraa (Original Motion Picture Soundtrack)                              9.999999e-01
## Tujhi Athwan-Ek Taraa (Original Motion Picture Soundtrack)                                  1.899617e-01
## Zenda (Original Motion Picture Soundtrack)-Ek Taraa (Original Motion Picture Soundtrack) 9.339304e-01
## Lalbaghchya Rajacha Vijay Aso-Hits of Avadhoot Gupte, Vol. 1                                4.200440e-01
## Mazhi Gani-Hits of Avadhoot Gupte, Vol. 1                                                   4.756312e-01
## Meri Madhubala-Hits of Avadhoot Gupte, Vol. 1                                               9.377142e-01
## Morya-Hits of Avadhoot Gupte, Vol. 1                                                        8.507284e-01
## Party Songs-Hits of Avadhoot Gupte, Vol. 1                                                  2.353684e-01
## Paus-Hits of Avadhoot Gupte, Vol. 1                                                         7.264032e-03
## Poorna - Ganesha-Hits of Avadhoot Gupte, Vol. 1                                             1.000000e+00
## Tujhi Athwan-Hits of Avadhoot Gupte, Vol. 1                                                 1.914000e-01
## Zenda (Original Motion Picture Soundtrack)-Hits of Avadhoot Gupte, Vol. 1                   9.942165e-01
## Mazhi Gani-Lalbaghchya Rajacha Vijay Aso                                                    7.140744e-05
## Meri Madhubala-Lalbaghchya Rajacha Vijay Aso                                                9.939272e-01
## Morya-Lalbaghchya Rajacha Vijay Aso                                                         9.999997e-01
## Party Songs-Lalbaghchya Rajacha Vijay Aso                                                   1.000000e+00
## Paus-Lalbaghchya Rajacha Vijay Aso                                                          8.482440e-08
## Poorna - Ganesha-Lalbaghchya Rajacha Vijay Aso                                              6.813683e-01
## Tujhi Athwan-Lalbaghchya Rajacha Vijay Aso                                                  5.044883e-05
## Zenda (Original Motion Picture Soundtrack)-Lalbaghchya Rajacha Vijay Aso                    9.966905e-01
## Meri Madhubala-Mazhi Gani                                                                   6.755428e-04
## Morya-Mazhi Gani                                                                            2.704718e-03
## Party Songs-Mazhi Gani                                                                      1.100302e-05
## Paus-Mazhi Gani                                                                             8.149565e-01
## Poorna - Ganesha-Mazhi Gani                                                                 4.924544e-01
## Tujhi Athwan-Mazhi Gani                                                                     9.998088e-01
## Zenda (Original Motion Picture Soundtrack)-Mazhi Gani                                       2.691734e-02
## Morya-Meri Madhubala                                                                        9.999990e-01
## Party Songs-Meri Madhubala                                                                  9.520324e-01
## Paus-Meri Madhubala                                                                         5.213540e-07
## Poorna - Ganesha-Meri Madhubala                                                             9.914509e-01
## Tujhi Athwan-Meri Madhubala                                                                 4.832559e-04
## Zenda (Original Motion Picture Soundtrack)-Meri