Supplementary Results for the Paper of Studying Platformization of Douyin

October 13, 2022

1 Statistics of Critical Features and Dependent Variables

Variables	Minimum/Maximum	Mean	Standard Deviation (Std. D)
peak	2/101	55.65	28.309
duration	1/56	7.655	10.353
timeline	0/1	0.12	0.328
classic	0/1	0.07	0.261
head effects	0.00127/1557.0	42.533	117.985
positive emotion	0/0.5	0.5	0.057
negative emotion	0/0.27	0.0445	0.037
self-similarity	0.067/0.067	0.0233	0.078
tempo harmonic	89.1/172.27	123.34	14.75
tempo percussive	69.84/172.27	123.64	21.97
female voice	0/1	0.510	0.500
male voice	0/1	0.330	0.471
child voice	0/1	0.045	0.207
chorus	0/1	0.104	0.306
pure music	0/1	0.011	0.103
pop	0/1	0.722	0.449
DJ	0/1	0.154	0.362
antique	0/1	0.041	0.199
rock	0/1	0.007	0.085
folk	0/1	0.014	0.119
hip-hop	0/1	0.020	0.139

Table 1: Descriptive Statistics for All Samples (N=557).

Table 1 presents descriptive statistics of all the sample. The expression of positive emotions (M=0.04) is more than that of negative emotions (M=0.02) in Douyin hot songs. Tempo in the harmonic portion (M=123.34) is close to that in the percussive portion (M=123.64) with the rhythm fluctuating around 123 beats per minute, however, the percussive section (Std. D=21.97) has a more dispersive rhythmic distribution with more low and high frequency rhythms than in the harmonic section (Std. D=14.75). Regarding music style, pop (M=0.44) and DJ (M=0.36) account for a higher portion of all style variables. In terms of status expression, majority of Douyin hot songs are sung in female voice (M=0.50) and male voice (M=0.47), with child voice (M=0.045), chorus (M=0.104), and absolute music (M=0.011) accounting for a small portion.

2   Multicollinearity Analysis

Prior to multiple linear regression analysis, we conducted a multicollinearity analysis to test whether variables are linearly correlated. A strong linear relationship between genre "pop" and singer identification "female" showed, which is probably because a big proportion of Douyin hit songs in *Pop* style are performed by female singers so that these two variables always appear simultaneously, resulting in high covariance. Thus, we eliminated these two dummy variables. The multicollinearity estimation and the diagnostic results showed that the variance inflation factor (VIF) values of other variables were lower than 1.6, which are much lower than the cutoff value of 5.0, the maximum level of VIFs value. Hence, remaining variables in our study can be considered free from collinearity issue.

Variables	Tolerance	VIF
tempo harmonic	0.652	1.533
tempo percussive	0.656	1.524
positive emotion	0.942	1.062
negative emotion	0.961	1.041
self-similarity	0.922	1.084
male voice	0.837	1.195
child voice	0.900	1.112
chorus	0.869	1.151
absolute music	0.917	1.090
DJ	0.935	1.070
antique	0.949	1.054
rock	0.947	1.056
folk	0.966	1.035
hip-hop	0.932	1.073
timeline	0.829	1.206
classic	0.864	1.158
head effects	0.900	1.111

Table 2: Multicollinearity Diagnosis in the Regression Model

3   Distribution of Dependent Variables

As shown in Figure 1, the distribution of peak position is approximately uniform but the distribution of cumulative popularity duration is highly skewed. The median is 4 indicating more than 50% of the song are popular for no more than 4 days.
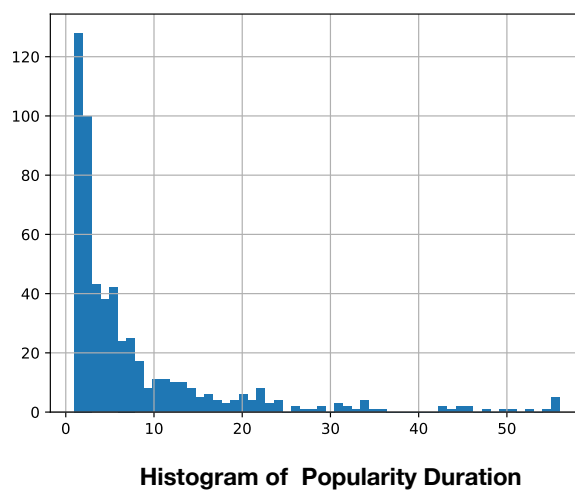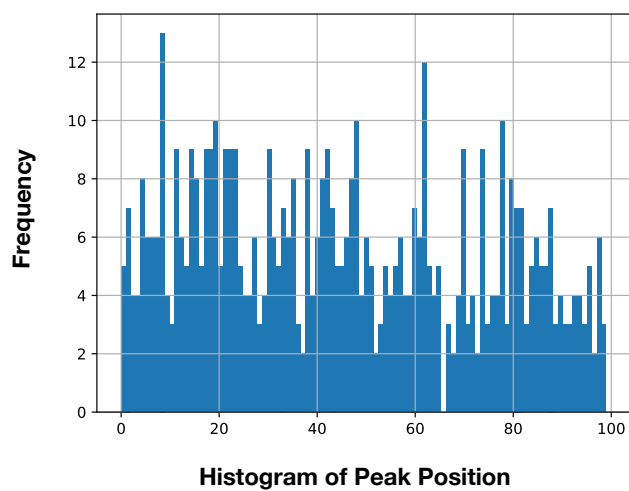
Figure 1: Distribution of Peak Position and Popularity Duration