Variables Selection :

keep :

"danceability", "energy", "key", "loudness", "mode", "speechiness", "acousticness", "instrumentalness", "liveness", "valence", "tempo", "duration\_ms", "time\_signature", "song\_name", "popularity", "genre"

remove :

Index, type, id, url, track\_href, analysis\_url, artist, classification, error

Linear regression model :

lm(formula = formula(model.final), data = s\_train)

Residuals:

Min 1Q Median 3Q Max

-72.154 9.074 0.332 9.374 45.630

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 59.85324 1.09482 54.669 < 2e-16 \*\*\*

danceability 3.20835 0.67060 4.784 1.72e-06 \*\*\*

loudness -0.11238 0.02596 -4.329 1.50e-05 \*\*\*

speechiness -10.53580 0.95817 -10.996 < 2e-16 \*\*\*

acousticness -2.19344 0.37919 -5.785 7.35e-09 \*\*\*

instrumentalness -10.54805 0.34993 -30.143 < 2e-16 \*\*\*

valence 1.95521 0.41823 4.675 2.95e-06 \*\*\*

time\_signature -1.01222 0.24604 -4.114 3.90e-05 \*\*\*

genreEDM -3.22377 0.30477 -10.578 < 2e-16 \*\*\*

genreHiphop 7.73666 0.35891 21.556 < 2e-16 \*\*\*

genrePop 13.18287 0.28803 45.769 < 2e-16 \*\*\*

genreR&B -6.31465 0.28981 -21.789 < 2e-16 \*\*\*

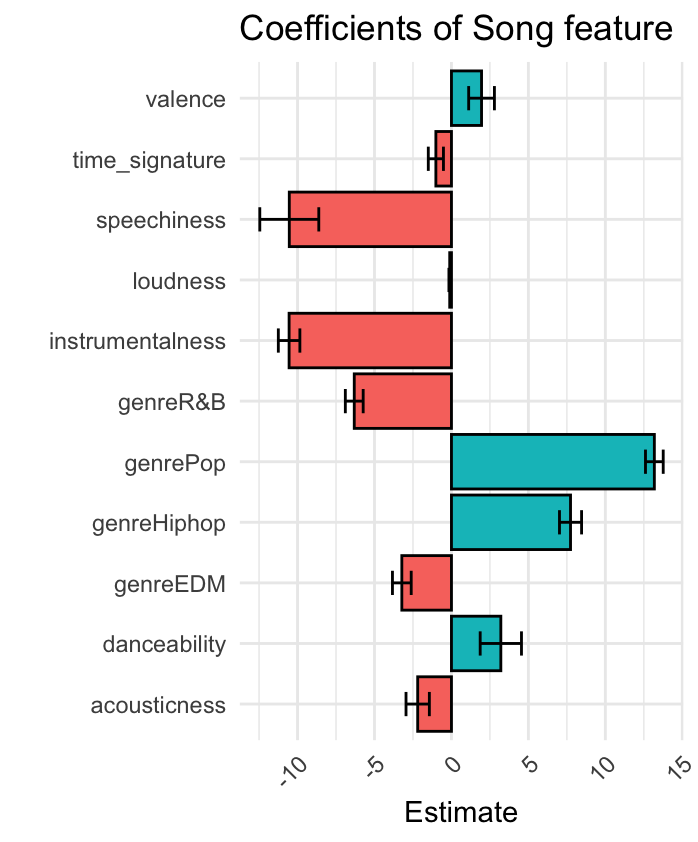
---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.92 on 27849 degrees of freedom

Multiple R-squared: 0.2775, Adjusted R-squared: 0.2772

F-statistic: 972.2 on 11 and 27849 DF, p-value: < 2.2e-16



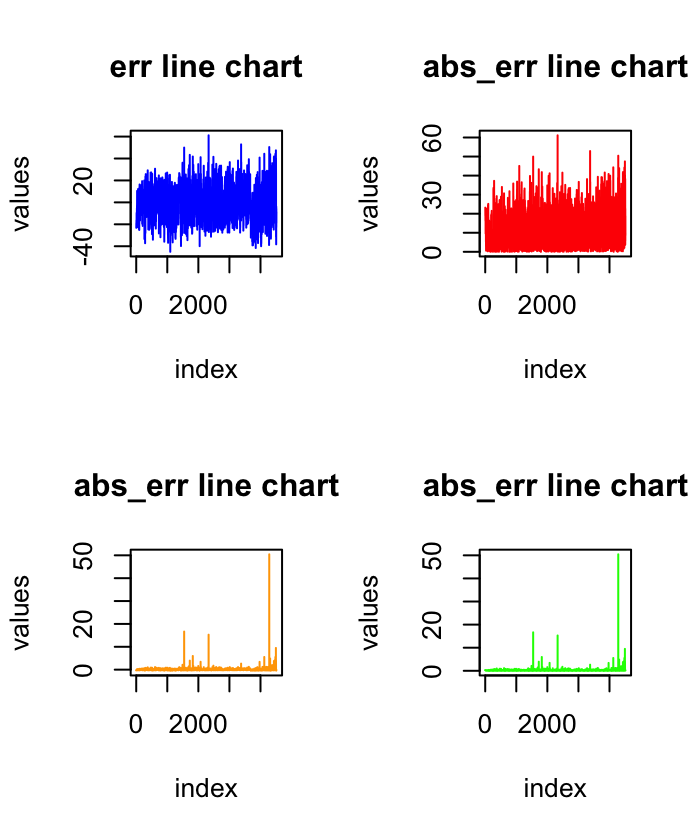
Explaination :

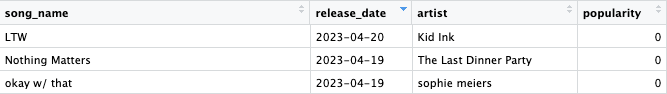
Accuracy :

[1] "New dataset err average 0.0487650962621291"

[1] "New dataset pcterr average 0.100251563290091"

[1] "New dataset abs\_err average 11.0048742989365"

[1] "New dataset abs\_pcter average 0.248453227450233"



We can use those model to predict new song’s popularity. By knowing how each variables would effect the song’s popularity, we can know what kind of music would impact the market well.