Madhubala                                   1.000000e+00
## Party Songs-Morya                                                                           9.999598e-01
## Paus-Morya                                                                                  6.456629e-06
## Poorna - Ganesha-Morya                                                                      9.546089e-01
## Tujhi Athwan-Morya                                                                          1.092418e-03
## Zenda (Original Motion Picture Soundtrack)-Morya                                            9.999974e-01
## Paus-Party Songs                                                                            1.064799e-08
## Poorna - Ganesha-Party Songs                                                                4.857309e-01
## Tujhi Athwan-Party Songs                                                                    1.111760e-05
## Zenda (Original Motion Picture Soundtrack)-Party Songs                                      9.778764e-01
## Poorna - Ganesha-Paus                                                                       1.132698e-02
## Tujhi Athwan-Paus                                                                           9.999063e-01
## Zenda (Original Motion Picture Soundtrack)-Paus                                             1.037719e-04
## Tujhi Athwan-Poorna - Ganesha                                                               2.040915e-01
## Zenda (Original Motion Picture Soundtrack)-Poorna - Ganesha                                 9.994204e-01
## Zenda (Original Motion Picture Soundtrack)-Tujhi Athwan                                     9.263560e-03
```

From the above graph and Tukey HSD (Honestly Significant Difference) test results, states that there are certain albums whose sample means are the same/similar and thus tend to overlap. On the other hand, there are certain albums whose means are much far away and thus have lesser p-values.

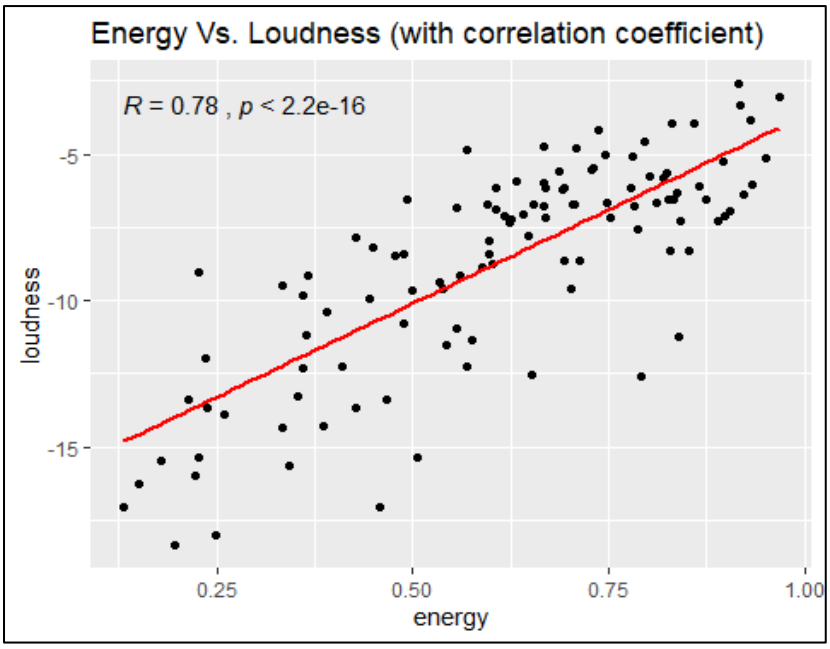
These results make sense because not all albums have the same loudness levels. There are certain albums that are totally loud (such as Aala Maharaja, Morya, Party Songs, etc.) whereas there is another category of albums that are relatively less loud (such as Paus, Tujhi Athwan, Mazhi Gani, etc.). The p-value of two means of polar loudness categories would be nearly 0 (Eg. p-value between Aala Maharaja & Mazhi Gani is 0.0000012), thus rejecting the null hypothesis. Conversely, the p-value of the same/similar loudness category albums has relatively higher p-values (Eg. p-value of Morya & Aala Maharaja is 0.9928106), thus

accepting the null hypothesis. These observations are as per the expectations and hence makes a lot of sense.

4. Linear Regression

Now, we fit a linear model of energy Vs. loudness. We expect the relationship to be positively linear, which means that as the energy increases loudness also increases. I felt that 'energy' is the independent variable and 'loudness' is the dependent variable since the energy level of the song decides its loudness. Hence, I chose predictor & response variables as 'energy' & 'loudness' respectively.

a) Scatterplot of the variables with the correlation between them



The variables energy & loudness follow the positive linear trend with some noise around the regression line. Thus, the correlation between them is 0.78, which is higher (ie. near to 1).

b) Fitting linear regression model and interpreting the slope & intercept

```
##
## Call:
## lm(formula = loudness ~ energy, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.4248 -1.2854  0.2168  1.5417  4.5619
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -16.5073     0.5779  -28.57   <2e-16 ***
## energy       12.8208     0.8847   14.49   <2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.179 on 134 degrees of freedom
## Multiple R-squared:  0.6105, Adjusted R-squared:  0.6076
## F-statistic:   210 on 1 and 134 DF,  p-value: < 2.2e-16
```

From the above results, we can interpret that the slope of the linear equation is 12.8208, and the intercept is -16.5073. Thus, the linear equation is –

$$y\_hat = (12.8208 * x) - 16.5073$$

The y-intercept & slope of the line is the parameters used for fitting the line well according to the data by minimizing the Sum of Squared Errors (SSE). The slope does not make sense as both the variables used are not scaled on the same scale. Thus, scaling will change the slope of the line. Also, the y-intercept suggests that even when energy is 0, loudness would be -16.5073 (which would be impractical). Hence, these parameters are subjective to the line fitting for the data.

c) Performing hypothesis test to test whether 'energy' is the good predictor

```
##
##  Welch Two Sample t-test
##
## data:  data$loudness and pred_loudness
## t = -1.4078e-14, df = 255.08, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 99 percent confidence interval:
##  -0.9824214  0.9824214
## sample estimates:
## mean of x mean of y
## -8.582882 -8.582882
```

The above t-test results with p-value = 1 suggest that the actual and predicted values of the loudness variable follow the same distribution. This provides the strong evidence that the two means are equal, and we accept the null hypothesis. Hence, the predictor variable 'energy' is a good predictor in the linear model.

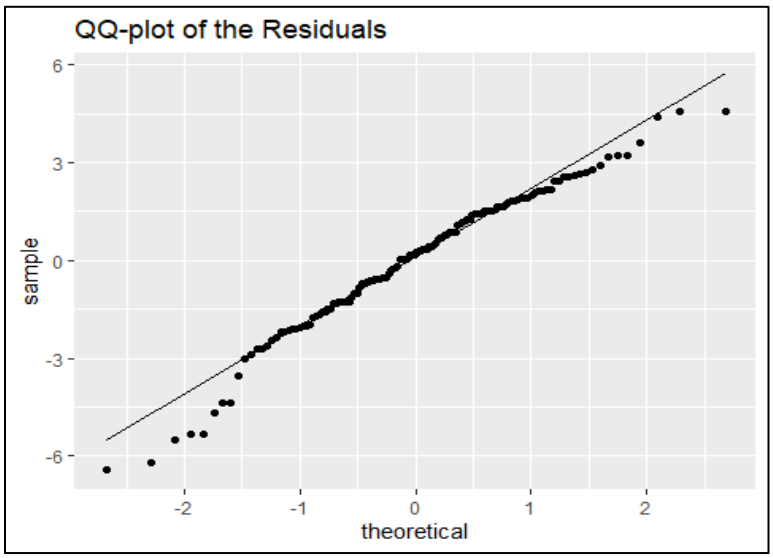
d) Predictions at 20th, 40th, 60th & 80th predictor values

```
##     energy loudness_predictions
## 20%  0.445           -10.802013
## 40%  0.588            -8.968638
## 60%  0.693            -7.622453
## 80%  0.826            -5.917286
```

The above table summarizes predicted loudness values for each percentile values of energy. From the table, we can observe that as the percentile increases energy value also increases with an increase in loudness predictions. This is because of the positive correlation between the variables: energy and loudness.

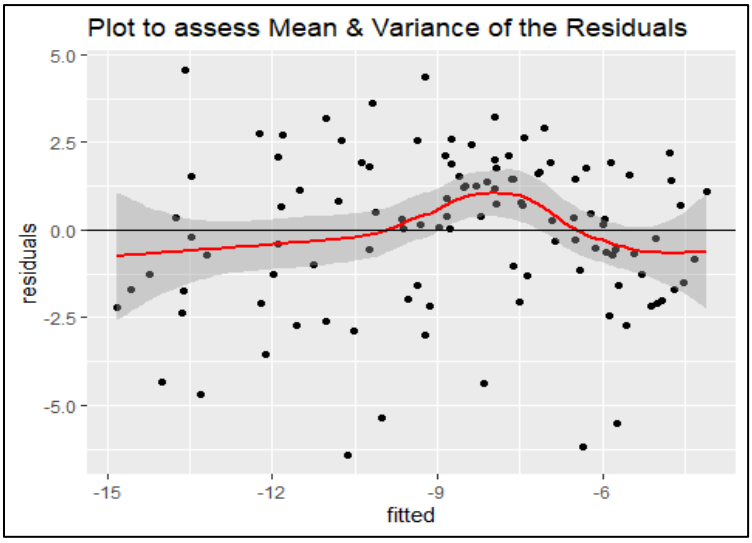
e) Checking the regression assumptions

i) Normality of the residuals



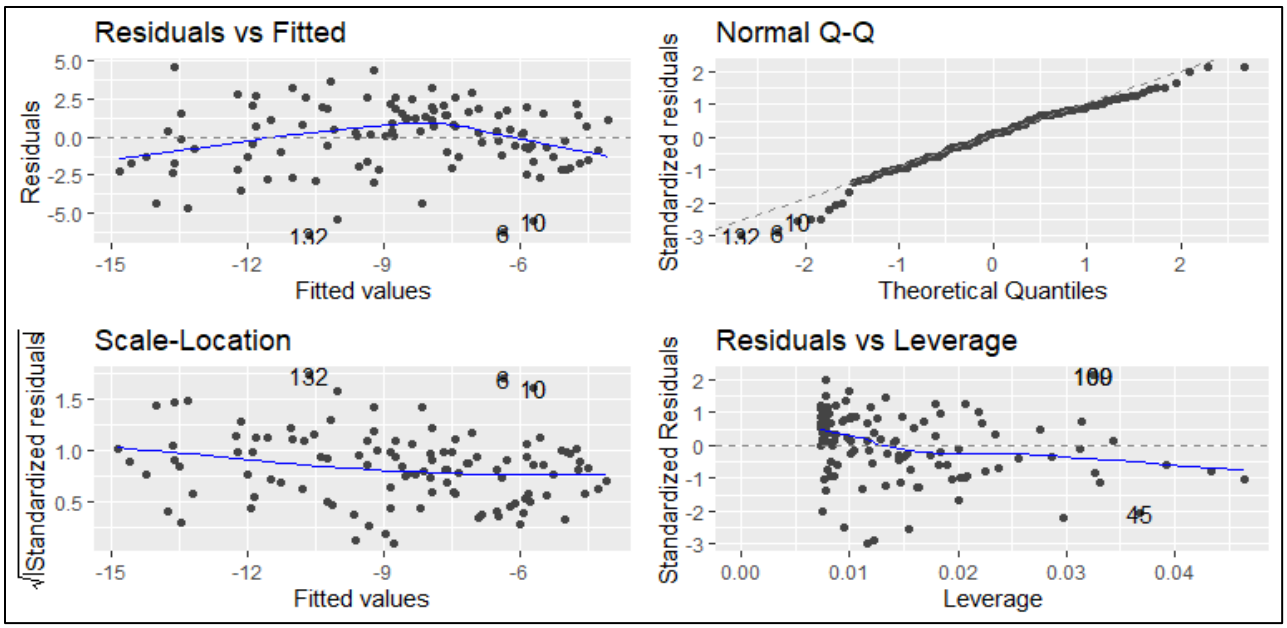
From the QQ-plot, we can interpret that the residuals are normally distributed. There is some deviation at the tails, but that is expected. Essentially there is not much of a pattern to the deviation from the line except for the tails. Hence, the residuals of our linear model follow the normality assumption.

ii) Assessing Mean & Variance of the residuals



The errors are normally distributed and thus the scatter plot has points homogenously distributed along the horizontal line (homoscedastic). Also, the red line almost follows the black line which means that the mean is approximately 0 across the regression line. Also, the plot does not showcase much larger deviations which indicates that the variance is constant across the whole plot.

Thus, we can confidently say that the linear model fitted using 'energy' and 'loudness' is valid based on the statistical inferences derived from the assumptions of the residuals.



From the above "Residuals vs Leverage" plot, we can say that both residuals and leverage are low and thus the model fits well with the predictor variable 'energy' to predict the dependent variable 'loudness'.

IV. Conclusion

In this study, we analyzed different parameters of the Spotify data for the artist "Avadhoot Gupte" and we found that attributes 'energy' & 'loudness' have the largest linear correlation. By knowing the artist, I can say that most of the songs in his albums are rock songs and thus have high energy & loudness. The results found out from different tests in this study are in line with our expectations. We observed that there is a positive linear relationship between energy and loudness. Moreover, the energy & loudness concentrate more towards the upper extreme values (indicating higher energy & loudness levels). This is also evident from the confidence intervals of the two variables.

It is also observed that the energy remains unchanged irrespective of the modes in the albums. However, the loudness levels change with the change in album names because all albums do not have uniform loudness levels. Some albums are comparatively louder than the rest of the albums. The Linear relation perfectly predicts the loudness levels using the predictor variable energy. Hence, the linear model used is valid and predictors are good in predicting the results with minimal errors.

V. Appendix (R code)

```
```{r setup, include = FALSE}
knitr::opts_chunk$set(echo = FALSE)
library(tidyverse)
library(spotifyr)
library(mosaic)
library(GGally)
library(ggpubr)
library(ggfortify)
library(gridExtra)
```

```{r}
Sys.setenv(SPOTIFY_CLIENT_ID = "b6eb08407bab41b4b3f8f5bb48a0f8f0")
Sys.setenv(SPOTIFY_CLIENT_SECRET = "73efc29aa9b442fdbc83019dfc8fc0ed")
access_token = get_spotify_access_token()
data = get_artist_audio_features("AVADHOOT GUPTE")
```

II. Exploratory Data Analysis
```{r fig.width = 12, message = FALSE, warning = FALSE}
ggpairs(data[c('duration_ms', 'valence', 'danceability', 'energy', 'acousticness', 'loudness',
'speechiness', 'instrumentalness', 'liveness')],
        mapping = aes(fill = sqrt(2)))
```

```{r}
cor(data[c('duration_ms', 'valence', 'danceability', 'energy', 'acousticness', 'loudness', 'speechiness',
'instrumentalness', 'liveness')])
```

```{r message = FALSE, warning = FALSE}
subset_data = data[c('energy', 'loudness')]
ggpairs(subset_data, mapping = aes(fill = sqrt(2)))
```

```{r fig.width = 8}
g1 = ggplot(data, aes(sample = energy)) +
  geom_qq() +
  geom_qq_line() +
  ggtitle("QQ plot of Energy")

g2 = ggplot(data, aes(sample = loudness)) +
  geom_qq() +
  geom_qq_line() +
  ggtitle("QQ plot of Loudness")
```
```

```
grid.arrange(g1, g2, ncol = 2)
```

III. Statistical Analyses
1. Confidence Intervals of ‘Energy’ & ‘Loudness’
```{r}
t.test(data$energy, conf.level = 0.99)$conf.int
t.test(data$loudness, conf.level = 0.99)$conf.int
```

2. Two-sample Hypothesis Test on ‘Energy’
```{r}
sample.means = aggregate(energy ~ mode_name, data = data, FUN = mean)

ggplot(data, aes(x = energy, y = ..density.., group = mode_name, fill = mode_name)) +
  geom_density(alpha = 0.4) +
  geom_vline(data = sample.means, aes(xintercept = energy, color = mode_name)) +
  ggtitle('Mode-wise Energy Distribution')
```

```{r}
major_data = dplyr::filter(data, mode_name == 'major')$energy
minor_data = dplyr::filter(data, mode_name == 'minor')$energy

t.test(major_data, minor_data, conf.level = 0.99)
```

3. ANOVA & Pairwise Hypothesis Test
```{r}
loudness.lm = lm(formula = loudness ~ album_name, data = data)
anova(loudness.lm)
```

```{r fig.width = 12, fig.height = 5}
sample.means = aggregate(loudness ~ album_name, data = data, FUN = mean)

ggplot(data, aes(x = loudness, y = ..density.., group = album_name, fill = album_name)) +
  geom_density(alpha = 0.4) +
  geom_vline(data = sample.means, aes(xintercept = loudness, color = album_name)) +
  ggtitle('Album-wise Loudness Distribution')
```

```{r}
## Tukey HSD Pairwise Hypothesis test
result = data.frame(TukeyHSD(loudness.lm)$album_name)
result["p.adj"]
```

4. Linear Regression
a) Scatterplot of the variables with the correlation between them
```{r message = FALSE, warning = FALSE}
ggplot(data, aes(x = energy, y = loudness)) +
  geom_point() +
  geom_smooth(method = 'lm', col = 'red', se = FALSE) +
  stat_cor() +
  ggtitle("Energy Vs. Loudness (with correlation coefficient)")
```

b) Fitting linear regression model and interpreting the slope & intercept
```{r}
data.lm = lm(formula = loudness ~ energy, data = data)
summary(data.lm)
```

c) Performing hypothesis test to test whether 'energy' is the good predictor
```{r}
pred_loudness = predict(data.lm, energy = data$energy)
t.test(data$loudness, pred_loudness, conf.level = 0.99)
```

d) Predictions at 20th, 40th, 60th & 80th predictor values
```{r}
percentile_predictions = data.frame(energy = quantile(data$energy, probs = c(0.2, 0.4, 0.6, 0.8)))
percentile_predictions$loudness_predictions = predict(data.lm, data.frame(energy =
percentile_predictions$energy))
percentile_predictions
```

e) Checking the regression assumptions
Normality of the residuals
```{r}
data$residuals = residuals(data.lm)

ggplot(data, aes(sample = residuals)) +
  stat_qq() +
  stat_qq_line() +
  ggtitle("QQ-plot of the Residuals")
```

Assessing Mean & Variance of the residuals
```{r message = FALSE, warning = FALSE}
data$fitted = fitted(data.lm)

ggplot(data, aes(x = fitted, y = residuals)) +
  geom_point() +
  geom_smooth(method = "lm", col = "red") +
```

```
  geom_hline(yintercept = 0) +
  ggtitle("Plot to assess Mean & Variance of the Residuals")
```

```{r fig.width = 8}
autoplot(data.lm)
```